Wiley Analytical Science

Free Virtual Conference

The 4th edition of the Wiley Analytical Science Conference starts April 26, 2022!

Join virtual seminars and connect with key experts during live Q&A sessions.

Featured Sessions in Lab Automation & Equipment:

- Differential Scanning Calorimerty – Fundamentals and Applications

- Part I: Why So Many Digitalization and Digital Transformation Projects Fail and How This Can Be Avoided
- Part II: Centering Consumables and Reagents in Customers' Awareness

Events.bizzabo.com/WASconferenceSpring22

MERCK METTLER TOLEDO WILEY

Check for updates

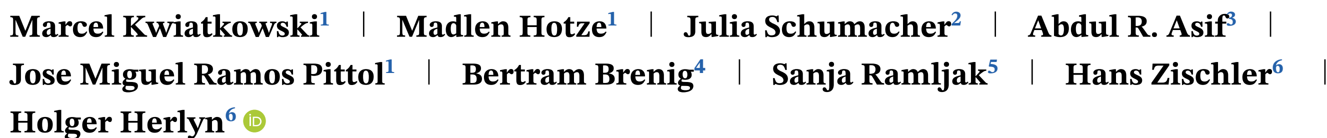
Received: 30 December 2020 | Revised: 30 November 2021 | Accepted: 2 February 2022

DOI: 10.1002/elps.202000393

ELECTROPHORESIS

RESEARCH ARTICLE

Protein speciation is likely to increase the chance of proteins to be determined in 2-DE/MS



Marcel Kwiatkowski[1] | Madlen Hotze[1] | Julia Schumacher[2] | Abdul R. Asif[3] | Jose Miguel Ramos Pittol[1] | Bertram Brenig[4] | Sanja Ramljak[5] | Hans Zischler[6] | Holger Herlyn[6]

[1]Department of Biochemistry and Center for Molecular Biosciences Innsbruck, University of Innsbruck, Innsbruck, Austria

[2]Abbott GmbH, Core Diagnostics, Wiesbaden, Germany

[3]Department of Clinical Chemistry/UMG-Laboratories, University Medical Center, Göttingen, Germany

[4]Department of Molecular Biology of Livestock, Institute of Veterinary Medicine, University of Göttingen, Göttingen, Germany

[5]Digital Diagnostics AG, Mainz, Germany

[6]Institute of Organismic and Molecular Evolution, Anthropology, University of Mainz, Mainz, Germany

Correspondence
Professor Holger Herlyn, Institute of Organismic and Molecular Evolution, Anthropology, University of Mainz, Anselm-Franz-von-Bentzel-Weg 7, 55128 Mainz, Germany.
Email: herlyn@uni-mainz.de

Funding information
University of Mainz

Abstract

Multiple spotting due to protein speciation might increase a protein's chance of being captured in a random selection of 2-DE spots. We tested this expectation in new (PXD015649) and previously published 2-DE/MS data of porcine and human tissues. For comparison, we included bottom-up proteomics studies (BU-LC/MS) of corresponding biological materials. Analyses of altogether ten datasets proposed that amino acid modification fosters multispotting in 2-DE. Thus, the number of 2-DE spots containing a particular protein more tightly associated with a peptide diversity measure accounting for amino acid modification than with an alternative one disregarding it. Furthermore, every 11th amino acid was a post-translational modification candidate site in 2-DE/MS proteins, whereas in BU-LC/MS proteins this was merely the case in every 21st amino acid. Alternative splicing might contribute to multispotting, since genes encoding 2-DE/MS proteins were found to have on average about 0.3 more transcript variants than their counterparts from BU-LC/MS studies. Correspondingly, resolution completeness as estimated from the representation of transcript variant-rich genes was higher in 2-DE/MS than BU-LC/MS datasets. These findings suggest that the ability to resolve proteomes down to protein species can lead to enrichment of multispotting proteins in 2-DE/MS. Low sensitivity of stains and MS instruments appears to enhance this effect.

KEYWORDS
alternative splicing, post-translational modification, protein abundance, protein species, two-dimensional gel electrophoresis

Abbreviations: AS, Alternative splicing; 2-DE/MS, 2-DE followed by mass spectrometry; BU-LC/MS, bottom-up proteomics based on liquid chromatography coupled to MS; IEF, isoelectric focussing; ΔPD, extent to which post-translational modifications increase peptide diversity; FA, formic acid; FDR, False discovery rate; GO, gene ontology; N_{TV}, number of transcript variants per gene; PD, peptide diversity with modifications disregarded; PMI, post-translational modification index; PTM, post-translational modification; RT, room temperature; TPM, transcripts per million.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2022 The Authors. *Electrophoresis* published by Wiley-VCH GmbH.

1 | INTRODUCTION

Two-dimensional gel electrophoresis (2-DE) has proven to be a powerful tool for resolving proteomes [1, 2]. However, it also appears that, depending on the details of a protocol, protein sets from 2-DE/MS can have specific properties. For example, underrepresentation of membrane proteins can occur [3], whereas highly abundant proteins might be overrepresented [4]. Samples from 2-DE/MS may also be enriched or depleted in proteins showing higher molecular mass [5, 6]. Such shifts in the composition are not attributable to 2-DE/MS *per se* [3, 6–8], which offers the advantage of high resolution down to the content of single spots instead [6–8]. The question addressed here is if the occurrence of single proteins in multiple spots might raise their chance of being collected in 2-DE and determined in downstream MS.

Leaving aside differential conformation and complex formation of proteins [9], protein speciation due to post-translational modifications (PTMs) will be a major factor causing multiple spotting of proteins in 2-DE [8, 10–14]. In fact, modifications, such as phosphorylation [15, 16], deamidation [17], oxidization [18], and glycosylation [19, 20], have the potential to drastically raise the number of protein species which partially will differ in MW and pI. Alternative splicing (AS) of mRNAs could be an additional source of multiple spotting [10–13]. The mechanism is likely to be important for cell and tissue differentiation [21–25] and strongly increases transcript diversity [26–28]. In humans, for example, AS produces more than 140 000 transcript variants [29–31] from about 20 300 [14] or 22 500 (Ensembl Genes 103) protein-coding genes. Notwithstanding the nonsense-mediated decay of many transcripts, some of the splice variants will be translated into alternative protein species [32–36] with specific migration patterns in 2-DE. Whether by AS or alternative amino acid modification, the resulting multispotting should increase the probability of a protein being detected in shotgun 2-DE/MS in at least one protein species. In the current study, we test this prediction in new 2-DE/MS data on the porcine testicular proteome. A meta-analysis of five datasets from 2-DE/MS and another five comparative datasets from bottom-up proteomics (BU-LC/MS) follows. Indeed, effects related to the 2D resolution of proteomes should not be reflected in the latter type of studies.

2 | MATERIALS AND METHODS

2.1 | Protein sets

We collected five protein sets from 2-DE/MS and BU-LC/MS studies, each, representing two species of Laurasiatheria (Mammalia), that is, domestic pig (*Sus scrofa*) and anatomical modern human (*Homo sapiens*) [37, 38]. The protein sets covered the soma and germline of single species (e.g., human testis and heart) and corresponding biological material of different species (e.g., human and porcine testis). In this study, we generated a new set of testicular proteins from boars by 2-DE/MS (Supporting information Table S1). The nine other protein sets were retrieved from tabular surveys of previous publications as detailed in Table 1 [39–47] and Supporting information Table S2.

