

**Characterization of chondrogenic progenitor cells
using mass spectrometry and
multidimensional separation approaches**

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

“Doktor rerum naturalium“

der Georg-August-Universität Göttingen

im Promotionsprogramm Biologie

der Georg-August-University School of Science (GAUSS)

vorgelegt von

Ilian Atanassov

aus Sofia/Bulgarien

Göttingen 2014

Mitglieder des Betreuungsausschusses:

Prof. Michael Kessel	Department of Developmental Biology, Max Planck Institute for Biophysical Chemistry, Göttingen
Prof. Henning Urlaub	Department of Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Göttingen
Prof. Nicolai Miosge	Department of Oral Biology and Tissue Regeneration, UMG Göttingen

Mitglieder der Prüfungskommission:

Referent: Prof. Michael Kessel	Department of Developmental Biology, Max Planck Institute for Biophysical Chemistry, Göttingen
Koreferent: Prof. Henning Urlaub	Department of Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Ahmed Mansouri	Department of Biological Cell Differentiation, Max Planck Institute for Biophysical Chemistry, Göttingen
Prof. Peter Rehling	Department of Biochemistry II, Georg- August University, Göttingen
Prof. Kai Tittmann	Department of Bioanalytics, Max Planck Institute for Biophysical Chemistry, Göttingen
Prof. Jürgen Wienands	Department of Cellular and Molecular Immunology, Georg-August University, Göttingen

Tag der mündlichen Prüfung: 11.07.2014

Table of contents

1	Introduction	3
1.1	Proteomics	3
1.1.1	Mass spectrometry as an analytical tool in proteomics	4
1.2	Mass spectrometry based proteomics	5
1.2.1	Electrospray ionization	5
1.2.2	Mass measurement	6
1.2.2.1	Mass measurement instruments	6
1.2.2.1.1	Quadrupole and linear ion trap	6
1.2.2.1.2	Orbitrap	7
1.2.2.1.3	Hybrid Linear Ion Trap – orbitrap	8
1.2.3	Protein and peptide fragmentation in the mass spectrometer	8
1.2.3.1	Fragmentation by collision induced dissociation	8
1.2.4	Protein and peptide identification by mass spectrometry	9
1.3	Quantitative mass spectrometry based proteomics	10
1.4	Protein and peptide separation in proteomics	11
1.4.1	Peptide separation	11
1.4.1.1	Reversed-phase liquid chromatography	11
1.4.1.2	Peptide isoelectric focusing	12
1.4.2	Protein separation	13
1.4.2.1	One-dimensional polyacrylamide gel electrophoresis	13
1.4.2.2	Two-dimensional electrophoresis	13
1.5	Cellular differentiation	14
1.6	Cellular environment, shape and differentiation	14
1.7	Chondrogenesis	16
1.8	<i>In vivo</i> chondrogenesis	16
1.8.1	Articular cartilage	18
1.8.2	<i>In vitro</i> chondrogenesis	18
1.8.3	Chondrogenic progenitor cells and osteoarthritis	19
1.9	Aims of the study	20
2	Materials and Methods	22
2.1	Chemicals	22
2.2	Laboratory equipment and instruments	22
2.3	Chromatography, commercial kits and solutions	22
2.4	Two-dimensional cell culture of CPCs (CPC-2D)	23

2.5	Metabolic labeling of CPCs.....	23
2.6	Three-dimensional CPC culture in alginate beads (CPC-3D)	23
2.7	Extraction of proteins from CPCs and protein separation.....	24
2.7.1	Biological replicate one; 1DE and 2DE analysis	24
2.7.2	Biological replicate two; 1DE analysis	25
2.8	Ethanol precipitation	25
2.9	HeLa cell nuclear extract sample preparation.....	26
2.10	In-gel trypsin digestion	26
2.11	Peptide extraction	27
2.12	In-solution digestion	27
2.13	pIEF-LC-MS/MS analysis.....	27
2.14	PAGE-pIEF-LC-MS/MS analysis	28
2.15	Comparison of 7 cm pIEF-LC-MS/MS and long LC gradient	28
2.16	PAGE-pIEF-LC-MS/MS analysis of CPC culture media	29
2.17	LC-MS/MS analysis	29
2.18	Protein identification and quantification from MS data	30
2.19	Identification of differentially abundant proteins	31
2.20	Functional enrichment analysis of differentially expressed proteins	31
2.21	1DE and 2DE protein intensity and ratio H/L ratio data analysis and comparison	32
2.22	Calculation of average protein GRAVY scores and pI	32
2.23	Protein sequence similarity analysis	32
3	Results.....	34
3.1	Quantification of proteome wide differences between CPCs cultured on plastic and in alginate	34
3.2	Evaluation of 2DE separation for large scale proteome profiling.....	42
3.3	Comparison of 1DE and 2DE approaches for proteome analysis.....	52
3.3.1	Comparison of the pIs of proteins identified using 1DE and 2DE	53
3.3.2	Comparison of the hydrophobicity of proteins identified using 1DE and 2DE	54
3.3.3	Comparison of the number of identified protein isoforms using 1DE and 2DE ..	56
3.3.4	Identification and quantification of protein species using 1DE and 2DE	66
3.4	Establishment of three dimensional PAGE-pIEF-LC-MS/MS separation method 86	
3.5	Identification of secreted signaling proteins from CPC cell culture medium	99
3.6	Optimization of the pIEF fractionation scheme.....	104

4	Discussion.....	111
4.1	Quantitative proteome differences between CPCs grown on 2D plastic and in 3D alginate.....	111
4.2	Protein identification and quantification using 2DE.....	113
4.3	Establishment and evaluation of PAGE-pIEF-LC-MS/MS.....	115
4.4	Identification of secreted proteins from CPC culture medium.....	115
4.5	Optimization of pIEF fractionation.....	117
5	References	118
6	Appendixes.....	133

List of Figures

FIGURE 1. QUANTIFICATION OF PROTEINS IN CPCs GROWN ON PLASTIC (CPC-2D) AND CPCs GROWN IN ALGINATE (CPC-3D).....	35
FIGURE 2 2DE ANALYSIS OF POOLED PROTEIN EXTRACTS OF HEAVY SILAC LABELED CPC-2D AND LIGHT SILAC LABELED CPC-3D.	44
FIGURE 3. NUMBER OF IDENTIFIED PROTEINS PER 2DE GEL FRACTION.	46
FIGURE 4. HISTOGRAM OF THE NUMBER OF IDENTIFIED PROTEINS PER 2DE GEL SLICE.....	47
FIGURE 5. HEATMAP OF THE NORMALIZED INTENSITY (A) AND RATIO H/L (B) OF PROTEIN ATP5O.	48
FIGURE 6. DISTRIBUTION OF NORMALIZED PROTEINS INTENSITIES OVER THE 2DE GEL.	49
FIGURE 7. IDENTIFIED PROTEINS PER 2DE GEL SLICE SHOWING MAXIMUM (100%) INTENSITY.....	50
FIGURE 8. NUMBERS OF PROTEINS SHOWING 100% INTENSITY IN EACH COLUMN (A) OR IN EACH ROW (B) OF THE 2DE GEL.	51
FIGURE 9. DISTRIBUTION OF IDENTIFIED PROTEINS SHOWING 100% NORMALIZED INTENSITY PER 2DE FRACTION.....	52
FIGURE 10. DISTRIBUTION THE PI OF PROTEINS IDENTIFIED USING THE 1DE AND 2DE APPROACHES.	53
FIGURE 11. DISTRIBUTION OF GRAVY SCORES OF ALL PROTEINS IDENTIFIED IN THE 1DE AND 2DE ANALYSIS.	54
FIGURE 12. SEQUENCE ALIGNMENT OF IDENTIFIED PROTEIN ISOFORMS FROM THE GENE HMGA1 (A) AND KGUA (B).....	58
FIGURE 13. DISTRIBUTION OF PROTEIN PEP VALUES FROM ALL ISOFORMS IDENTIFIED IN THE 1DE ANALYSIS OF FIRST BIOLOGICAL CPC-2D/CPC-3D REPLICATE.	61
FIGURE 14. INTENSITY DISTRIBUTION OF TWO ISOFORMS OF COL6A2.	63
FIGURE 15. INTENSITY DISTRIBUTION OF TWO ISOFORMS OF PFN2.	65
FIGURE 16. IDENTIFICATION OF PROTEIN SPECIES OF VIMENTIN (VIM), 2DE EXPERIMENT.	68
FIGURE 17. IDENTIFICATION OF PROTEIN SPECIES OF VIMENTIN (VIM), 1DE EXPERIMENT.	69
FIGURE 18. IDENTIFICATION OF PROTEIN SPECIES OF PLASTIN 3 (PLS3), 2DE EXPERIMENT.....	70
FIGURE 19. IDENTIFICATION OF PROTEIN SPECIES OF PLASTIN 3 (PLS3), 1DE EXPERIMENT.....	71
FIGURE 20. IDENTIFICATION OF PROTEIN SPECIES OF T-COMPLEX PROTEIN 1 SUBUNIT ALPHA (TCP1), 2DE EXPERIMENT.....	72
FIGURE 21. IDENTIFICATION OF PROTEIN SPECIES OF T-COMPLEX PROTEIN 1 SUBUNIT ALPHA (TCP1), 1DE EXPERIMENT.....	73
FIGURE 22. IDENTIFICATION OF PROTEIN SPECIES OF LAMIN-B1 (LMNB1), 2DE EXPERIMENT.	74
FIGURE 23. IDENTIFICATION OF PROTEIN SPECIES OF LAMIN-B1 (LMNB1), 2DE EXPERIMENT.	75
FIGURE 24. IDENTIFICATION OF PROTEIN SPECIES OF LAMIN A/C (LMNA), 2DE EXPERIMENT.....	76
FIGURE 25. IDENTIFICATION OF PROTEIN SPECIES OF LAMIN A/C (LMNA), 1DE EXPERIMENT.....	77
FIGURE 26. IDENTIFICATION OF PROTEIN SPECIES OF LAMIN A/C (LMNA), 2DE EXPERIMENT.....	78
FIGURE 27. IDENTIFICATION OF PROTEIN SPECIES OF LAMIN A/C (LMNA), 1DE EXPERIMENT.....	79
FIGURE 28. IDENTIFICATION OF PROTEIN SPECIES OF F-ACTIN-CAPPING PROTEIN SUBUNIT BETA (CAPZB), 2DE EXPERIMENT.....	80