2.2 | Porcine testis proteome: 2-DE, tryptic digestion, and LC-MS/MS analysis

Tissue pieces were cut from the ultra-frozen testicles of three boars. Thawing at 4°C was followed by solubilization of the samples at room temperature (RT) in lysis buffer containing 7 M urea, 2 M thiourea, 4% CHAPS, 1.5% DTT (w/v), 1% ampholytes (pH 3–10; Bio-Rad), and 10 µL protease inhibitor cocktail (Sigma-Aldrich). After pelleting of cell debris by centrifugation (1 min at maximum speed), we purified the supernatant using the ReadyPrep 2-D Cleanup Kit (Bio-Rad) according to the manufacturer's protocol. Proteins were rehydrated in the same buffer as above but without protease inhibitor cocktail. Upon determination of protein concentration (Bradford assay; Bio-Rad), 250–400 µg of total protein was placed on immobilized pH gradient strips (7 cm IPG ReadyStrip, pH 5–8; Bio-Rad), for passive rehydration (2 h at RT). Subsequently, strips were covered with mineral oil for active rehydration for 14 h at 50 V. One-dimensional separation by isoelectric focussing (IEF) was conducted at 20°C in a Protean IEF cell (Bio-Rad) according to a previously published protocol [48]: 2 h at 200 V, 2 h at 500 V, and 5 or 6 h at 4000 V (increments with rapid ramp). We equilibrated IPG strips for 25 min in a buffer containing 6 M urea, 0.375 M Tris (pH 8.8), 30% glycerol, 2% SDS, and 2% DTT. Following this, the strips were exposed for 25 min to the same buffer supplemented with 2.5% iodoacetamide instead of DTT, and 20 µL of 1% bromophenol blue. Equilibrated strips were placed on 10% polyacrylamide gels, followed by electrophoresis at 120 or 140 V. We ran three gels (technical replicates) for each of the three boars (biological replicates). The MW standard used was Kaleidoscope (Bio-Rad). Following staining with colloidal Coomassie-Brilliant Blue G250 (Roti-Blue, Roth), gels were scanned with a GS-800 calibrated densitometer (Bio-Rad).

We excised arbitrarily selected 2-DE spots using biopsy punches. Gel slices were vacuum dried, soaked with digestion buffer supplemented with trypsin (0.01 µg/µL), and incubated overnight at 37°C in the same buffer

TABLE 1 References and methods applied for gathering the protein sets analyzed

Species	Tissues	Methods	Instruments	References
Pig	Testis	2-DE/MS	LC-MS/MS	This study
		BU-LC/MS	LC-MS/MS	[39]
Human	Testis	2-DE/MS	MALDI-TOF	[40]
		BU-LC/MS	LC-MS/MS	[41]
Human	Prefrontal cortex	2-DE/MS	MALDI-TOF	[42]
		BU-LC/MS	LC-MS/MS	[43]
Human	Heart	2-DE/MS	MALDI-TOF	[44]
		BU-LC/MS	LC-MS/MS	[45]
Human	Kidney	2-DE/MS	MALDI-TOF	[46]
		BU-LC/MS	LC-MS/MS	[47]

Note. BU-LC/MS, bottom-up proteomics; 2-DE/MS, two-dimensional gel electrophoresis followed by proteomic analysis of single spots.

devoid of trypsin (5 mM $CaCl_2$, 25 mM NH_4HCO_3). Peptides were extracted the next day by sonication in solvents with increasing ACN contenteration. Following vacuum-centrifugation and subsequent reconstitution in 0.1% formic acid (FA), peptides were separated by RPLC (C18; buffer A: 0.1% FA dissolved in HPLC-H_2O; buffer B: 0.1% FA, dissolved in ACN) using a flow rate of 0.4 µL/min and a gradient of 2–30% in 30 min. Eluting peptides were introduced via an ESI interface into a Q-TOF mass spectrometer (one boar, Q-TOF Ultima, Micromass/Waters, Manchester, UK) and an ion-trap mass spectrometer (two other boars, XCT ion-trap, Agilent Technologies, Waldbronn, Germany), and analyzed as described elsewhere [49, 50].

The XCT ion-trap was equipped with an Agilent HPLC-Chip Cube interface integrated into an 1100 series HPLC system (Agilent Technologies). The HPLC-chip (Large capacity chip, Agilent Technologies) was fitted with an enrichment column (internal volume 160 nL, 5 µm Zorbax 300 SB-C18 material), a separation column (150 mm × 0.075 mm, 5 µm Zorbax 300 SB-C18 material), and a nanospray emitter. In detail, samples were loaded (8 µL/sample) on the enrichment column using a flow rate of 4 µL/min with the mix of the two following mobile phases at a ratio 98:2 (mobile phase A: 0.2% FA in H_2O; mobile phase B: 100% ACN). Tryptic peptides trapped on the enrichment column were disseminated on the separation column using a flow rate of 0.4 µL/min and a linear gradient of 2–40% eluent B within 40 min. For IT-MS analysis, the following parameters were used: MS scanning range from 300 to 2000 m/z, electrospray voltage was set to −1750 V, nitrogen was used as drying gas for desolvation at a flow rate of 4 L/min, and 325°C. Precursor ion mass spectra were acquired in positive ion mode with automated data-dependent MS/MS of the three most intense ions of each precursor MS scan. Doubly charged ions were isolated preferentially. Isolation width was set to 4 Da and MS/MS fragmentation amplitude to 1.25 V. Precursor ions were actively excluded for 1 min. Peak lists for MS/MS database search (*.mgf) were generated using Data Analysis Software for 6300 Series Ion Trap LC/MS version 3.4 (Agilent Technologies) by applying default software settings, except for the following variations: (i) for mass spectrum calculation, additional background subtraction was performed using peak start and end spectra; (ii) after extraction of MS/MS data, spectra were deconvoluted using default parameters.

For Q-TOF analysis, the MS instrument was coupled to a CapLC system (Waters, Eschborn, Germany). A total of 1 µL per sample was loaded onto a µ-precolumn cartridge (C18 pepMap, 300 µm × 5 mm; 5 µm particle size, LC Packings, Germering, Germany) using a flow rate of 5 µL/min and 5% buffer B (buffer A: 0.1% FA prepared in 5% ACN; buffer B: 0.1% FA prepared in 95% ACN). The peptides were separated on an analytical column (C18 pepMap100 nano Series, 75 µm × 15 cm; 3 µm particle size, LC Packings) using a flow rate of 0.25 µL/min and a gradient from 10 to 95% buffer B in 50 min. The Q-TOF Ultima MS system was equipped with a nanoflow ESI Z-spray source. MS analysis was performed in positive ion mode. The nanospray needle was held at 2000 V and the source temperature at 40°C. MS/MS analysis was performed in DDA mode. MS1 scans were recorded over a m/z range of 400–1600. Multiply charged precursor ions were selected in the quadrupole using a selection window of 2 Da, fragmented in the hexapole collision cell using a collision ramp from 20 to 35 eV and their fragment ions were recorded in the TOF mass analyzer over an m/z range of 100–1600. Fragmented precursor ions were excluded from DDA analysis for 1 min. Data acquisition was performed using MassLynx (v 4.0, Waters). MS/MS peak lists (*.pkl) were generated using ProteinLynx Global Server bioinformatics tool (PLGS; v 2.2; Waters) and the following parameters: raw spectra were smoothed (savitzky golay, smooth

window channel: 3, number of smooths: 2), centered (min peak width at half height: 4, centroid to: 80%), noise reduction 10%, medium deisotoping with 3% threshold. The generated pkl-files were converted to mgf-files.

We used ProteomeDiscoverer 2.4 (Thermo Fisher Scientific, San Jose, USA) for peptide and protein identification. Peak lists (*.mgf) were searched with Sequest HT against the *S. scrofa* reference proteome database (UniProt Proteome ID: UP000008227, 49 793 proteins). The searches were performed with precursor mass tolerance of 30 ppm (Q-TOF) or 2 Da (IT) and fragment mass tolerance of 0.1 Da (Q-TOF) or 0.5 Da (IT). For identification, the following variable modifications were considered: oxidation of methionine, deamidation of glutamine and asparagine, phosphorylation of threonine, serine and tyrosine, acetylation of lysine, protein *N*-terminal acetylation, protein *N*-terminal methionine loss, protein *N*-terminal methionine loss plus acetylation. A carbamidomethylation of cysteine residues was considered as a static modification. The Percolator function implemented in Sequest was used to calculate posterior error probabilities and q-values for the identified peptide-spectrum matches. Threshold for correct peptide identification was set to a q-value of 0.05. The peptides determined are listed in Supporting information Table S1. The results were verified exemplarily for some of the samples in the Mascot search engine. We matched the proteins derived in the present study and two previous 2-DE/MS analyses (MALDI-TOF and MALDI-TOF/TOF) of the porcine testicular proteome [51, 52], to recognize reproduced proteins.

2.3 | Bioinformatics and statistics

The applications outlined below were run with Ensembl and UniProt IDs as well as gene symbols. These were either copied directly from publications referenced in Table 1 or had been retrieved beforehand using Ensembl's Biomart, UniProt's ID mapping tool, STRING version 11.0 (https://string-db.org/), or ShinyGO version 0.61 which relies on annotations in Ensembl Genes 96 [53]. For easier handling, we confined analyses of larger gene and protein lists to 2000 randomly selected IDs per study. Thus, the gene lists (Supporting information Table S2) that entered the current bioinformatics pipeline were smaller than in the references from which they were extracted. In addition, some genes and proteins were not matched in single approaches. We considered this unproblematic as it should not have affected the results.

We started with partial correlation analyses of the newly generated porcine data. In the case of two proteins, which had been identified in two protein species, each, we included the longer ones. First, we examined in the newly generated data on the porcine testicular proteome which measure closer associates with the number of 2-DE spots containing a particular protein, peptide diversity with modifications disregarded (PD), or the extent to which modifications increased peptide diversity (ΔPD). Both variables and their basic data are reported in Supporting information Table S1. For the inference of PD, we divided the number of peptides with modifications disregarded (thus, when considering EGWsDSTYGVTK and EGWSDSTYGVTK as a single peptide) by the amino acid span of the particular protein species determined. By doing so, we accounted for the expectation of larger peptide diversity in longer amino acid sequences. For deriving ΔPD, we subtracted the number of peptides with modifications disregarded from the peptide number inclusively modifications (thus, when regarding EGWsDSTYGVTK and EGWSDSTYGVTK as two peptides), prior to division by amino acid span. Partial rank correlations were then carried out between 2-DE spot number on the one hand and either PD or ΔPD on the other hand, using SPSS 23 V5 R. The influence of the uncorrelated third parameter was always controlled for.

Using ShinyGO, we continued with investigating whether frequencies in the number of transcript variants per gene (N_{TV}) differed between the genes encoding porcine testicular proteins from the current 2-DE/MS analysis (query set) and the rest of porcine genes (Chi-square test). We also evaluated whether the N_{TV} frequency distribution in our query set differed from the genome-wide expectation. For comparison, ShinyGO was additionally run on porcine testicular proteins from BU-LC/MS analysis. An analogous procedure was applied to four pairs of 2-DE/MS and BU-LC/MS studies on human testis, kidney, liver, and prefrontal cortex (Table 1).

Using Biomart (Ensembl Genes 103), we retrieved N_{TV} values for single genes in the 2-DE/MS and BU-LC/MS studies which are reported in Supporting information Table S3 along with the IDs of the encoded transcript variants. Based on this, we quantified the difference between mean N_{TV} ($\bar{N}_{TV}$) in 2-DE/MS and BU-LC/MS datasets, separately for each species. We additionally used the N_{TV} values for derivation of the proportion of transcript variant-rich genes, separately for the species and analysis pipelines. For definition of transcript variant-rich genes, we arbitrarily set thresholds between the lower 80% and upper 20% of the N_{TV} ranges in the aforementioned frequency distributions from ShinyGO analyses. Thus, porcine genes having >4 and human genes having >11 protein-coding transcript variants were regarded as transcript variant-rich. Their numbers were divided by the total numbers of protein-coding genes in 2-DE/MS and BU-LC/MS datasets, to obtain proportions of transcript variant-rich genes. Corresponding values were compared

between 2-DE/MS and BU-LC/MS datasets, but also between each of these and shares of transcript variant-rich genes in the entire porcine and human protein-coding genes, respectively. By doing so, we estimated the level of resolution attained in MS analyses.

Downstream steps focused on human proteins due to better data availability. Thus, we collected naturally occurring PTMs from ProteomeScout (http://proteomescout.wustl.edu/compendia; visited on 03/04/2021) [54]. Their number divided by amino acid number was taken as a proxy for the propensity to form protein species due to PTM (Supporting information Table S4). We refer to this proxy as to the PTM index (PMI). In addition, transcript abundances (transcripts per million, TPM) were extracted for human genes (Supporting information Table S5) from the gtexGeneV8 database at the European mirror site of the UCSC Table Browser (http://genome-euro.ucsc.edu/cgi-bin/hgTables). Levels of PMI and TPM were then contrasted between human 2-DE/MS and BU-LC/MS datasets by the MWU test in SPSS.

For the human data, we further tested for correlations between the three parameters collected: PMI, TPM, and N_{TV}. For this purpose, we pooled the proteins from the 2-DE/MS and BU-LC/MS studies of the human tissues and kept only those for which all three parameters were available. Subsequent analyses were again performed as partial rank correlations in SPSS, thus, controlling for the uncorrelated variable once again.

All tests conducted were two-sided. Alpha error rates (*p* values) were converted into false discovery rates (FDRs) applying the method of Benjamini and Hochberg [55], accounting for altogether 17 tests (ten Chi-squared tests, five partial rank correlations, and two MWU tests).

Finally, we performed gene ontology (GO) analyses to assess the physiological involvements of the proteins in the two porcine and four human dataset pairs. For this purpose, we called PANTHER GO-Slim Molecular Function annotations (http://pantherdb.org/, visited on 05/12/2021) for the IDs reported in current Supporting information Table S2.

3 | RESULTS

Present 2-DE/MS analysis of porcine testis led to the detection of 248 protein species from 82 2-DE spots (Figure 1; Supporting information Figures S1-S11). Average coverage was 24%, with a span of 3 to 93% (Supporting information Table S1). Theoretical pI and MW ranged from 4.77 to 11.85 and 14.2 to 116.1 kDa, respectively. Moreover, two proteins (cathepsin D and pre-mRNA-processing factor 19) were determined in two protein species, each, which were derived from alternative transcripts. For 22 of the proteins, we found matches in previous 2-DE/MS studies on porcine testis (Supporting information Table S1).