FIGURE 29. IDENTIFICATION OF PROTEIN SPECIES OF F-ACTIN-CAPPING PROTEIN SUBUNIT BETA (CAPZB), 1DE EXPERIMENT.....	81
FIGURE 30. IDENTIFICATION OF PROTEIN SPECIES OF UPF0568 PROTEIN C14ORF166 (C14ORF166), 2DE EXPERIMENT.....	82
FIGURE 31. IDENTIFICATION OF PROTEIN SPECIES OF UPF0568 PROTEIN C14ORF166 (C14ORF166), 1DE EXPERIMENT.....	83
FIGURE 32. SEPARATION EFFICIENCY OF PIEF AND 1D PAGE.....	87
FIGURE 33. PROTEOME PROFILING BY PAGE-PIEF-LC-MS/MS.....	88
FIGURE 34. ORTHOGONALITY PLOTS OF RP-HPLC RETENTION TIME AND IPG STRIP SLICE.....	90
FIGURE 35. COMPARISON OF NUMBERS OF IDENTIFIED PROTEINS USING SHORT LC GRADIENT, LONG LC GRADIENT AND PIEF.	91
FIGURE 36. COMPARISON OF PROTEOME COVERAGE USING GELC-MS/MS, PIEF-LC-MS/MS AND PAGE- PIEF-LC-MS/MS.	92
FIGURE 37. PROTEINS IDENTIFIED FOR EACH PAGE SLICE FROM THE GELC-MS/MS AND PAGE-PIEF-LC- MS/MS APPROACHES.....	94
FIGURE 38. REPRODUCIBILITY OF PIEF.	95
FIGURE 39. REPRODUCIBILITY OF THE DISTRIBUTION OF IDENTIFIED PEPTIDES ACROSS THE IPG STRIP.....	97
FIGURE 40. QUANTIFICATION OF PROTEINS FROM THE CPC CULTURE MEDIUM.	101
FIGURE 41. ANALYSIS OF THE DISTRIBUTION OF PEPTIDES ACROSS AN 18CM IPG STRIP.	106
FIGURE 42. REPRODUCIBILITY OF THE STANDARD AND OPTIMIZED 18CM PIEF FRACTIONATION SCHEME. .	109

List of tables

TABLE 1. PROTEINS SHOWING HIGHER EXPRESSION IN CPC-2D COMPARED TO CPC-3D, SIGNIFICANT IN BOTH BIOLOGICAL REPLICATES	37
TABLE 2. GENE ONTOLOGY ENRICHMENT ANALYSIS OF SIGNIFICANT PROTEINS WITH HIGHER EXPRESSION IN CPC-2D COMPARED TO CPC-3D.	38
TABLE 3. PROTEINS SHOWING HIGHER EXPRESSION IN CPC-3D COMPARED TO CPC-2D, SIGNIFICANT IN BOTH BIOLOGICAL REPLICATES	39
TABLE 4. FUNCTIONAL ENRICHMENT ANALYSIS OF PROTEINS UPREGULATED IN CPC-3D RELATIVE TO CPC-2D.....	42
TABLE 5. PROPERTIES OF PROTEINS IDENTIFIED ONLY WITH 1DE AND NOT 2DE, HAVING GRAVY SCORE LARGER THAN 0.25.	55
TABLE 6. NUMBER OF IDENTIFIED PROTEINS, IDENTIFIED ISOFORMS AND NUMBER OF PEPTIDES PER ISOFORM.....	59
TABLE 7. NUMBERS OF IDENTIFIED PEPTIDES AND PROTEINS IN PIEF REPLICATES.....	95
TABLE 8. REPRODUCIBILITY OF GELC-MS/MS AND PAGE-PIEF-LC-MS/MS.....	97
TABLE 9. LIST OF SECRETED PROTEINS IDENTIFIED FROM CPC CELL CULTURE MEDIUM AND NOT IN ANY OF THE WHOLE CELL BASED ANALYSES IN THIS STUDY.....	103
TABLE 10. OPTIMIZATION OF THE FRACTIONATION SCHEME FOR 18 CM PIEF.	108

List of abbreviations

1D	one dimensional
1DE	1D PAGE; one dimensional electrophoresis
2D	two dimensional
2DE	two dimensional electrophoresis
3D	three-dimensional
ACN	acetonitrile
AGC	automatic gain control
CID	collision induced dissociation
CPC	chondrogenic progenitor cells
CPC-2D	chondrogenic progenitor cells cultured in plastic dishes
CPC-3D	chondrogenic progenitor cells cultured in alginate beads
Da	dalton
ECM	extracellular matrix
EO	endochondral ossification
ESI	electrospray ionization
FA	focal adhesion
FAK	focal adhesion kinase
FBS	fetal bovine serum
GO	gene ontology
GeLC-MS/MS	protein analysis by one-dimensional electrophoresis followed by liquid chromatography and tandem mass spectrometry
GRAVY	grand average of hydropathy
H/L	heavy/light
HPLC	high performance liquid chromatography
ID	identifier
IEF	isoelectric focusing
IPG	immobilized pH gradient
IPI	international protein index
LC	liquid chromatography
LC-MS/MS	liquid chromatography – tandem mass spectrometry
LIT	linear ion trap

mRNA	messenger ribonucleic acid
MS/MS	tandem mass spectrometry
MSCs	mesenchymal stem cells
MW	molecular weight
NE	nuclear extract
OA	osteoarthritis
PAGE	polyacrylamide gel electrophoresis
PCM	pericellular matrix
PEP	posterior error probability
pI	isoelectric point
pIEF	peptide isoelectric focusing
QIT	quadrupole ion trap
PAGE	polyacrylamide gel electrophoresis
RF	radio frequency
RNA	ribonucleic acid
RP	reversed phase
RP-HPLC	reversed phase liquid chromatography
RT	retention time
SDS	sodium dodecyl sulfate

Abstract

Modern day mass spectrometry (MS) has evolved into a powerful analytical technique, central to the field of proteomics. Currently, the identification and quantification of proteins from different biological samples can be performed routinely and on a large-scale. Despite the availability of fast and accurate mass spectrometers, deciphering the complex proteome of an organism necessitates the use of various separation techniques prior to the mass spectrometric analysis. Each separation technique offers a set of advantages and disadvantages and has its own power and limitations. Evaluation of the performance of the different separation approaches prior to the mass spectrometric analysis is a crucial step in improving the analytical power of the technique.

The main aim of this study is to perform a proteomic characterization of chondrogenic progenitor cells (CPCs) using mass spectrometry and multidimensional separation approaches. CPCs are progenitor cells of mesenchymal origin, characterized by high chondrogenic potential, that migrate into the degenerating cartilage in late stages of osteoarthritis (Koelling, Kruegel et al. 2009). The mobilization of these cells from the bone marrow can be seen as a regenerative attempt of the cartilage, a tissue with very low reparative capacity. While the migratory and differentiation potential of CPCs has been studied in detail, additional information regarding the phenotype and the cellular fate of CPCs is needed. Results from this work show that CPCs upregulate proteins associated with the pericellular matrix and downregulate mesenchymal stem cells markers, when grown in a three dimensional cell culture, indicating ongoing chondrogenesis. In addition, this study reports multiple examples of the identification and quantification of not only single protein species but of a number of variants of the same protein which have different abundance in differentiating and undifferentiated CPCs. Moreover, a number of secreted signaling factors were discovered. The characterization of these different aspects of the protein expression profile of CPCs was done using different multidimensional separation methods. While the analysis of chondrogenic differentiation of CPCs could be performed using single dimension separation at the protein and peptide level, the identification of multiple protein species and secreted signaling molecules demanded the application of methods with higher separation power such as two-dimensional electrophoresis and a novel three-dimensional separation approach combining protein separation by molecular weight and peptide separation by isoelectric focusing and reversed phase liquid chromatography prior to the mass spectrometric analysis.

Overall, this work provides a rich body of information about the protein expression profile of CPCs and its changes during chondrogenesis in addition to the application and critical evaluation of several multidimensional separation approaches for proteomic analysis.

1 Introduction

1.1 Proteomics

Proteins are essential components of the cell and the multicellular organism. They participate in a multitude of functions such as catalysis, transport, cell communication and signaling. Proteins are also the structural components of the cell (cytoskeleton), hormones, antibodies and antigens, and receptors. In addition to these diverse functions, proteins control the phenotype of a cell, which is essential in the context of the multicellular organism. Multicellular organization is based on the generation and coexistence of a variety of cell types (organized into tissues) that share common embryonic progenitors. With few exceptions, all different cell types share the same DNA content while, at the same time, cells can have large phenotypic differences. Proteins contribute to these differences due to their diverse chemical nature and through differences in their abundance and structure. The process of generating protein diversity from a static DNA sequence (gene expression) is based on the cellular control over a number of biological processes such as transcription, RNA splicing, translation and post-translational modification. First, the protein abundance is affected by the rate of mRNA transcription, mRNA translation and the mRNA and protein half-life. It has been shown that the rates of mRNA transcription and translation have major influence on protein abundance while mRNA stability and protein degradation have lesser effects (Schwanhausser, Busse et al. 2011). Second, protein diversity can be generated as a result of alternative splicing (Stamm, Ben-Ari et al. 2005). This process in which specific exons of a gene are included or excluded from the final mRNA, results in the production of different mRNA and protein species that are coded by the same gene. Moreover, alternative splicing can alter the function of the protein product as the inclusion/exclusion of exons can remove or add protein domains or regions that are essential for the protein's function and interactions (Buljan, Chalancon et al. 2012). Third, proteins are also a product of a wide variety of post-translational modifications (Mann and Jensen 2003). The chemical nature of proteins is such that even small modifications, such as single protein phosphorylation, acetylation, or methylation, can have a major impact on the protein function and interactions. Most of these modifications are reversible and their dynamic control (addition or removal) allows the cells to perform cellular signaling (e.g. through phosphorylation events connecting signaling at the plasma membrane with dynamics of transcription factors in the nucleus) (Miyazono 2000) or enforce epigenetic changes (e.g. changes in gene expression due to histone acetylation) (Kouzarides 2007).

Clearly, proteins, their abundance, and modification state have to be studied systematically and on a large-scale in order to gain understanding of protein function and its relation to the cellular phenotype. The field of proteomics (Zhu, Bilgin et al. 2003) is dedicated to studying the proteome – the set of all expressed proteins in a specific cell type, tissue or organism, at a certain time (Wasinger, Cordwell et al. 1995). Proteomics is a post-genomics discipline that builds up on the wealth of knowledge generated from large-scale DNA sequencing (Mardis 2011). Genomics has allowed defying the basic set of encoded genes (around 23000 for a human) for a large and increasing number of organisms - 3699 sequenced genomes by September 2012 (Genomes Online Database, Joint Genome Institute). Generating such wealth of data has been possible due to a constant increase in the power of sequencing techniques and a decreasing price (Metzker 2010). Nevertheless, the set of genes encoded in an organism cannot by itself explain multicellular phenotypes. This can only be achieved by acquiring and integrating knowledge from other cellular components, such as the proteome.

While studying the proteome of the cell is an exciting opportunity, this task faces a number of methodological challenges. One problem is the very wide protein expression range. Proteome wide expression levels can differ more than 100 000 fold with very abundant cellular proteins like ribosomal and proteasomal proteins are being co-expressed with very low abundant proteins like protein kinases and transcription factors (Nagaraj, Wisniewski et al. 2011). Moreover, the expressed cellular proteins have very different physico-chemical properties such as size, isoelectric point and hydrophobicity. Evidently, an analytical approach dedicated to studying the proteome of a cell would need to overcome those challenges.

1.1.1 Mass spectrometry as an analytical tool in proteomics

Over time, mass spectrometry (MS) has emerged as a fundamental tool in (large-scale) proteomic analysis (Aebersold and Mann 2003). The ability to identify and quantify thousands of proteins driven by development of new technologies and instrumentation (Hu, Noll et al. 2005), application of peptide and protein separation (Mallick and Kuster 2010), introduction to isotope labeling for protein quantification (Gevaert, Impens et al. 2008) and advancements in data analysis (Kumar and Mann 2009). In addition, the development of novel hybrid mass analyzers have allowed for significant improvements in proteomic analysis.