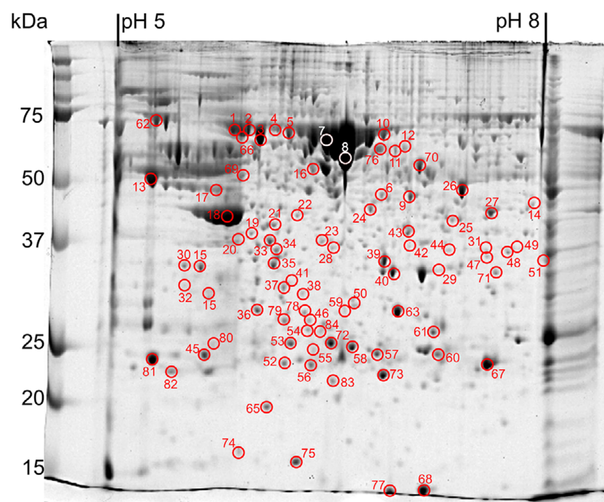


FIGURE 1 Two-dimensional SDS-PAGE gel run with testicular lysate of boar 1. The gel image corresponds to technical replicate A in this boar and is representative of three technical replicates per each of the three boars analyzed (compare Supporting information Figures S2, S3). Circles highlight 2-DE spots examined by MS. Their numbering matches with respective entries in Supporting information Table S1. Raw scans of all technical replicates are provided as Supporting information Figures S3-S11 [Color figure can be viewed at www.electrophoresis-journal.com]

A closer look at the data revealed that amino acid modifications might have raised a protein's chance of being captured. In fact, 46% of the proteins were determined from two or more 2-DE spots (Figure 2; Supporting information Table S1). Subsequent partial rank correlations underscored that variation due to amino acid modifications could have fostered multispotting in 2-DE of the porcine testis proteome (Table 2). Thus, the number of 2-DE spots from which a protein was recovered more tightly associated with ΔPD, that is, the extent to which modifications increased peptide diversity, than with PD, that is, the diversity of peptides with modifications disregarded. This was reflected in a more than threefold larger coefficient in partial rank correlation of 2-DE spot number with ΔPD (rho = 0.429; FDR < 0.001) relative to a corresponding correlation with PD (rho = 0.128; FDR < 0.05). In line with this, peptides with alternative modification patterns were repeatedly detected in different 2-DE spots. For example, TFTDcFNcLPIAAIVDEK and TFTDcFncLPIAAIVDEK were recovered from 2-DE spots nos. 23 and 28, respectively (A0A287A6R4 in Supporting information Table S1).

According to ShinyGO analysis (Ensembl Genes 96), AS could have played a contributory role in the multispotting of porcine testicular proteins. Thus, distributions of N_{TV}

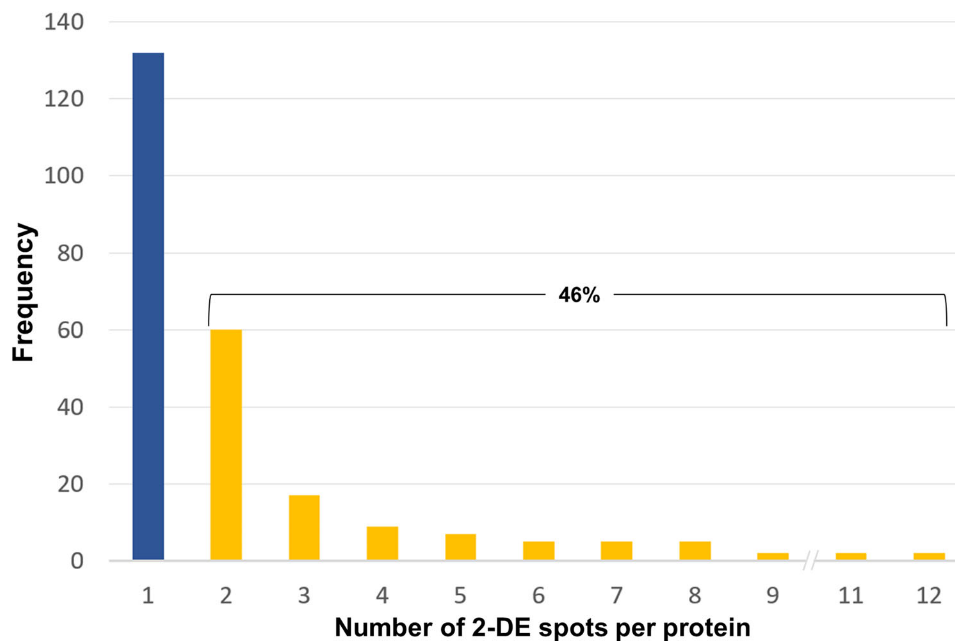


FIGURE 2 Frequency distribution of the number of 2-DE spots, from which a protein was determined, in present 2-DE/MS analysis of the porcine proteome. Out of the 246 proteins determined, 46% occurred in more than a single 2-DE spot (light bars) [Color figure can be viewed at www.electrophoresis-journal.com]

TABLE 2 Results of partial rank correlations

Parameter pair	Controlled variable	$N_{Proteins}$	rho	FDR[a]
ΔPD vs. $N_{2\text{-DE Spots}}$	PD	246	0.429	<0.001
PD vs. $N_{2\text{-DE Spots}}$	ΔPD	246	0.128	<0.05
TPM vs. N_{TV}	PMI	5365	0.133	<0.001
TPM vs. PMI	N_{TV}	5365	0.363	<0.001
N_{TV} vs. PMI	TPM	5365	−0.055	<0.001

Note. ΔPD, degree by which amino acid modifications increased peptide diversity; $N_{2\text{-DE Spots}}$, number of 2-DE spots from which a protein was recovered; N_{TV}, number of transcript variants (per coding gene); PD, peptide diversity; PMI, post-translational modification index; rho, correlation coefficient; TPM, transcripts per million.

[a]False discovery rates as derived from partial rank correlations.

frequencies significantly differed between the genes coding for the proteins determined by us and the rest of the genes in the pig (FDR < 0.001, Chi-square test; Table 3). Compared to genome-wide frequencies, the distribution was shifted toward higher N_{TV} values in genes encoding porcine testicular proteins from present 2-DE/MS analysis (Figure 3A). In their counterparts coding for pig proteins from a BU-LC/MS study, the frequency distribution was also shifted toward higher N_{TV} values, relative to the genome-wide expectation and the rest of porcine genes (FDR < 0.001, Chi-square test; Table 3). Nevertheless, the shift was stronger in the 2-DE/MS than BU-LC/MS sample (Figure 3). This was verified in Ensembl Genes 103 annotations, according to which genes underlying 2-DE/MS-determined porcine proteins ($\bar{N}_{TV}$ = 3.253) on average had 0.265 more transcript variants than the ones encoding proteins from BU-LC/MS analysis of porcine testis ($\bar{N}_{TV}$ = 2.988) (Supporting information Table S3). Consistently, the proportion of transcript variant-rich genes was higher in the 2-DE/MS than in the BU-LC/MS study, although this proportion was increased in both experimental datasets when compared to the genome-wide expectation (Table 4).

We observed similar patterns in 2-DE/MS and BU-LC/MS representing human tissues, particularly testis, but also heart, kidney, and prefrontal cortex. In all corresponding query sets, N_{TV} frequencies differed from the rest of genes in the human genome (FDR < 0.01, Chi-square test, each; Table 3). Furthermore, frequency distributions persistently were more strongly shifted toward higher N_{TV} values in human query sets from 2-DE/MS than BU-LC/MS studies (Supporting information Figure S12). Correspondingly, a human gene coding for a 2-DE/MS-determined protein ($\bar{N}_{TV}$ = 5.806) had on average 0.330 more protein-coding transcript variants than its counterpart encoding a BU-LC/MS protein ($\bar{N}_{TV}$ = 5.476; Supporting information Table S3). Also, the representation of transcript variant-rich genes more strongly exceeded the genome-wide expectation in human genes coding for 2-DE/MS-determined proteins than in the ones encoding proteins from BU-LC/MS studies (Table 4).

TABLE 3 Sample sizes and comparison of per-gene number of transcript variants in query sets with the rest of genes in a given species

Species	Biological material	Methods	$N_{proteins}$[a]	FDR[b]
Pig	Testis	2-DE/MS	230	<0.001
		BU-LC/MS	1396	<0.001
Human	Prefrontal cortex	2-DE/MS	146	<0.001
		BU-LC/MS	2000	<0.001
Human	Heart	2-DE/MS	97	<0.001
		BU-LC/MS	2000	<0.01
Human	Kidney	2-DE/MS	180	<0.001
		BU-LC/MS	1953	<0.001
Human	Testis	2-DE/MS	433	<0.001
		BU-LC/MS	2000	<0.001

Note. BU-LC/MS, bottom-up proteomics based on liquid chromatography coupled to MS; 2-DE-MS, two-dimensional gel electrophoresis followed by MS.
[a]Number of protein-coding genes as recognized by ShinyGO v. 0.61.
[b]False discovery rates as derived from Chi-squared tests comparing N_{TV} distributions between query sets and the rest of genes in a given biological species.

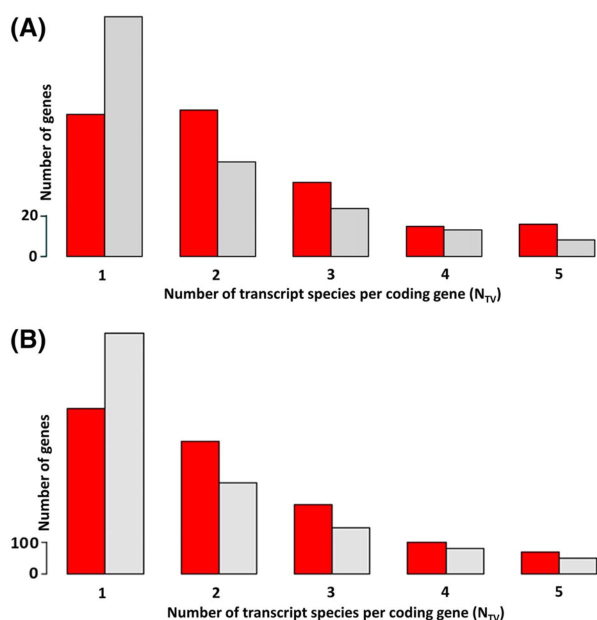


FIGURE 3 Frequency distribution of the per-gene number of transcript variants (N_{TV}) in porcine testicular proteins. (A) The query set contained 230 genes matched by ShinyGO version 0.61, out of 246 ones coding for 2-DE/MS-determined proteins. Distributions of N_{TV} differed between query set (red bars) and the genome-wide expectation (grey bars). (B) The shift toward higher N_{TV} values was less pronounced in a query set including 1396 genes that coded for porcine testicular proteins from a BU-LC/MS reference study [39]. The results shown here are representative of analyses carried out on altogether five study pairs (Supporting information Figure S12) [Color figure can be viewed at www.electrophoresis-journal.com]

Subsequent analyses, which focused on the aforementioned human protein samples due to better data availability, underscored that amino acid modifications might promote multispotting in 2-DE. In particular, pooled human proteins from the four 2-DE/MS studies overall exceeded their BU-LC/MS counterparts in the level of PMI, that is, the number of PTMs per amino acid (FDR < 0.001, MWU test; Table 5; Supporting information Table S4). With a mean PMI ($\overline{PMI}$) of 0.092, every 11th amino acid of a 2-DE/MS protein was a PTM candidate. In contrast, every 21st amino acid in BU-LC/MS proteins might carry a PTM ($\overline{PMI}$ = 0.049). In addition, transcripts had an overall higher abundance (TPM) when encoding 2-DE/MS rather than BU-LC/MS proteins (FDR < 0.001, MWU test; Table 5). On average, transcript abundance was 67% higher

TABLE 4 Proportions of transcript variant-rich genes

Species	Data	Proportion of transcript variant-rich genes
Domestic pig	Whole genome	0.170
	2-DE/MS	0.201
	BU-LC/MS	0.185
Human	Whole genome	0.052
	2-DE/MS	0.093
	BU-LC/MS	0.087

Note. For abbreviations see legend to Table 5.

TABLE 5 Comparison of variation and abundance parameters between human proteins from 2-DE/MS and BU-LC/MS and respective transcripts

Species	Parameter	Method	$N_{Proteins}$	FDR[a]
Human	TPM	2-DE/MS	823	<0.001
		BU-LC/MS	7824	
Human	PMI	2-DE/MS	808	<0.001
		BU-LC/MS	7358	

Note. BU-LC/MS, bottom-up proteomics based on liquid chromatography coupled to MS, 2-DE/MS, two-dimensional gel electrophoresis followed by MS; PMI, post-translational modification index; TPM, transcripts per million.
[a]False discovery rates as derived from MWU tests comparing parameter levels between 2-DE/MS and BU-LC/MS datasets.

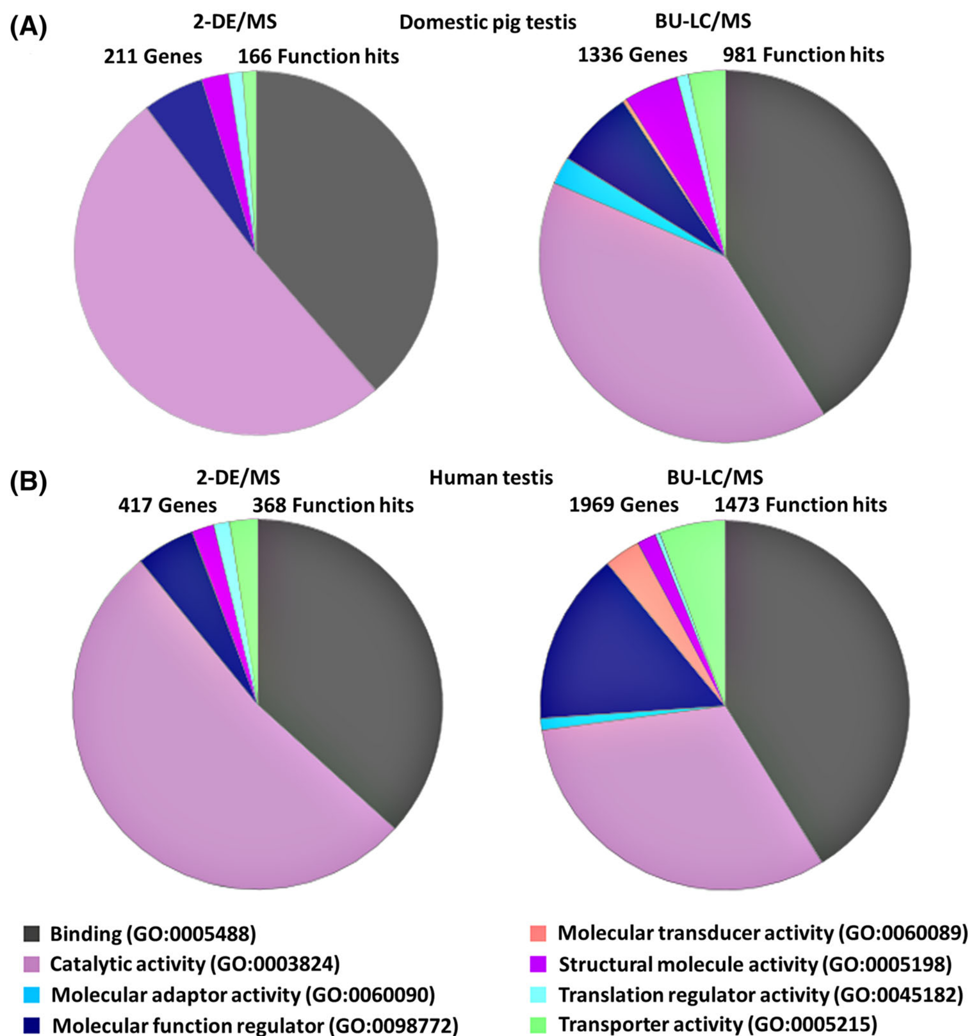


FIGURE 4 Representation of molecular functions in porcine (A) and human testicular proteins (B) as determined by 2-DE/MS (left) and BU-LC/MS. Fewer gene ontology (GO) terms were obtained for 2-DE/MS than BU-LC/MS samples. The emergence of additional GOs in BU-LC/MS datasets (molecular adaptor activity and molecular transducer activity) was accompanied by smaller shares of GOs relating to binding and catalytic activity, when compared to 2-DE/MS datasets. The pattern described was reproduced in altogether five study pairs (see Supporting information Figure S13)