1.2 Mass spectrometry based proteomics

A mass spectrometer consists of three essential parts: the ion source, the mass analyzer, and the detector. The following sections provide examples of sample ionization at the ion source and mass measurement.

1.2.1 Electrospray ionization

Sample ionization has a very important role in the mass spectrometric analysis as neutral molecules have to acquire a charge in order to be separated and detected in the mass analyzer. The ion source is used to generate stable ions of the molecules of interests. In addition, the ion source is the place where molecules are transferred from liquid (or solid) state into gas phase and enter the vacuum of the mass analyzer. The introduction of electrospray ionization (ESI) in mass spectrometric analysis (Yamashita and Fenn 1984) has played an essential role in the development of mass spectrometric analysis as an analytical technique in proteomics. ESI is a soft ionization technique that results in very little fragmentation of the ionized molecules. Due to this, ESI can be used for the analysis of molecules that contain weak bonds such as post-translational modifications as they remain intact during the process of ionization. In fact, because ESI is such a gentle ionization method, it has also been used for the ionization of large non-covalently bound complexes (Heck 2008), including membrane protein complexes (Morgner, Montenegro et al. 2012). For ionization using ESI, the sample of interest is dissolved in a polar, volatile solvent and migrates through the tip of a needle that is maintained at a high (usually positive) potential, relative to the entrance of the mass spectrometer (Fenn, Mann et al. 1989). The high electric potential causes the fluid to form a so called Taylor cone (Taylor 1964) at the tip of the needle. Due to the electric field, small charged droplets are formed at the Taylor cone, which sequentially shrink due to evaporation. In this manner ions are formed at atmospheric pressure and then enter the vacuum region of the mass spectrometer. While ESI is sensitive to the presence of salts, the sample only needs to be dissolved in a mixture of water and organic solvent such as methanol or acetonitrile, including a small percentage of formic or acetic acid, which aids protonation. Due to the possibility for ionization at atmospheric pressure and continuous flow of the solvent, ESI can be successfully interfaced with separation techniques such as liquid chromatography for reduction of the sample complexity (Abian, Oosterkamp et al. 1999).

The introduction of ESI can be seen as a big step up from the pre-existing ionization techniques for protein analysis. Its application has contributed to solving the problem of transferring the analytical sample from a liquid state to a gas state and into the vacuum of

the mass spectrometer without notable fragmentation. The intact molecules can later be fragmented in a predictable manner in order to gain structural information. The development of ESI has contributed to major advancements in sample analysis and therefore has been honored by a Nobel Prize in chemistry (2002) to John Fenn.

1.2.2 Mass measurement

After ionization at the ion source, the ions enter the vacuum of the mass analyzer and are further separated according to their mass to charge ratio (m/z). Upon separation, the ions reach the detector which converts the energy of the incoming particles into a signal that is recorded and transferred to the computer of the mass spectrometer. A number of mass spectrometers exist that are based on different separation principles. The following sections will describe the principles of ion separation of two different types of MS instruments, which can be operated in a hybrid manner.

1.2.2.1 Mass measurement instruments

1.2.2.1.1 Quadrupole and linear ion trap

In the quadrupole ion trap (QIT) mass spectrometer, ions are trapped and stored in an electric field. The mass analyzer is also called Paul trap after Wolfgang Paul, who developed the fundamental theory of the device (Paul and Steinwedel 1953) and for which work shared the Nobel Prize in physics in 1989. In the QIT, a potential well is established within a ring electrode and two end-cap electrodes. Ions from the ion source enter through the end-cap electrode into the trapping field of the ring electrode and become trapped by holding the end-cap electrodes at ground potential and application of radio frequency (RF) potential to the ring electrode. The energy of the ions confined to the center of the trap is lowered using collision with a damping gas. The use of a damping gas, commonly helium, has been shown to be beneficial for the performance of the instrument by improving the mass resolution, sensitivity and detection limit (Stafford Jr, Kelley et al. 1984). The helium gas can further be used to induce fragmentation and obtain structural information. After trapping the ions and damping their energy, the mass spectrum (a diagram of the intensity of the m/z values of the ions) is recorded. The m/z measurement is based on increasing the RF amplitude (low amplitude is used to trap all ions above certain m/z). When the RF amplitude is increased, ions with increasing m/z become unstable, leave the ion trap through the exit end-cap and reach the detector as a result of which the m/z values are recorded. This mode of mass measurement using the QIT is called the mass selective instability mode (Stafford Jr, Kelley et al. 1984).

Based on the principles of QIT, another ion trap has been developed – the linear ion trap (LIT). The LIT operates by trapping ions radially by a two-dimensional RF field, and axially by stopping potentials applied to end electrodes (Schwartz, Senko et al. 2002). Compared to the QIT, the LIT has an improved trapping efficiency and increased ion capacity (Schwartz, Senko et al. 2002). Direct comparisons between QIT and LIT have shown that the LIT identifies more peptides and proteins compared with the QIT (Mayya, Rezaul et al. 2005). Overall, LITs are characterized by high scanning speed and high sensitivity. In addition, the LIT allows for several stages of fragmentation of an ion of interest, which can be used to obtain increasing amounts of structural information. The LIT has certain disadvantages – it is not well suited for quantitative analysis of complex samples. One reason is that the number of ions that can be stored simultaneously in an ion trap is limited (Wong and Graham Cooks 1997). In the case of complex biological samples, many different ions are stored in the ion trap at the same time and the limit of detection can be severely reduced by the interference of ions with high abundance. Furthermore the low accuracy of the instrument (Wong and Graham Cooks 1997) can lead to the incorrect determination of the charge state of the ion and inference of a wrong m/z (Colinge, Magnin et al. 2003).

1.2.2.1.2 Orbitrap

The orbitrap is the newest invented mass analyzer (Makarov 2000). It is composed of an outer barrel-like electrode and a central spindle like central electrode along the axis of the instrument. Ions are trapped (oscillate) in stable trajectories around the central electrode. This is possible due to a balance between an electrostatic attraction between the oscillating ions and the central electrode on one side, and the centrifugal force of the oscillating ions on the other (Hu, Noll et al. 2005). The frequency of axial oscillations (oscillations along the central electrode) is independent of the initial velocities and the coordinates of the ions. For this reasons, this frequency can be used to determine the m/z of the ions. This is done using the following formula: $w = \sqrt{k / (m/z)}$ where w is the oscillating frequency and k is an instrument constant. Axial oscillations are directly detected from the image current on the outer electrodes. This is done by recording the time-domain signal and transforming it to a m/z spectrum using Fourier transformation (Senko, Canterbury et al. 1996). Since the frequencies of these oscillations can be measured very precisely, very high accuracy of measurement of m/z can be achieved. The accuracy of m/z measurement and the speed of the instrument are related in a manner that higher accuracy of measurement needs higher measurement time leading to lower speed. In addition, the orbitrap has a high dynamic

range over which the m/z can be measured with high precision (Makarov, Denisov et al. 2006).

1.2.2.1.3 Hybrid Linear Ion Trap – orbitrap

Both mass analyzers (LIT and orbitrap) can be operated together in a hybrid manner (Makarov, Denisov et al. 2006). In such an instrument the measurement of the peptide/protein precursor m/z is performed in the orbitrap analyzer with very high accuracy of the measurement, while peptide fragmentation is performed in the LIT in order to gain structural information (Scigelova and Makarov 2006). The LIT performs fast and sensitive analysis of the resulting fragments, which can later be used to unambiguously identify the protein/peptide. This mode of m/z analysis of the analyte (precursor) and the products of its fragmentation is called tandem mass analysis (MS/MS). In 2009, an improved version of the hybrid orbitrap-LIT instrument was introduced (Olsen, Schwartz et al. 2009).

1.2.3 Protein and peptide fragmentation in the mass spectrometer

Peptide fragmentation (sequencing) by MS/MS is not straightforward because various covalent bonds can get broken during the process. The breaking bonds can be either backbone linkages or side chain groups. In order to precisely describe these events, a special nomenclature is used. A common convention for description of peptide fragments formed in the mass spectrometer has been proposed (Roepstorff and Fohlman 1984, Steen and Mann 2004). Ions derived from the N-terminus of the original peptide are termed a, b, or c (the charge is retained on the N-terminus), while those originating from the C-terminus are named x, y, or z.

1.2.3.1 Fragmentation by collision induced dissociation

One common method for peptide fragmentation of the is using collision induced/activated dissociation (CID/CAD) (McLuckey 1992). Using this technique the precursor ions are accelerated to some kinetic energy and then collide with gas molecules of nitrogen, argon, or helium, and fragment. In the collision, a part of the kinetic energy of the ion is converted into vibrational/rotational energy. If the internal gained energy is high enough, the precursor ion will fragment and the fragment ions will be observed in the mass spectrometer. Depending on the type of mass analyzer, either high-energy CID (kiloelectron volt collision energy) or low-energy CID (<100 eV) is performed (Sleno and Volmer 2004). Low-energy CID is characteristic to ion trap instruments (March 1997). In low-energy CID, only a small amount of energy is added to the internal energy of the precursor ion in each collision and

energy from hundreds of collisions will have to be accumulated in order to induce fragmentation (Westman-Brinkmalm and Brinkmalm 2008). In the case of peptide fragmentation, low-energy CID often causes the amide bond of the backbone to dissociate, thereby producing b- and y-type of fragment ions (Biemann 1988).

1.2.4 Protein and peptide identification by mass spectrometry

The identity of a protein in a complex mixture can be revealed from mass measurement of intact proteins and structural information derived from its fragmentation (Meng, Cargile et al. 2002). Such measurements can be very advantageous for the identification of post-translational modifications and protein isoforms (Siuti and Kelleher 2007). Mass spectrometric analysis of whole proteins has been performed on a large-scale leading to the identification of more than 1000 proteins using multidimensional separation (Tran, Zamdborg et al. 2011). Nevertheless, there are a number of limitations associated with this type of analysis which include protein fractionation, ionization and fragmentation in the gas phase. Alternatively, proteins can be identified based on the peptides they generate after hydrolysis by a specific protease. The resulting peptides can be fractionated and fragmented more easily. This mode of analysis called “bottom up” or “shotgun” is widely applied in the field of proteomics (Yates 2004).

The digestion of a protein with a specific protease generates a unique peptide pattern, which can be employed for protein identification. For example, in his seminal work on sickle cell anemia, Vernon M. Ingram compared the trypsin generated “finger-prints” of normal and sickle cell hemoglobin using two-dimensional (2D) peptide separation and discovered that they differed, eventually pointing to the disease causing mutation (Ingram 1957). In modern proteomic analysis the 2D peptide maps have been substituted by detection with a mass spectrometer.

The peptide fingerprint generated by the hydrolysis with a specific protease can be used for protein identification, because it is a reasonably unique characteristic for a protein, even among a large collection of protein sequences such as a sequence database (Henzel, Billeci et al. 1993, Yates, Speicher et al. 1993). The protease of choice is usually trypsin because this protease has a number of characteristics, which make it useful for large-scale proteomic experiments. Trypsin cleaves carboxy-terminal (C-terminal) to Arg and Lys (Olsen, Ong et al. 2004) and this cleavage pattern results in peptides with average length of 9 amino acids (Vandermarliere, Mueller et al. 2013). Such length is very suitable for analysis by mass spectrometry while the presence of C-terminal basic residue supports peptide protonation and subsequent fragmentation (Vandermarliere, Mueller et al. 2013). In addition, the digestion of a complex protein mixture reduces the different physico-chemical

properties of the present proteins into a set of peptides with more uniform population. This facilitates the analysis of very different proteins present in the same sample.