for genes of human 2-DE/MS proteins relative to genes underlying human BU-LC/MS proteins (Supporting information Table S5).

The findings outlined above were paralleled by trends in partial rank correlation analyses of pooled values from 2-DE/MS and BU-LC/MS studies (FDR < 0.001, each; Table 2). In particular, TPM significantly correlated with N_{TV} and PMI in 5365 genes/proteins, for which all three parameters were available. Thus, the higher the transcript abundance was, the more transcript variants a gene tended to have (rho = 0.133). In addition, higher transcript amounts associated with elevated PMI values (rho = 0.363). Last, N_{TV} and PMI were correlated, although with a negligible coefficient (rho = –0.055).

Complementary functional characterization (PANTHER: GO-Slim Molecular Function) revealed five recurring GOs in porcine and human study sets. These were GOs relating to binding, catalytic activity, molecular function regulation, structural molecule activity, and transporter activity (Figure 4, Supporting information Figure S13). Involvements in binding and catalytic activity accounted for at least about 75% of the GO terms. Beyond these similarities, there were clear differences between 2-DE/MS and BU-LC/MS datasets. Specifically, BU-LC/MS sets persistently had more GOs than the corresponding 2-DE/MS sets. These additional GOs in BU-LC/MS samples were molecular transducer activity and molecular adaptor activity. We take the results of the GO analyses as post-hoc affirmation that the 2-DE/MS and BU-LC/MS studies were each sufficiently homogeneous to conduct the comparisons between both groups as outlined above.

4 | DISCUSSION

The usefulness of 2-DE for separating proteins [1, 2] and protein species [10–13, 56, 57] is well established. In the current investigation (Figure 1; Supporting information Figures S1–S11), 2-DE/MS enabled the detection of 248 protein species representing 246 proteins (Supporting information Table S1), of which 22 were previously discovered in porcine testis using 2-DE/MS [51, 52]. More importantly, the present findings add to previous hints that, depending on the detailed protocol applied, 2-DE/MS might affect the composition of protein samples [3–5]. Thus, present results suggest that shotgun 2-DE/MS could preferentially determine multispotting proteins. Consistent with such a possibility, a considerable percentage of proteins from porcine testis were identified from more than a single 2-DE spot (Supporting information Table S1). The exact percentage of such multiple determinations (Figure 2; 46%) was in the range of previous 2-DE/MS-based investigations, including those on glioblastoma (32%) [10], HeLa cells (50%) [58], and *Mycobacterium tuberculosis* (55%) [8]. In this respect, the present study could be representative of a general trend.

As mentioned before, protein speciation through amino acid modification and AS can cause multispotting in 2-DE [56]. Consistently, we found the number of spots from which a porcine testicular protein was recovered to be more closely associated with a peptide diversity measure which accounted for alternative amino acid modification patterns than with a measure which did not (Table 2). It should be noted though that, with a maximum of 93%, the detected peptides only incompletely covered the protein species determined in the current 2-DE/MS. Potentially, this reflects their overall larger size compared to protein species for which 100% sequence coverage was achieved elsewhere. In fact, complete coverage was attained in previous 2-DE/MS studies for protein species of up to about 14 kDa [8, 59], whereas such a MW marked the lower boundary in the present examination of the porcine testicular proteome. Yet, almost all protein species determined by us had larger MWs of up to about 115 kDa (Supporting information Table S1). In addition, the resolution of 2-DE was higher in the reference studies [8, 59] than in the present one (Figure 1; Supporting information Figures S1–11). Furthermore, a combination of cleavage reagents is considered to be favorable to attain 100% sequence coverage [60] while the porcine testis proteome was merely trypsin-digested in the present investigation. Either way, we do not expect that complete sequence coverage would change the results of the present correlation analyses. This is because the current proxies for addressing the issue do not depend on complete sequence coverage. The same applies to our proxy for assessing the level of protein speciation due to AS, that is, N_{TV}.

In further support of a prominent role of amino acid modifications in multiple spotting, we noticed that human proteins from 2-DE/MS are more likely to occur in multiple PTM variants than their counterparts from BU-LC/MS analyses (PMI in Table 2; Supporting information Table S4). Thus, every 11th amino acid site was found to be a PTM candidate in human 2-DE/MS proteins on average, while it was every 21st site in BU-LC/MS-determined ones. Compared to this, the contribution of AS to multispotting is probably small. Indeed, many proteins have only a single dominant isoform despite the occurrence of additional transcript variants [31, 33]. This is matched in the present study by only a slight increase in the N_{TV} (per coding gene) in the 2-DE/MS samples compared to the BU-LC/MS samples (Figure 3; Table 3; Supporting information Figure S12; Supporting information Table S3). Nevertheless, the increase was traceable (plus ~0.3) and our transcript-based proxy giving resolution in electrophoresis was also higher in the 2-DE/MS than BU-LC/MS studies (Table 4). Thus, elevated protein speciation parameters probably reflect an overrepresentation of protein species-rich proteins in 2-DE/MS and not their underrepresentation in BU-LC/MS.

In addition to the occurrence in variants, present findings (Table 5) suggest that higher protein abundance, as approximated by TPM, can raise the determination probability of a protein in 2-DE/MS [4]. However, such an effect is not an inherent property of 2-DE/MS but reflects modest sensitivity of a particular pipeline rather. Indeed, the detection limits of the staining protocols applied in the 2-DE/MS studies included here were in the low nanogram range ([40, 42, 44, 46], present study). Compared to this, the sensitivity of MS instruments—which is the limiting factor in BU-LC/MS—is an order of magnitude higher. For example, ESI-QTOF spectrometers may detect peptides down to 10–100 fmol and modern Orbitrap MS instruments can trace peptides in concentrations of 1–10 amol [11, 61]. It is meanwhile even possible to determine more than 400 proteins from a single 2-DE spot with modern MS instruments [10–13]. This is far from what has been reached in the 2-DE/MS analyses considered here ([40, 42, 44, 46], Supporting information Table S1). Probably for this reason, fewer GO terms were represented in protein sets from 2-DE/MS than from BU-LC/MS pipelines included in the current meta-analysis (Figure 4; Supporting information Figure S13). In fact, the additional GOs in the BU-LC/MS studies referred to molecular adapter activity and molecular transducer activity and, thus, to processes which usually are exerted by low-abundance proteins.