While peptide maps can be used to infer the presence of a protein in a sample, the protein identification process can become impossible when the sample is a complex protein mixture. In this case, more information is needed in order to confidently identify most of the proteins present in the sample. As described, MS/MS analysis produces information specific to the sequence of a peptide and is a general solution to the problem of protein mixtures. In this manner all proteins present in a sample are digested, the resulting peptides are analyzed by MS/MS, and their fragmentation pattern is used to derive the peptide amino acid sequence and finally, the identity of the proteins. This sequencing approach is based on the predictable nature of the fragmentation of peptides using CID. In addition, by determining the mass of a peptide with high accuracy (e.g. in the orbitrap), the possible amino acid composition is narrowed, facilitating identification (Eng, Searle et al. 2011).

Large-scale peptide and protein identification is carried out using a search engine (Marcotte 2007). The engine predicts the fragmentation pattern of the peptide, based on the established nomenclature, and the instrument/fragmentation technique used in the experiment and calculates the masses (m/z) of the expected peptide fragments. In the next step the similarity of the peptide fragment masses and the masses calculated by the software is scored. Finally, the identity of the proteins is inferred from the sets of identified peptides (Baldwin 2004). This step (protein inference) is rather crucial since peptides can be either uniquely assigned to a single protein or shared by more than one protein species (Nesvizhskii and Aebersold 2005). In order to deal with this, proteins are usually grouped together based on the set of identified peptides, while the size of these groups depends on the similarity of the proteins and the number of identified peptides.

1.3 Quantitative mass spectrometry based proteomics

In addition to identification of proteins in complex biological samples, mass spectrometry provides the opportunity to quantify the protein abundance across samples. This can be done by comparing the intensity of the MS signal or the numbers of identified peptides across experiments, or by the introduction of stable isotope labeling (Bantscheff, Schirle et al. 2007). One of the most powerful methods for quantitative analysis is the introduction of stable isotope labeling by amino acids in cell culture (SILAC) (Ong, Blagoev et al. 2002). In this method, cells are grown using cell culture medium containing Lys and Arg which contain the heavier ^{13}C , ^{15}N , ^2D isotopes. The introduction of the amino acids into peptides generates predicted mass shifts of +4, +6 or +8 Da for Lys and +6 or +10 for Arg.

Combined with the use of trypsin as a protease of choice and the ability to pool samples very early in the sample preparation, this techniques is a powerful method for quantitative proteomics (Mann 2006).

1.4 Protein and peptide separation in proteomics

While in certain cases peptide/protein identification can be achieved without sample separation prior to the mass spectrometric analysis (Wilm, Neubauer et al. 1996), usually some degree of separation is performed. There are many advantages to combining MS and separation methods, including increased sensitivity, dynamic range, and selectivity (Wu and Han 2006). Various separation methods can be used to fractionate the sample offline or online before they are introduced into the mass spectrometer. Separation using one or a combination of different methods is essential for complex mixtures as shotgun analyses usually result in a very high complexity of the sample. Besides generating a large number of peptides, there are additional factors that complicate the identification of all proteins in a complex mixture. These include the dynamic range of the proteins (and the protease generated peptides), protease missed cleavages and chemical modifications induced during sample preparation (Lubec and Afjehi-Sadat 2007). In order to overcome these issues, one needs to use separation methods that can accommodate the sample complexity. Overall, separation techniques can be applied on peptide or protein level.

1.4.1 Peptide separation

Large-scale protein identification is largely based on shotgun proteomic analysis (Yates 2013). In this mode of analysis the protein sample is translated into higher complexity at the peptide level as proteins generate multiple peptides after protease digestion. Complex peptide mixtures can be separated based on different peptide characteristics, such as polarity (hydrophobicity), charge and isoelectric point (pI) (Di Palma, Hennrich et al. 2012).

1.4.1.1 Reversed-phase liquid chromatography

The most common technique for peptide separation prior to MS/MS is reversed-phase liquid chromatography (RP-HPLC) (Ducret, Van Oostveen et al. 1998, Shen, Zhao et al. 2002). This technique relies on the hydrophobic interactions between the analyte (peptide mixture) and stationary hydrophobic surface. A liquid mobile phase, which is more polar than the stationary phase and is usually comprised of water and an organic solvent like acetonitrile, continuously competes with the peptides for hydrophobic interactions. The

stationary phase is usually made of capillaries packed with silica material to which long chain alkyl groups (usually n-octadecyl) are covalent attached (Hsieh and Jorgenson 1996). By increasing the percentage of the organic solvent in the mobile phase, the interaction of more hydrophobic peptides with the stationary phase is disturbed and they migrate through the capillary. By using a gradient of organic solvent the separation of peptides with different hydrophobicity takes place (Lee, Yi et al. 2004). RP-HPLC has a number of advantages that have made it the final analytical chromatography step before MS analysis and a standard in peptide separation in shotgun proteomics. One of the main advantages is that it is seamlessly interfaced with ESI - the mobile phase used for peptide separation is the same as the solution used for ionization. In addition, the tip of the capillary turns into an ion source when potential is applied between the tip and the entrance of the mass spectrometer. RP-HPLC can also be very robust, reliable and reproducible, which allows its use in large-scale proteomic studies. In addition, RP-HPLC can attain very high separation resolution (Shen and Smith 2002). Separation can further be enhanced by using longer organic solvent gradients (longer analysis time), increased length of the packed capillary or decrease in the size of the packing material (Thakur, Geiger et al. 2011).

1.4.1.2 Peptide isoelectric focusing

While peptide separation prior to MS/MS by RP-HPLC (further abbreviated as LC-MS/MS) is widespread, it can rarely overcome the high complexity generated using shotgun proteomics. In order to increase the separation power, other separation steps can be performed. These separation steps can be interfaced either online or offline. One example of efficient first dimension separation (prior to LC-MS/MS) is peptide isoelectric focusing (pIEF) (Cargile, Talley et al. 2004). In this technique peptide separation is usually carried in immobilized pH gradients (IPGs) - acrylamide gels that are made by copolymerization of acrylamide with acrylamide derivatives containing carboxyl or tertiary amine groups, which act as buffers (Strahler and Hanash 1991). By mixing the acrylamide and its derivatives in defined ratios, a gel containing a pH gradient is created. When a potential is applied at the ends of the gel, peptides start to migrate because of the resulting current until they reach a pH region where their charge is zero, i.e. their pI (Rabilloud, Vaezzadeh et al. 2009). The pI of a peptide/protein is based on the dissociation constants of the ionizable side chains of one of the following amino acids: Arg, Lys, Asp, Glu, His, Tyr and Cys and on the dissociation constants of the peptide carboxy group at the C-terminus and the amino group at the N-terminus (Sillero and Ribeiro 1989). Overall, pIEF can achieve focusing of the peptides in very sharp regions of the pH gradient, resulting in very high resolution of separation (Branca, Orre et al. 2014). Depending on the mode of operation, peptides can

be recovered from the gel either by cutting it into pieces following extraction (Cargile, Talley et al. 2004), or from the liquid phase of an off-gel isoelectric focusing apparatus (Michel, Reymond et al. 2003).

1.4.2 Protein separation

Besides separation of peptides prior to LC-MS/MS analysis, some degree separation can be performed at the protein level prior to protease digestion. Separations can be based on one of the protein properties like molecular weight (MW), pI and hydrophobicity.

1.4.2.1 One-dimensional polyacrylamide gel electrophoresis

One very advantageous technique for reducing sample complexity is protein separation using one-dimensional (1D) polyacrylamide gel electrophoresis (PAGE) (Laemmli 1970). After electrophoresis, selected protein bands or the whole lane is cut into a number of fractions which then can be in-gel digested using trypsin (Shevchenko, Tomas et al. 2006) and analyzed using LC-MS/MS. In this method, commonly known as GeLC-MS/MS (Schirle, Heurtier et al. 2003), proteins are efficiently separated and mostly identified in a single fraction. Large-scale protein analysis using GeLC-MS/MS has been applied to different samples, e.g. murine embryonic stem cells (Graumann, Hubner et al. 2008) and cerebrospinal fluid (Zougman, Pilch et al. 2008). One advantage of GeLC-MS/MS analysis is that large proteins, which generate a high number of peptides after protease digestion, are separated from smaller proteins, which generate fewer peptides. In addition, this method is also applicable to cases in which proteins of specific MW range are of interest.

1.4.2.2 Two-dimensional electrophoresis

Two-dimensional gel electrophoresis (2DE) is a powerful technique that resolves complex protein mixtures by employing protein separation by pI in the first dimension and by MW in the second dimension (O'Farrell 1975, Görg, Weiss et al. 2004). Protein separation in the first dimension is carried out as described for pIEF. One of the main advantages of this technique is that a large number of protein species can be resolved and visualized within the same gel (Reed, Densmore et al. 2012). It is not uncommon to see trains of spots that correspond to the same protein (Halligan 2009). Such patterns can be generated from post-translational modifications such as phosphorylation, acetylation, and deamidation as these modifications influence the pI of the proteins and their migration pattern within the two dimension of 2DE (Girardet, Miclo et al. 2006). In addition, different glycosylation patterns or proteolytic processing by cellular proteases can generate separate protein species which

can be visualized on a 2DE gel but have to be identified by MS/MS in order to confirm the identity of the protein (Barrabes, Sarrats et al. 2010). The separation resolution of the technique can be enhanced by employing narrow pH ranges in the first dimension, so called “zoom-gels” (Hoving, Gerrits et al. 2002). Furthermore if a sample is split and analyzed using overlapping narrow pH ranges a very large number of spots can be observed (Hoving, Voshol et al. 2000). Although such high number of protein spots is very impressive, it is clear that not all spots correspond to individual proteins, thus the achievable depth of proteomic analysis is unclear. So far no study has evaluated how many proteins are identifiable from a whole 2DE gel.

1.5 Cellular differentiation

As described in the previous sections, large-scale proteomic experiments are possible due to the availability of fast and accurate mass spectrometers and a variety of fractionation methods for the reduction sample complexity prior to LC-MS/MS analysis. One exciting field of application of mass spectrometry based proteomics is in the field of cellular differentiation (van Hoof, Krijgsveld et al. 2012). The process of cellular differentiation can be described as a sequence of events that lead to the commitment and specialization of stem cells and cellular progenitors to a specific cell fate (Keller 2005). Cellular differentiation is usually accompanied by major changes in cellular gene expression and phenotype (Tosh and Slack 2002). The process involves cell signaling events that are initiated at the plasma membrane and finally reach the nucleus in order to rewrite the epigenetic landscape and turn on/off the activity of transcription factors, finally resulting in altered gene expression (Pawson 1993). One of the most well studied aspects of cellular differentiation is the action of secreted molecules (factors) which upon binding to receptors at the plasma membrane of the target cells induce cell signaling, differentiation and changes in gene expression. Examples of such mechanisms are the signaling cascades initiated by the binding of Transforming growth factor beta (TGF- β) (Shi and Massagué) and Wnt (Kim, Kim et al. 2013) to their receptors at the plasma membrane. Parallel to secreted factors, cells also respond to the physical nature of their environment by changes in cell shape. This response can also regulate cellular differentiation (Watt, Jordan et al. 1988, McBeath, Pirone et al. 2004).

1.6 Cellular environment, shape and differentiation

The cellular environment has an essential role in determining the cell fate in a multicellular organism (Huang and Ingber 1999, Mammoto, Mammoto et al. 2013). One example for the

impact of the cellular environment comes from studies on the differences between two-dimensional (2D) and three-dimensional (3D) *in vitro* cell cultures. In this respect, studies have shown differences in cellular adhesion (Cukierman, Pankov et al. 2001) and cellular gene expression (Kumar, Zhong et al. 2008, Konze, van Diepen et al. 2014) as part of large overall functional differences between the 2D and 3D cell culture systems (Pampaloni, Reynaud et al. 2007, Baker and Chen 2012).

For most studies 2D plastic dishes are a standard and thus have been used extensively for cell culture. Most often, polystyrene is the material from which the plastics of the dishes are molded (Ryan 2008). Since polystyrene is very hydrophobic and non-adherent, most cells do not attach/spread easily on the surface and for this reason the polystyrene surface is treated, e.g. with corona discharge, leading to the generation of hydroxyl groups on the polystyrene surface (Curtis, Forrester et al. 1983). Treated polystyrene is much more hydrophilic and readily binds adhesive molecules, such as fibronectin, vitronectin and collagens to which cells can attach (Grinnell and Feld 1982). Overall, the standard 2D culture provides flat but very adhesive surface on which cells usually adopt a spread flat shape (Mather and Roberts 1998). This is in contrast to suspension/3D culture conditions, in which case the cells would adopt a round cell shape (Mather and Roberts 1998).

Importantly, numerous studies have shown that (changes in) the cell shape can have a major impact on the cellular phenotype including cell growth (Folkman and Moscona 1978) and differentiation (Benya and Shaffer 1982, Watt, Jordan et al. 1988, McBeath, Pirone et al. 2004).

The main reason for the culture dependent changes in the cellular shape is that cells can feel and respond to the tension of the extracellular space (Discher, Janmey et al. 2005). There are three components of this response: 1) the cell has an internal tension generating mechanism; 2) there are receptors for adhesive extracellular proteins, located at the plasma membrane; and 3) a complex cellular machinery exists which connects the internal tension generator with the receptors at the plasma membrane.

The most well studied component of the tension generating mechanism is the force generated from actomyosin (Shin, Swift et al. 2013). This protein complex is made of actin, a main filamentous component of the cytoskeleton (Janmey 1998) and the tension is generated by the myosin motor proteins (the second component) that move across, crosslink and pull actin using ATP hydrolysis (Pollard and Korn 1973). Connected to this tension generating system are cellular receptors of adhesive extracellular proteins, located at the plasma membrane. The most important class of these receptors are the integrins (Hynes 2002). Integrins exist as two noncovalently bound α and β subunits. The integrin family is composed of 18 α and 8 β subunits that form heterodimers. Different heterodimers bind different adhesion molecules, for example the integrin heterodimers $\alpha3 \beta1$ and $\alpha6 \beta1$

bind laminin, while $\alpha 1 \beta 1$ and $\alpha 2 \beta 1$ bind collagen (Hynes 2002). One distinct feature of integrins is that their cytoplasmic domains are short (10-70 amino acids) which defines the lack of catalytic activity of the receptors. Integrin signaling involves binding of the receptor ligand, induction of conformational changes and clustering of the receptors (Giancotti 1999). Sustained integrin signaling results in the formation of focal adhesion (FA) complexes at the plasma membrane and the formation of actin stress fibers (Wozniak, Modzelewska et al. 2004). The FA is a large, dynamic, multi-protein complex composed of more than 150 proteins (Zaidel-Bar, Itzkovitz et al. 2007) that links the extracellular matrix with the actin cytoskeleton and acts as a cell signaling hub (Burrige and Chrzanowska-Wodnicka 1996). Two prominent examples of FA proteins, which link integrins with the actin cytoskeleton, are talin and vinculin (Humphries, Wang et al. 2007). In addition to connecting proteins, a number of signaling proteins (such as kinases, phosphatases and GTPases) are also associated with the FA (Zaidel-Bar, Itzkovitz et al. 2007). These include the integrin-linked kinase (Li, Zhang et al. 1999) and the tyrosine phosphatase PTEN (Tamura, Gu et al. 1998). A prominent member of the FA associated GTPases that can transduce changes in the tension of the cellular environment into changes in gene expression and influence a number of cellular processes such as cell proliferation and differentiation is RhoA (Burrige and Wennerberg 2004).

1.7 Chondrogenesis

One example for cellular differentiation which is guided by environmental queues and accompanied by large changes in the cellular phenotype is chondrogenesis, the process of differentiation of mesenchymal stem cells (MSCs) into chondrocytes. *In vivo*, this multistage process is initiated by condensation of MSCs (Hall and Miyake 2000) and accompanied by changes in cell shape (Fell 1925, Solursh, Linsenmayer et al. 1982) and in the network of transcription factors governing the differentiation to chondrocytes (de Crombrughe, Lefebvre et al. 2000). The differentiated chondrocytes remain as resting cells to form the articular cartilage or undergo proliferation, terminal differentiation to chondrocyte hypertrophy, and apoptosis in order to form transient cartilage, the template for future bones (Goldring 2012).

1.8 *In vivo* chondrogenesis

The process of chondrogenesis of MSCs is essential for bone formation and includes the production of specialized extracellular matrix (ECM), the cartilage which is replaced by bone, a process called endochondral ossification (EO) (Mackie, Ahmed et al. 2008). The

process of cartilage formation and EO is initiated by the condensation of MSCs. The cells aggregate to a high cell density and form cell-cell interactions. Specific cell-cell adhesion molecules such as N-cadherin and N-CAM are associated with this cellular stage (Oberlender and Tuan 1994). MSCs divide in the center of the condensations and turn off the expression of mesenchymal and condensation markers (Wezeman 1998). Next, the cells undergo chondrogenesis and change their shape from elongated to a characteristic spherical cell shape (Woods, Wang et al. 2007). The newly differentiated chondrogenic cells divide rapidly leading to a large increase in cell numbers and deposition of ECM which later becomes cartilage. At the same time, several characteristic chondrogenic transcription factors are expressed – Sox9 (Bi, Deng et al. 1999), Sox5 and Sox6 (Smits, Li et al. 2001). These transcription factors are characteristic of the chondrocyte phenotype and modulate a change in the ECM protein expression during the chondrogenesis of MSCs. While undifferentiated MSCs express collagen type I and fibronectin (Dessau, von der Mark et al. 1980), and tenascin (Mackie, Thesleff et al. 1987) chondrocytes decrease the expression of these proteins and start to express collagen II (Lefebvre, Huang et al. 1997), collagen type IX and XI (Swiderski and Solursh 1992), and aggrecan (Sekiya, Tsuji et al. 2000). The next step of the process is the maturation of the chondrocytes into hypertrophic chondrocytes. This is characterized by an exit of the cell cycle, increase of the cell size, decrease of collagen type II expression and increase of in the expression of collagen type X and vascular endothelial growth factor (VEGF) (Lefebvre and Smits 2005). The transcription factor RUNX2 is crucial for the induction of chondrocyte hypertrophy (Kim, Otto et al. 1999) and induces the expression of collagen X (Kim, Otto et al. 1999) and metalloproteinase-13 (Jiménez, Balbín et al. 1999). In the final step of EO, hypertrophic chondrocytes differentiate to terminal chondrocytes. This phase is characterized by chondrocyte apoptosis, vascular invasion and finally ossification (calcification) (Colnot 2005). A key angiogenic factor in EO is VEGF (Gerber, Vu et al. 1999), the expression of which is associated with hypertrophic chondrocyte phenotype as explained above. After invasion of blood vessels, osteoblasts also invade the cartilage and deposit bone matrix onto the remnants of the cartilage ECM (Mackie, Ahmed et al. 2008). The invasion of blood vessels and osteoblasts completes the process of chondrogenesis of MSCs, secretion of cartilage ECM and finally transformation of transient cartilage to bone. The chondrocytes that produce this cartilage template are also termed transient since they exhibit the chondrocyte phenotype temporarily, secrete extensively ECM proteins and further proceed to maturation.

1.8.1 Articular cartilage

Besides the transient cartilage described in the previous section, *in vivo* chondrogenesis results in the production of another type of cartilage at the ends of long bones and synovial joints – the articular cartilage (Kuettner 1992). Articular cartilage is composed of an extensive ECM and embedded chondrocytes. Contrary to the fate of the chondrocyte involved in bone formation, chondrocytes from the articular cartilage do not progress to hypertrophy and terminal differentiation (Onyekwelu, Goldring et al. 2009), but instead are dedicated to the turnover of ECM proteins and maintain the structure of the articular cartilage (Muir 1995). At the same time the cells exhibit low metabolic activity and do not divide actively, unless challenged by a pathological condition such as osteoarthritis (Pearle, Warren et al. 2005).

Chondrocytes of the articular cartilage are characterized by their shape and order. In the superficial layer chondrocytes are ordered in parallel to the surface while in the deeper zones, clusters of chondrocytes are found. The tidemark defines a boundary between the cartilage and the underlying bone (Lyons, Stoddart et al. 2005). Since articular cartilage does not proceed to transform to bone through EO, the tissue remains avascular and aneural. The exchange of nutrients and oxygen is through diffusion (Maroudas, Bullough et al. 1968). The nature of the articular cartilage makes it susceptible to degenerative diseases because it is not able to mobilize a regenerative response (Huey, Hu et al. 2012).

1.8.2 *In vitro* chondrogenesis

Chondrogenesis of MSCs can be induced *in vitro*. This can be done when MSCs are forced to aggregate together in the form of a pellet culture (Johnstone, Hering et al. 1998). This culture system allows for cell–cell interactions, which are similar to those induced in the MSC condensation step prior to chondrogenesis and transient cartilage production. In addition to the pellet culture, chondrogenesis of MSCs can also be induced by culturing the cells in suspension. Suitable cultures for chondrogenesis are agarose (Huang, Reuben et al. 2004), alginate (Ma, Hung et al. 2003, Xu, Wang et al. 2008) and hyaluronan hydrogel (Aulin, Bergman et al. 2011). The signaling factor TGF- β is commonly added to culture media in order to promote chondrogenesis and inhibit chondrocyte hypertrophy (Mehlhorn, Schmal et al. 2006). Usually several chondrogenic markers are used to test if chondrogenesis takes place: collagen type II, collagen type X and aggrecan (Mehlhorn, Schmal et al. 2006). Besides the influence of secreted signaling molecules such as TGF- β , another important factor that influences the chondrogenesis of MSCs is the cell shape (Daniels and Solursh 1991). It has been shown that RhoA signaling, through its

downstream effector Rho-associated protein kinase (Leung, Chen et al. 1996) can direct the differentiation of MSCs in a cell shape and soluble secreted factor dependent manner (McBeath, Pirone et al. 2004). It has also been shown that MSCs that are allowed to spread differentiate into osteoblast, while MSCs that remain round differentiate into adipocytes (McBeath, Pirone et al. 2004). In the context of chondrogenesis, it has been shown that RhoA can regulate the activity of the transcription factor Sox9, a master regulator of chondrogenesis (Akiyama, Chaboissier et al. 2002). The function of Sox9 is essential for chondrogenesis as the transcription factor has binding sites in genes coding for a number of ECM proteins, transcription factors and signaling molecules (Oh, Maity et al. 2010). The activity of Sox9 is positively regulated by downregulation or blocking of RhoA (Kumar and Lassar 2009). In addition, overexpression of RhoA can block both Sox9 expression and chondrogenesis (Woods, Wang et al. 2005).