5 | CONCLUDING REMARKS

Present findings suggest that random or arbitrary selection of 2-DE spots can result in an overrepresentation of multispotting proteins in downstream MS analysis. The deeper cause for this is probably protein speciation due to alternative amino acid modification patterns and, to a lesser extent, AS. Although demonstrated here for the pig and human, we expect the effect to be valid beyond the taxonomic group they represent, Laurasiatheria [37, 38]. This is because PTM and AS have probably evolved, at least in their basic characteristics, prior to the divergence of Eukaryota [26, 62–67]. The occurrence of protein speciation in other taxa, such as *Helicobacter pylori*, could support this view [3]. However, the explanatory approach taken here should not depend on whether multispotting is due to naturally occurring protein speciation which has evolved once in a common stem line or severalfold independently in the tree of life. Rather, artificial amino acid modifications, which can occur during sample preparation, may also contribute to multispotting. We have accounted for both possibilities in present analysis of mass spectra by including natural (e.g., STY phosphorylation) and artificial modifications (e.g., M oxidation).

Furthermore, we do not claim that preferential determination of multispotting proteins is reflected in every 2-DE/MS study. In particular, the combination of high-resolution 2-DE with high-sensitivity LC-MS/MS raises the prospect that proteomes will be detected entirely, or at least almost completely [10–13, 58]. This should counteract potential biases in the composition of 2-DE/MS-determined protein samples. Still, awareness that shotgun 2-DE/MS using less sensitive staining protocols and MS instruments can lead to an overrepresentation of protein species-rich proteins might prove useful in the future. Future investigations might also address whether high-abundance proteins in general tend to be rich in protein species, much as the present results suggest (Table 2). Finally, the proportion of transcript variant-rich genes could turn out to be a useful measure for assessing the degree of resolution reached. It should be advantageous here that 100% sequence coverage in MS analyses is not a prerequisite for using this measure and that representative data on transcript variants are publicly available for numerous biological species.

ACKNOWLEDGEMENTS

Funding was provided by University of Mainz (Stufe-1 to HH). We thank two anonymous reviewers and the editor handling the review process for constructive comments and useful suggestions. We are further grateful to Mario Dejung (IMB, Mainz) for technical assistance.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

New MS data on the porcine testicular proteome have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with record ID PXD015649. Identifiers and symbols of the corresponding proteins and genes are listed in Supporting information Tables S1 and S2. Supporting information Table S2 additionally contains ID lists for the other nine protein sets analyzed, all of which were taken from Tables in previous publications [39–47]. Transcript IDs are reported in Supporting information Table S3.

ORCID

Holger Herlyn https://orcid.org/0000-0003-3157-7243

REFERENCES

1. Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik. 1975;26:231–43.
2. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. J Biol Chem. 1975;250:4007–21.
3. Jungblut PR, Schiele F, Zimny-Arndt U, Ackermann R, Schmid M, Lange S, et al. *Helicobacter pylori* proteomics by 2-DE/MS, 1-DE-LC/MS and functional data mining. Proteomics. 2010;10:182–93.
4. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci USA. 2000;97:9390–5.
5. Klont F, Bras L, Wolters JC, Ongay S, Bischoff R, Halmos GB, et al. Assessment of sample preparation bias in mass spectrometry-based proteomics. Anal Chem. 2018;90:5405–13.
6. Schmidt F, Donahoe S, Hagens K, Mattow J, Schaible UE, Kaufmann SH, et al. Complementary analysis of the *Mycobacterium tuberculosis* proteome by two-dimensional electrophoresis and isotope-coded affinity tag technology. Mol Cell Proteomics. 2004;3:24–42.
7. Jungblut P, Thiede B, Zimny-Arndt U, Müller EC, Scheler C, Wittmann-Liebold B, et al. Resolution power of two-dimensional electrophoresis and identification of proteins from gels. Electrophoresis. 1996;17:839–47.
8. Lange S, Rosenkrands I, Stein R, Andersen P, Kaufmann SH, Jungblut PR. Analysis of protein species differentiation among mycobacterial low-Mr-secreted proteins by narrow pH range Immobiline gel 2-DE-MALDI-MS. J Proteomics. 2014;97:235–44.
9. Deng X, Hahne T, Schröder S, Redweik S, Nebija D, Schmidt H, et al. The challenge to quantify proteins with charge trains due to isoforms or conformers. Electrophoresis. 2012;33:263–9.
10. Zhan X, Yang H, Peng F, Li J, Mu Y, Long Y, et al. How many proteins can be identified in a 2DE gel spot within an analysis of a complex human cancer tissue proteome? Electrophoresis. 2018;39:965–80.
11. Zhan X, Li N, Zhan X, Qian S. Revival of 2DE-LC/MS in proteomics and its potential for large-scale study of human proteoforms. Med One. 2018;3:e180008.