In summary, the culturing of MSCs in a low binding hydrogel like alginate or agarose leads to a change in the cell shape from spread/flat (on 2D plastic) to spherical/round (in suspension), which induces changes in gene expression and promotes chondrogenesis.

1.8.3 Chondrogenic progenitor cells and osteoarthritis

Osteoarthritis (OA) is a degenerative disease of the articular cartilage, characterized by the progressive degradation of the tissue and the inability of the residing chondrocytes to generate repair tissue (Goldring 2012). Another mark of the disease are breaks in the tidemark and deep fissures in the cartilage which otherwise has a smooth surface (Koelling, Kruegel et al. 2009).

A migratory population of progenitor cells has been identified in the late stages of OA (Koelling, Kruegel et al. 2009). These cells were identified to migrate through the tidemark, but were absent in healthy cartilage samples indicating that the cells responded to the pathologic condition of the tissue. The cells were termed chondrogenic progenitor cells (CPCs) due to their ability to differentiate into cells with chondrogenic phenotype when embedded in alginate (Koelling, Kruegel et al. 2009). Besides chondrogenic differentiation, the cells are able to differentiate into two other main MSCs lineages (Pittenger, Mackay et al. 1999) – adipogenic and osteogenic (Koelling, Kruegel et al. 2009), indicating that the cells are of mesenchymal origin. When CPCs are incorporated into alginate they change their shape (relative to 2D culture) from flat to round and upregulate collagen type II (Koelling, Kruegel et al. 2009) indicating ongoing chondrogenesis. Collagen type II expression can further be induced by the addition of TGF- β 3 and BMP-6. Although upregulation of collagen type II indicates chondrogenic differentiation (Kosher, Kulyk et al. 1986), a much more informative approach is needed to characterize in detail the process of

chondrogenesis of CPC and the basis of the phenotypic changes induced by culturing the cells in alginate. Using large-scale mass spectrometry based proteomic investigation of the differences in protein abundance between CPC cells grown on plastic and cells grown in alginate, the process of chondrogenesis of the cells can be characterized in more detail. In addition, since CPCs are progenitor cells from mesenchymal origin, large-scale protein quantification during chondrogenesis can reveal more detail about the nature of the process. This would be beneficial as other studies focused on MSCs differentiation have either focused only on a few markers (Ma, Hung et al. 2003) or have used solely microarray and real time PCR analysis techniques (Xu, Wang et al. 2008).

CPCs have been proposed as a suitable candidate for cell based *in situ* cartilage repair (Muhammad, Schminke et al. 2013). It remains to be shown that CPCs are able to produce a cartilage like ECM with a high resistance to mechanical stress (Koelling, Kruegel et al. 2009). Since the structure and the mechanical properties of the ECM depend on the composition and the interactions of the matrix components (Frantz, Stewart et al. 2010), namely proteins and proteoglycans (Schaefer and Schaefer 2010), an evaluation of the protein content of the ECM produced by CPCs can reveal information about its structure and function.

1.9 Aims of the study

The main aim of this work is to perform a large-scale mass spectrometry based proteomic characterization of CPCs undergoing chondrogenic differentiation and to discover low abundant factors which are secreted by CPCs *in vitro*. Due to the complexity of the cellular proteome this task necessitates the use of different multidimensional separation approaches. In this aspect, this work aims to establish a novel three dimensional separation approach for deep proteome coverage and to evaluate the performance of the widely used two-dimensional electrophoresis. More specifically, this work encompasses five main objectives. The first objective is to characterize the changes in protein expression during chondrogenic differentiation of CPCs. This is based on quantifying proteins between CPCs grown in a two-dimensional plastic culture and in three-dimensional alginate cell culture. The second objective is to evaluate the performance of a large-scale two-dimensional electrophoresis, in which case the whole protein content of the two-dimensional gel is analyzed instead of single protein species, and its application to the protein quantification of CPCs undergoing chondrogenesis. The third objective of this work is to assay the performance of a novel multidimensional separation method based on protein separation by molecular weight in the first dimension followed by isoelectric focusing of peptides in the second dimension prior to mass spectrometric analysis. The

forth objective is to apply the established three-dimensional separation method to the discovery of low abundant signaling factors, secreted by CPCs *in vitro*. The final aim of this work is to provide a straightforward approach for optimization the fraction collection after isoelectric focusing in order to achieve a similar depth of analysis in less time.

2 Materials and Methods

2.1 Chemicals

Acetonitrile, LiChrosolv	Merck, Darmstadt (DE)
Ammonium hydrogen carbonate	Fluka, Buchs (CH)
Cleland's reagent (DTT, for MS analysis)	Calbiochem, Darmstadt (DE)
Coomassie Brilliant Blue G-250	Fluka, Buchs (CH)
Ethanol	Merck, Darmstadt (DE)
Formic acid	Fluka, Buchs (CH)
Iodacetamide (IAA)	Sigma-Aldrich, Steinheim (DE)
Methanol, LiChrosolv	Merck, Darmstadt (DE)
n-hexane	Merck, Darmstadt (DE)
Sodium dodecyl sulfate	Serva, Heidelberg, (DE)
Urea	Merck, Darmstadt (DE)
Water, LiChrosolv	Merck, Darmstadt (DE)

2.2 Laboratory equipment and instruments

LTQ Orbitrap XL	Thermo Fischer Scientific, Bremen
LTQ Orbitrap Velos	Thermo Fischer Scientific, Bremen
SpeedVac Savant SPD121P	Thermo Fischer Scientific, Bremen
Thermomixer Comfort	Eppendorf, Hamburg
Branson Sonifier	Emerson
Fresco microcentrifuge	Thermo Scientific
Heraeus microcentrifuge	Thermo Scientific
IPGphor	Pharmacia Amersham
pH meter	Metler-Toledo, Giesen
XCell Sure Lock Mini NuPAGE cell	Invitrogen, Karlsruhe

2.3 Chromatography, commercial kits and solutions

18 cm IPG strips	GE Healthcare
7 cm IPG strips	GE Healthcare
C18 Empore disk	Sigma-Aldrich, Steinheim
IPG strip cover fluid	Pharmacia Biotech

NuPAGE Antioxidant	Invitrogen, Karlsruhe
NuPAGE LDS Sample Buffer (4x)	Invitrogen, Karlsruhe
NuPAGE MOPS SDS Running Buffer, 10x	Invitrogen, Karlsruhe
NuPAGE Novex 4-12% Bis-Tris gels	Invitrogen, Karlsruhe
NuPAGE Sample Reducing Agent (10x)	Invitrogen, Karlsruhe
ReproSil-Pur C18 AQ 3 μ m resin	Dr. Maisch
ReproSil-Pur C18 AQ 5 μ m resin	Dr. Maisch
Sep-Pac C18 Cartridge	Waters

2.4 Two-dimensional cell culture of CPCs (CPC-2D)

Clonal CPCs, #241 (Koelling, Kruegel et al. 2009), derived from single cell and immortalized using hTERT (Docheva, Padula et al. 2010), were grown in 75 cm² flasks (Sarstedt), starting at 1x10⁵ cells per flask. Cells were cultured in 10 ml Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % fetal bovine serum (FBS) (Gibco) supplemented with penicillin/streptomycin (50,000 U/50 mg) and L-glutamine (10 mM). The cell medium was changed every 3-4 days and cells were split when they reached 70-90 % confluence. For 10-12 days the cell numbers reached values between 5x10⁵ and 1x10⁶. The CPC culturing and cell collection was performed by Christa Bode at the department of Prof. Nicolai Miosge (UMG, Göttingen).

2.5 Metabolic labeling of CPCs

For metabolic labeling, clonal CPCs, #241 (Koelling, Kruegel et al. 2009), were cultured as described in the previous section with the following modifications. For cell culture, SILAC DMEM medium (Thermo Scientific) was used, supplemented with (heavy) Arg +6 or +10 (Eurisotop) at concentration 15 mg / 500 ml and (heavy) Lys +4 or +8 (Eurisotop) at concentration 30 mg / 500 ml and 10% dialyzed FBS (PAA Laboratories). Cells were cultured for 5 weeks reaching amino acid incorporation of more than 95%. The CPC culturing and cell collection was performed by Christa Bode at the department of Prof. Nicolai Miosge (UMG, Göttingen).

2.6 Three-dimensional CPC culture in alginate beads (CPC-3D)

CPCs were grown in monolayer until they reached 70 % confluence. Next, cells were detached from the cell culture surface using trypsinization for 5 min. Trypsin activity was blocked by adding 5 ml DMEM supplemented with 10% FBS. Next, cells were pelleted for

10 min at 1000 rpm, washed with PBS and pelleted again. For cell counting cells were dissolved in 1ml of PBS and counted using a cell counter (Cellometer, Nexcelom Bioscience). For chondrogenic differentiation pelleted cells were dissolved in 1.2 % alginate solution. The solution was prepared by mixing equal volumes of 0.15 M NaCl and 2.4 % alginate (Keltone) m/v solution. 3D cell cultures were initiated by dissolving 5×10^4 cells in 20 μ l of 1.2 % alginate solution and drop wise pipetting of the mixture into a 6 well plate containing 4 ml / well of 0.1 M CaCl_2 leading to alginate polymerization and cell embedment. In total 8 alginate beads per well were prepared and allowed to polymerize for 10 min. Next, the CaCl_2 solution was removed and the alginate beads were washed with 2 ml of DMEM medium and incubated in 4 ml DMEM medium for 3 days. Afterwards the DMEM was substituted with media inducing chondrogenesis. This media was composed of DMEM containing high glucose DMEM containing 10^{-7} M dexamethasone, 50 mg/ml ascorbate-2-phosphate, 40 mg proline, 100 mg/ml pyruvate, and 50 mg/ml ITS+Premix (Koelling, Kruegel et al. 2009). Cells were washed every 3-4 days for a total of 24 days. In order to collect the embedded CPCs, alginate beads were lysed in a solution containing 10mM HEPES and 55mM EDTA pH 7.4. The solution was centrifuged for 10 min at 1200 rpm and the resulting CPCs pellet was washed with PBS, pelleted again and stored at -80°C for proteomic analysis. Cell culture was performed by Christa Bode at the department of Prof. Nicolai Miosge (UMG, Göttingen).

2.7 Extraction of proteins from CPCs and protein separation

2.7.1 Biological replicate one; 1DE and 2DE analysis

For the quantification of relative abundance of proteins, heavy and light Arg/Lys labelled CPCs were pooled in equal amounts and protein extraction was performed a solution of 7M urea, 2M thiourea, 4% CHAPS, 2% DTT, 2% Pharmalyte 3-10. Next, the lysate was cleared at 100000 g ultracentrifugation and the supernatant was used for 2DE analysis. 2DE analysis was carried out using 500 μ g of protein extract which were separated using isoelectric focusing in the pH range 4 to 9 in the first dimension, followed by SDS-PAGE in the second dimension. For 1DE analysis, the protein extract was diluted with LDS solution and was separated on a NuPAGE gel (Invitrogen). After 1DE separation, the whole lane was fractionated into 24 slices and subjected to trypsin in-gel digestion. After 2DE analysis, the whole gel was cut into 24 columns and 48 rows resulting in total of 1152 gel slices, which were subjected to trypsin digestion. The sample preparation of biological replicate 1 was performed in the laboratory of Prof. Bernhard Kuester (TU Muenchen). The peptides from each 2DE slice were dissolved in 22 μ l 1% formic acid, 5% ACN from which 10 μ l

were injected for LC-MS/MS analysis using either long (30 min) or short (15 min) gradient, depending on the Coomassie staining of the slice.