12. Zhan X, Li B, Zhan X, Schlüter H, Jungblut PR. Innovating the concept and practice of two-dimensional gel electrophoresis in the analysis of proteomes at the proteoform level. Coorssen J R Proteomes. 2019;7:36.
13. Naryzhny S, Klopov N, Ronzhina N, Zorina E, Zgoda V, Kleyst O, et al. A database for inventory of proteoform profiles: "2DE-pattern". Electrophoresis. 2020;41:1118–24.
14. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, et al. How many human proteoforms are there? Nat Chem Biol. 2018;14:206–14.
15. Lim YP. Mining the tumor phosphoproteome for cancer markers. Clin Cancer Res. 2005;11:3163–9.
16. Raggiaschi R, Gotta S, Terstappen GC. Phosphoproteome analysis. Biosci Rep. 2005;25:33–44.
17. Sarioglu H, Lottspeich F, Walk T, Jung G, Eckerskorn C. Deamidation as a widespread phenomenon in two-dimensional polyacrylamide gel electrophoresis of human blood plasma proteins. Electrophoresis. 2000;21:2209–18.
18. Kleinert P, Kuster T, Arnold D, Jaeken J, Heizmann CW, Troxler H. Effect of glycosylation on the protein pattern in 2-D-gel electrophoresis. Proteomics. 2007;7:15–22.
19. Barrabés S, Sarrats A, Fort E, De Llorens R, Rudd PM, Peracaula R. Effect of sialic acid content on glycoprotein pI analyzed by two-dimensional electrophoresis. Electrophoresis. 2010;31:2903–12.
20. Zhou H, Liu Y, Chui J, Guo K, Shun Q, Lu W, et al. Investigation on glycosylation patterns of proteins from human liver cancer cell lines based on the multiplexed proteomics technology. Arch Biochem Biophys. 2007;459:70–8.
21. Venables JP. Aberrant and alternative splicing in cancer. Cancer Res. 2004;64:7647–54.
22. Tazi J, Bakkour N, Stamm S. Alternative splicing and disease. Biochim Biophys Acta. 2009;1792:14–26.
23. Ward AJ, Cooper TA. The pathobiology of splicing. J Pathol. 2010;220:152–163.
24. Scotti MM, Swanson MS. RNA mis-splicing in disease. Nat Rev Genet. 2016;17:19–32.
25. Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO. Alternative splicing and the evolution of phenotypic novelty. Philos Trans R Soc London B Biol Sci. 2017;372:20150474.
26. Chen SY, Li C, Jia X, Lai SJ. Sequence and evolutionary features for the alternatively spliced exons of eukaryotic genes. Int J Mol Sci. 2019;20:3834.
27. Schreiber K, Csaba G, Haslbeck M, Zimmer R. Alternative splicing in next generation sequencing data of *Saccharomyces cerevisiae*. PLoS One. 2015;10, e0140487.
28. Grau-Bové X, Ruiz-Trillo I, Irimia M. Origin of exon skipping-rich transcriptomes in animals driven by evolution of gene architecture. Genome Biol. 2018;19:135.
29. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40:1413–5.
30. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456:470–6.
31. Gonzàlez-Porta M, Frankish A, Rung J, Harrow J, Brazma A. Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. Genome Biol. 2013;14, R70.
32. Irimia M, Roy SW. Origin of spliceosomal introns and alternative splicing. Cold Spring Harb Perspect Biol. 2014;6, a016071.
33. Ezkurdia I, Rodriguez JM, Carrillo-de Santa Pau E, Vázquez J, Valencia A, Tress ML. Most highly expressed protein-coding genes have a single dominant isoform. J Proteome Res. 2015;14:1880–7.
34. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet. 2010;11:345–55.
35. Mudge JM, Frankish A, Fernandez-Banet J, Alioto T, Derrien T, Howald C, et al. The origins, evolution, and functional potential of alternative splicing in vertebrates. Mol Biol Evol. 2011;28:2949–59.
36. Tress ML, Abascal F, Valencia A. Alternative splicing may not be the key to proteome complexity. Trends Biochem Sci. 2017;42:98–110.
37. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. Molecular phylogenetics and the origins of placental mammals. Nature. 2001;409:614–8.
38. Stuart GW, Moffett K, Leader JJ. A Comprehensive vertebrate phylogeny using vector representations of protein sequences from whole genomes. Mol Biol Evol. 2002;19:554–62.
39. Ma R, Zhang Y, Liu H, Ning P. Proteome profile of swine testicular cells infected with porcine transmissible gastroenteritis coronavirus. PLoS One. 2014;9:e110647.
40. Guo X, Zhao C, Wang F, Zhu Y, Cui Y, Zhou Z, et al. Investigation of human testis protein heterogeneity using 2-dimensional electrophoresis. J Androl. 2010;31:419–29.
41. Zhang Y, Li Q, Wu F, Zhou R, Qi Y, Su N, et al. Tissue-based proteogenomics reveals that human testis endows plentiful missing proteins. J Proteome Res. 2015;14:3583–94.
42. McManus CA, Polden J, Cotter DR, Dunn M. Two-dimensional reference map for the basic proteome of the human dorsolateral prefrontal cortex (dlPFC) of the prefrontal lobe region of the brain. J Proteomics. 2010;10:2551–5.
43. Cabello-Arreola A, Ho AM, Ozerdem A, Cuellar-Barboza AB, Kucuker MU, Heppelmann CJ, et al. Differential dorsolateral prefrontal cortex proteomic profiles of suicide victims with mood disorders. Genes. 2020;11:256.
44. Westbrook JA, Wheeler JX, Wait R, Welson SY, Dunn MJ. The human heart proteome: Two-dimensional maps using narrow-range immobilised pH gradients. Electrophoresis. 2006;27:1547–55.
45. Doll S, Dreßen M, Geyer PE, Itzhak DN, Braun C, Doppler SA, et al. Region and cell-type resolved quantitative proteomic map of the human heart. Nat Commun. 2017;8:1469.
46. Yoshida Y, Miyazaki K, Kamiie J, Sato M, Okuizumi S, Kenmochi A, et al. Two-dimensional electrophoretic profiling of normal human kidney glomerulus proteome and construction of an extensible markup language (XML)-based database. Proteomics. 2005;5:1083–96.
47. Sigdel TK, Piehowski PD, Roy S, Liberto J, Hansen JR, Swensen AC, et al. Near-single-cell proteomics profiling of the proximal tubular and glomerulus of the normal human kidney. Front Med. 2020;7:499.
48. Ramljak S, Asif AR, Armstrong VW, Wrede A, Groschup MH, Buschmann A, et al. Physiological role of the cellular prion protein (PrP c): protein profiling study in two cell culture systems. J Proteome Res. 2008;7:2681–95.

49. Trusch M, Tillack K, Kwiatkowski M, Bertsch A, Ahrends R, Kohlbacher O, et al. Displacement chromatography as first separating step in online two-dimensional liquid chromatography coupled to mass spectrometry analysis of a complex protein sample–the proteome of neutrophils. J Chromatogr A. 2012;1232:288–94.
50. Kwiatkowski M, Krösser D, Wurlitzer M, Steffen P, Barcaru A, Krisp C, et al. Application of displacement chromatography to online two-dimensional liquid chromatography coupled to tandem mass spectrometry improves peptide separation efficiency and detectability for the analysis of complex proteomes. Anal Chem. 2018;90:9951–58.
51. Huang S-Y, Lin J-H, Chen Y-H, Chuang C-K, Lin E-C, Huang M-C, et al. A reference map and identification of porcine testis proteins using 2-DE and MS. Proteomics. 2005;5:4205–12.
52. Huang SY, Lin JH, Teng SH, Sun HS, Chen YH, Chen HH, et al. Differential expression of porcine testis proteins during postnatal development. Anim Reprod Sci. 2011;123:221–33.
53. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics 2020;36:2628–29.
54. Matlock MK, Holehouse AS, Naegle KM. ProteomeScout: a repository and analysis resource for post-translational modifications and proteins. Nucleic Acids Res. 2015;43:D521–30.
55. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57:289–300.
56. Aslebagh R, Channaveerappa D, Arcaro KF, Darie CC. Comparative two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) of human milk to identify dysregulated proteins in breast cancer. Electrophoresis. 2018;39:1723–34.
57. Hentschker C, Dewald C, Otto A, Büttner K, Hecker M, Becher D. Global quantification of phosphoproteins combining metabolic labeling and gel-based proteomics in B. pumilus. Electrophoresis. 2018;39:334–43.
58. Thiede B, Koehler CJ, Strozynski M, Treumann A, Stein R, Zimny-Arndt U, et al. High resolution quantitative proteomics of HeLa cells protein species using stable isotope labeling with amino acids in cell culture (SILAC), two-dimensional gel electrophoresis (2DE), and nano-liquid chromatograpohy coupled to an LTQ-OrbitrapMass spectrometer. Mol Cell Proteomics. 2013;12:529–38.
59. Okkels LM, Müller EC, Schmid M, Rosenkrands I, Kaufmann SH, Andersen P, et al. CFP10 discriminates between nonacetylated and acetylated ESAT-6 of *Mycobacterium tuberculosis* by differential interaction. Proteomics. 2004;4:2954–60.
60. Meyer B, Papasotiriou DG, Karas M. 100% protein sequence coverage: a modern form of surrealism in proteomics. Amino Acids. 2011;41:291–310.
61. Miller I, Crawford J, Gianazza E. Protein stains for proteomic applications: which, when, why? Proteomics. 2006;6:5385–5408.
62. Irimia M, Rukov JL, Penny D, Roy SW. Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. BMC Evol Biol. 2007;4:188.
63. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. Nature. 2010;463:457–63.
64. Roy SW, Irimia M. Splicing in the eukaryotic ancestor: form, function and dysfunction. Trends Ecol Evol. 2009;24:447–55.
65. Rogozin IB, Carmel L, Csuros M, Koonin EV. Origin and evolution of spliceosomal introns. Biol Direct. 2012;7:11.
66. Beltrao P, Bork P, Krogan NJ, van Noort V. Evolution and functional cross-talk of protein post-translational modifications. Mol Syst Biol. 2013;9:714.
67. Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB. Identification and analysis of alternative splicing events conserved in human and mouse. Proc Natl Acad Sci USA. 2005;102:2850–5.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Kwiatkowski M, Hotze M, Schumacher J, Asif AR, Pittol JMR, Brenig B, et al. Protein speciation is likely to increase the chance of proteins to be determined in 2-DE/MS. Electrophoresis. 2022;1–12.
https://doi.org/10.1002/elps.202000393