2.7.2 Biological replicate two; 1DE analysis

CPC-2D and CPC-3D cell pellets were thawed on ice for 10 min and cells were dissolved in ice cold PBS by gently pipetting up and down at a concentration of 1 million cells per 100 μ l. Next, 100 μ l from the CPC-2D/CPC-3D cell suspension was pooled together and spin down at 1000 g for 10 min using a Heraeus microcentrifuge. The supernatant was discarded and cells were overlaid with 200 μ l of 4% SDS solution and heated for 10 min at 70°C in a Thermomixer Comfort (Eppendorf). In order to reduce viscosity and completely disrupt cells, the cell solution was sonicated for a total of 9 seconds using a Branson Sonifier (Emerson). The cells were sonicated for 3x3 sec with 3 second pauses in between. Minimum heating of the sample was experienced during sonication. Prior to SDS-PAGE analysis, the sample was cleared from debris by centrifugation at 13000 rpm on a Heraeus microcentrifuge (Thermo Scientific). The protein extract was frozen at -80°C or used for further analysis. For 1DE analysis, 20 μ l of the protein extract were mixed with 3 μ l NuPAGE Sample Reducing Agent (10x), 2 μ l NuPAGE LDS Sample Buffer (4x) (both Invitrogen) and 5 μ l H₂O to a final volume of 30 μ l. The sample was heated for 10 min at 70°C in a Thermomixer Comfort (Eppendorf) and spin down at 13000 rpm for 5 min. Next the whole volume was applied to a NuPAGE Novex 4-12% Bis-Tris gel and protein separation was performed at constant 200V for 50 min. The protein sample from three separate CPC-2D/CPC-3D cell pools was used for 1DE analysis. After completion of the electrophoretic separation, the 1DE gel was stained overnight with Colloidal Coomassie (Neuhoff, Arold et al. 1988) and then destained with water for 2h. The whole lane was fractionated into 23 slices using an in-house developed gel cutter (Schmidt and Urlaub 2009). All 23 slices were subjected to in-gel trypsin digestion. Extracted peptides were dissolved in 20 μ l 1% formic acid, 5% ACN from which 5 μ l were injected for LC-MS/MS analysis.

2.8 Ethanol precipitation

Ethanol precipitation of protein mixtures was carried out as described (Schmidt, Lenz et al. 2010). A protein solution was mixed with 3x volume ice cold ethanol (Merck) and 1/10 volume 0.5 M sodium acetate buffer (pH 5.3). The mixture was incubated for at least two hours at -20°C and centrifuged for 30 min at 4°C, 13300 rpm on a Fresco microcentrifuge (Thermo Scientific). After centrifugation, the supernatant was removed and the protein pellet was overlaid with 200 μ l 80% ethanol (Merck). The sample was centrifuged again 30

min at 4°C, 13300 rpm, the supernatant was discarded and the protein precipitate was dried for 5 min in SpeedVac (Thermo Scientific) after which it was used for analysis or frozen at -80°C.

2.9 HeLa cell nuclear extract sample preparation

HeLa cell nuclear extract (NE) was prepared according to Dignam, Lebovitz et al. (1983) and was provided by Hossein Kohansal from the laboratory of Prof. Reinhard Luhrmann, Max Planck Institute for Biophysical Chemistry. NE mixture corresponding to 70 µg of protein was precipitated with ethanol, and the resulting pellet was dissolved in 30 µL 1× LDS sample buffer (Invitrogen). The protein sample was separated on a gradient 4%–12% NuPAGE Novex Bis-Tris gel (Invitrogen). Colloidal Coomassie-stained gels were fractionated into 23 slices. All slices showing Coomassie staining (slices 2–22) were subjected to in-gel trypsin digestion.

2.10 In-gel trypsin digestion

In-gel trypsin digestion was performed as described (Schmidt and Urlaub 2009). All incubation steps were performed at 26°C in thermomixer (Eppendorf) at 1050 rpm for 15 min, unless otherwise stated. Gel slices were washed with 150 µl water for 5 min and dehydrated with 150 µl acetonitrile (ACN). Gel pieces were dried for 5 min and then rehydrated with 100 µl reducing solution (100 mM DTT in 0.1 M ammonium carbonate, pH 8.0). Gel slices were incubated at 56°C for 50 min and then dehydrated with 150 µl ACN. After reduction, the proteins were alkylated with 100 µl 60 mM IAA in 0.1 M ammonium carbonate) for 20 min at 26°C. The gel pieces were then washed with 150 µl ammonium carbonate (pH 8.0), followed by addition of 150 µl ACN. Next, the solution was removed and the gel slices were dehydrated with 150 µl ACN and dried in a hood for 10 min. The dried gel slices were subjected to in-gel digestion. For this gel slices were incubated for 15 min with 10 µl of digestion buffer. The digestion buffer was composed of 45 µl trypsin (12.5 µg/ml, Roche), 150 µl 0.1 M ammonium carbonate (pH 8.0), 15 µl CaCl₂ (5 mM) and 150 µl H₂O. After 15 min the gel slices were overlaid again with 10 µl of digestion buffer. Afterwards the gel slices were covered with buffer to prevent drying. The buffer composition was 150 µl 0.1 M ammonium carbonate (pH 8.0), 15 µl CaCl₂ (5 mM) and 195 µl H₂O. The trypsin digestion was carried out overnight at 37°C.

2.11 Peptide extraction

Peptides from in-gel digestion were extracted as described (Schmidt and Urlaub 2009). All incubation steps were performed at 1050 rpm in a Thermomixer Comfort (Eppendorf) at 37°C for 15 minutes. Overnight digested gel slices were incubated with 30 µl water to which, after 5 min incubation, 80 µl ACN were added. The supernatant was transferred to a new tube and the dehydrated gel pieces were incubated with 50 µl 5 % formic acid, followed by the addition of 50 µl ACN. The supernatant step was pooled with the solution from the first extraction step. Finally the pooled supernatants were dried using a SpeedVac (Thermo Scientific) stored at -20°C or used for further analysis.

2.12 In-solution digestion

In-solution trypsin digestion was performed as described (Schmidt, Lenz et al. 2010). For in-solution digestion, 70 or 100 µg of ethanol-precipitated HeLa NE were used. Overall 5 in-solution digestion experiments were performed, one with 70 µg and four using 100 µg HeLa NE. The 100 µg HeLa NE protein sample was used for the optimization of the 18 cm pIEF fractionation scheme.

The precipitated protein pellet was dissolved in 8 M urea (Serva) and digested with Lys-C (Roche) (1:50 protease to protein ratio) for 3 h. The sample was diluted with 0.1 M ammonium bicarbonate solution (pH 8.0) to 2 M urea and digested with trypsin (Promega) (1:50) overnight at 25°C at 850 rpm in SpeedVac. The resulting peptides were desalted using STAGE tip (Rappsilber, Ishihama et al. 2003) with 12 Empore plugs (Sigma-Aldrich) or using Sep-Pac C18 cartridge (Waters).

2.13 pIEF-LC-MS/MS analysis

For the pIEF analysis, dried peptides were dissolved in 350 µL of 8 M urea, 0.2% IPG buffer (GE Healthcare), and then applied to the strip holder. 18 cm IPG strip (pH 3-10; GE Healthcare) was used. The sample was pipetted onto the strip holder and the 18 cm IPG strip was applied by lowering and raising it three times. Then, the strip was positioned on the incline at the cathodic end of the strip holder and pushed with forceps, until the IPG gel reached the bottom of the strip holder. Next, the IPG strip was covered with around 1.5 ml cover fluid (Pharmacia Biotech) and was rehydrated overnight at 20°C without applied voltage. Following rehydration, peptides were separated on an IPGphor (Pharmacia Amersham) for a total of 30 000 Vh at max 50 µA per strip. The following parameters were used: overnight rehydration at 0 V then 500 Vh at 500 V, then 1750 Vh at gradient 500 V to

3000 V, then 27750 Vh at 8000 V all steps performed at 20 °C. After pIEF, IPG strips were cleaned of the cover fluid by immersing them for 10 sec in n-hexane (Merck) and manually stripped from IPG plastic into 36 pieces (0.5 cm each). Peptides were then extracted by sequentially incubating the IPG strip slices for 30 min each with 1% formic acid; 50% ACN, 1% formic acid; and 99% ACN, 1% formic acid. Extracted peptides were then desalted using STAGE tips (Rappsilber, Ishihama et al. 2003) using 2 plugs. Peptides were dried using SpeedVac (Thermo Scientific) for around 2 hours. Extracted peptides were used for LC-MS/MS analysis or frozen at -20°C. For LC-MS/MS analysis peptides were dissolved in 20 µl 1% formic acid, 5% ACN from which 5 µl per replicate were injected for LC-MS/MS analysis.

2.14 PAGE-pIEF-LC-MS/MS analysis

For the PAGE-pIEF-LC-MS/MS analysis 70 µg ethanol precipitated HeLa NE were dissolved in 20µL 1x LDS NuPAGE buffer (Invitrogen) and separated on a NuPAGE Novex 4-12% Bis-Tris gel (Invitrogen). All steps were identical to those described in session 2.9, except that peptides extracted from the PAGE slices were subsequently separated on a 7 cm IPG strips (pH 3-10, GE Healthcare). Tryptic peptides were extracted from the PAGE slices as explained, dried and dissolved in 150 µL solution of 8 M urea, 0.2% IPG buffer. After overnight rehydration without applied voltage, separation was carried out using IPGphor (GE Healthcare) for a total of 9950 Vh. IEF was performed using the following parameters: max 50 µA per strip, 250 Vh at 500 V, then 6500 Vh at gradient 500 V to 6000 V, and then 3200 Vh at 6000 V. All steps were performed at 20 °C. Following pIEF, the IPG strip was fractionated into 13 slices (0.5 cm each), and peptides were extracted and desalted as described in the previous section.

2.15 Comparison of 7 cm pIEF-LC-MS/MS and long LC gradient

For the comparison between 7 cm pIEF and a long LC gradient two samples of 70 µg HeLa NE were separated in two separate lanes on a NuPAGE Novex 4-12% Bis-Tris gel. PAGE slice #10 was chosen for analysis. After in-gel digestion peptides were extracted and dried and redissolved in 22 µL 5% ACN with 0.1% formic acid. The sample was analyzed twice using of 5 µL of the sample using a 33 min LC gradient and twice using 5 µL using a 221 min LC gradient; 5%–38% buffer B. In parallel, total peptides derived from PAGE slice #10 from the second lane were separated using pIEF. Tryptic peptides were extracted from the

pIEF strip and desalted. Dried peptides were redissolved in 22 μL 5% ACN with 0.1% formic acid and 5 μL of the sample was analyzed twice by LC-MS/MS.

2.16 PAGE-pIEF-LC-MS/MS analysis of CPC culture media

CPCs conditioned cell culture medium was collected 24 hours after removing the old culture medium, washing the cells with PBS and culturing the cells in medium without FCS. Cell culture medium was collected by Christa Bode at the department of Prof. Nicolai Miosge (UMG, Göttingen). Cell media was TCA precipitated (Chen, Lin et al. 2005). For this purpose, 1500 μl of culture medium were precipitated in a single Eppendorf tube; a total of 15000 μl of cell culture medium were precipitated into 10 tubes. The protein precipitates were separated using 1D PAGE (section 2.9). All 10 lanes were fractionated into 23 slices (Schmidt and Urlaub 2009) and all gel fractions with numbers 16 to 22 from all 10 lanes were combined and digested with trypsin except that Ca^{2+} was not added to the digestion mixture. TCA precipitation, 1D PAGE analysis and digestion were performed by Sarah Gaida (Bioanalytical Mass Spectrometry group, MPI-BPC). Peptides were later separated using pIEF as explained. After peptide extraction, peptides were dissolved in 20 μl 1% formic acid, 5% ACN from which 10 μl were injected for LC-MS/MS analysis using an Orbitrap-Velos mass spectrometer (Thermo Scientific).

2.17 LC-MS/MS analysis

LC separation was performed on an Agilent 1100-LC system (Agilent Technologies). Peptides were first loaded onto a trap column packed in-house (2 cm, 360 μm o.d., 150 μm i.d.; ReproSil-Pur, C18, AQ 5 μm , Dr. Maisch HPLC GmbH). Peptide separation was performed at a flow rate of 300 nL/min on a 12 cm analytical column (75 μm internal diameter, packed with ReproSil-Pur C18 AQ 5 μm resin; Dr. Maisch, Germany). The same analytical column dimensions were used for the analysis of CPC-2D/CPC-3D biological replicate 1 (both 1DE and 2DE analysis), HeLa NE analysis using 1DE, 18 cm pIEF and PAGE-pIEF-LC-MS/MS and CPC secretome analysis.

Alternatively, an analytical column with 75 μm ID packed with ReproSil-Pur C18 AQ 3 μm resin (Dr. Maisch, Germany) was used for the analysis of CPC-2D/CPC-3D biological replicate 2 and for the pIEF fractionation optimization experiments.

Reversed-phase separation of peptides was performed using a gradient mixture of 0.1% formic acid in water (buffer A). Buffer B was composed of 95% ACN, 0.1% formic acid. For the sample analysis, a 33 min or a 15 min, 5%–38% buffer B gradient was used. Eluting peptides were analyzed on a hybrid LTQ-Orbitrap-XL mass spectrometer (Thermo

Scientific), operating in a data-dependent mode. Survey full scan MS spectra were acquired in the orbitrap (m/z 350–1600) with a resolution of 30000. The automatic gain control (AGC) target was at 1E6. The five most intense ions with were selected for CID MS/MS fragmentation in the linear ion trap at an AGC target of 3E5. This mode of operation was used for the analysis of CPC-2D / CPC-3D biological replicate 1 (both 1DE and 2DE analysis), HeLa NE analysis using 1DE, 18 cm pIEF and PAGE-pIEF-LC-MS/MS analysis. Alternatively, the eight most intense ions with were selected for CID MS/MS fragmentation in the linear ion trap. This was the case for the analysis of CPC-2D/CPC-3D biological replicate 2 and the pIEF fractionation optimization experiments. In both cases ions with unrecognized charge state and with charge state of one were excluded. Detection in the linear ion trap of previously selected ions was dynamically excluded for 60 s. Internal calibration of the orbitrap was performed using the lock mass option (Olsen, de Godoy et al. 2005).

For the secretome study, the analysis was done using a LTQ Orbitrap Velos mass spectrometer. The instrument was operated in the following manner: full scan MS spectra were acquired in the orbitrap (m/z 350–1600) with a resolution of 30,000. The 15 most intense ions with were selected for CID MS/MS fragmentation in the linear ion trap and the dynamic exclusion was set to 30s.

2.18 Protein identification and quantification from MS data

In this study, protein identification and quantification was carried out using MaxQuant (Cox and Mann 2008). For the PAGE-pIEF-LC-MS/MS study, proteins were identified with MaxQuant (version 1.0.13.13), using the Mascot search engine (version 2.3.03, Matrix Science). The International Protein Index (IPI) (Kersey, Duarte et al. 2004) human protein sequence database (version 3.72) was used. The database was concatenated with typical contaminants and reversed sequences using the SequenceReverser.exe tool from the MaxQuant software package (Cox and Mann 2008). The default search parameters were used for the MaxQuant analysis. For the main search, 7 ppm MS tolerance, 0.5 Da MS/MS tolerance, maximum 2 missed cleavages, 1% protein and peptide false discovery rate was used. The option “keep low scoring version of identified peptides” was turned off. Methionine oxidation and cysteine carbamidomethylation were set as variable and fixed modifications, respectively. The raw files from the HeLa NE 1DE reproducibility experiment were analyzed with MaxQuant (version 1.3.0.5) using the Andromeda search engine (Cox, Neuhauser et al. 2011). The data was searched against the IPI human protein sequence database (version 3.72).

The data from the CPC-2D/CPC-3D replicate experiments, CPC secretome analysis and 18cm pIEF fractionation optimization were searched against the Uniprot (The UniProt Consortium 2014) human proteome reference dataset, downloaded on 24/11/2012. For those experiments, data analysis was done using MaxQuant version 1.0.13.13, using the default parameters.

For the protein quantification of metabolically labeled samples, the multiplicity parameter of MaxQuant was set to two. The heavy labels were set to Arg10/Lys8 or Arg6/Lys4 and the minimum ratio count was 2. For quantification unique and razor proteins were considered. The re-quantify option was turned on for all analyses, except the analysis of the data from the analysis of proteins *in vitro* secreted by CPCs in order to evaluate the effect of the option on the quantification results.

2.19 Identification of differentially abundant proteins

Analysis for significance of differential protein expression was carried out using the Perseus statistical suite (Cox and Mann 2012). For the analysis, the raw files of CPC-2D / CPC-3D biological replicates 1 and 2 were searched together using MaxQuant. The MaxQuant file ProteinGroups.txt was loaded in Perseus, and contaminants and reverse database hits were removed. The ratio H/L was transformed to log₂ scale and the protein intensity was transformed to log₁₀ scale. The statistical test used for testing for differential abundance was significance B (Cox and Mann 2008). A threshold p value of 0.05 was set. Finally, the overlap of the significant proteins was determined and only proteins termed significant in both biological replicates were selected for functional enrichment analysis.

2.20 Functional enrichment analysis of differentially expressed proteins

The sets of proteins which were significant in the analyses of both biological replicates were analyzed for functional enrichment using the DAVID bioinformatics suite (Huang da 2007). The gene names of the up or downregulated proteins were loaded and “Homo sapiens” was selected as species. Next, an enrichment analysis for the following three categories was performed: GOTERM_BP_FAT, GOTERM_CC_FAT and GOTERM_MF_FAT. The results of the “Functional annotation chart” were used and values with functional enrichment p values of more than 0.05 were considered in this work.

2.21 1DE and 2DE protein intensity and ratio H/L ratio data analysis and comparison

In order to compare the distribution of protein intensities/ratios over the 1DE or 2DE gel, the results from the evidence.txt results file of MaxQuant were used. For both separation approaches, the slice specific information for each identified peptide and protein was calculated. This was done using the following calculations: sum of the intensity, product of PEP, log₂ transformation of the H/L ratios, median of the transformed H/L ratios. The newly calculated values were plotted using the R statistical software (R Development Core Team 2010).

2.22 Calculation of average protein GRAVY scores and pI

GRAVY (grand average of hydropathy) scores were calculated using the amino acid hydrophobicity values from Kyte-Doolittle hydrophobicity scale (Kyte and Doolittle 1982). The sum of the hydrophobicity values of all amino acids in a protein was calculated and the sum was divided by the length of the protein in order to obtain the GRAVY score.

The protein pI was calculated using the seqinr package, R (R Development Core Team 2010).

2.23 Protein sequence similarity analysis

For the sequence similarity analysis, the Uniprot protein identifiers (IDs) were extracted from the ProteinGroups.txt (MaxQuant output) file. Since for most protein groups several protein IDs are provided, the first reported ID (explaining the highest number of identified peptide) was extracted. Protein isoforms were selected based on the entries in the “Gene names” column of the ProteinGroups.txt file. Protein groups which had the same gene name were used for the analysis as protein isoforms are separate protein species which are encoded by the same protein. Next the protein sequences which correspond to the selected protein IDs were extracted from Uniprot (www.uniprot.org). The sequence similarity was calculated using two different methods.

First, the similarity was calculated without gap penalty. For this, extracted protein sequences were aligned using the Clustal Omega multiple sequence alignment suite (www.ebi.ac.uk/Tools/msa/clustalo) (Sievers, Wilm et al. 2011). The default parameters for sequence alignment were used. The values from the “Percent Identity Matrix” were used and reported in this study.

Second, sequence similarity was calculated with gap penalty. For this, the protein sequences were aligned using the “Align” tool from Uniprot (www.uniprot.org/align). This tool is also based on Clustal Omega (Sievers, Wilm et al. 2011) but the sequence identity is reported with gap penalty.

3 Results

3.1 Quantification of proteome wide differences between CPCs cultured on plastic and in alginate

This section is dedicated to the large-scale quantification of proteins from CPC's grown either on 2D plastic culture (CPC-2D) or in 3D alginate beads (CPC-3D). The encapsulation of CPCs in alginate beads results in a change of the cell shape from ellipsoid (2D cell culture) to round (alginate culture) and the cells initiate chondrogenic differentiation (Koelling, Kruegel et al. 2009). The differentiation is characterized by the presence of mRNA of a major chondrogenic marker – collagen type II. This is in contrast to the 2D culture, in which case, CPCs express very little collagen type II. While examining the expression of collagen type II can reveal changes in the cellular differentiation state, comparison of the global protein profiles would reveal more detail of the differentiation process and the nature of CPCs. In order to compare the protein expression on large-scale, CPC-2D cells were labeled with heavy SILAC amino acids (Arg +10 and Lys +8). In a second biological replicate the CPC-2D cells were labeled with Arg +6 and Lys +4. In both cases, CPC-3D cells were labeled with light version of the Arg and Lys. Next, cells were pooled together, lysed and the extracted protein content was analyzed downstream using 1D PAGE separation, gel fractionation, in-gel trypsin digestion and finally LC-MS/MS analysis. By combining the cells, all sample processing steps (i.e. cell lysis, fractionation, protein digestion) are performed simultaneously. In addition, the signal from peptides derived from cellular proteins from both CPC-2D and CPC-3D will be present in the mass spectrometer and measured at the same time. A calculation of the ratio of the signal from the CPC-2D (heavy) peptides and the signal from the CPC-3D (light) peptides can be used to derive the relative differences in the abundance of proteins of CPCs undergoing chondrogenesis. Proteins with high heavy/light (H/L) ratio would indicate that the protein is more abundant/upregulated in the CPC-2D relative to CPC-3D. At the same time, low H/L ratios will indicate higher abundance/upregulation of the protein in CPC-3D, relative to CPC-2D. The results of the relative quantification of proteins between CPC-2D and CPC-3D are shown in (Figure 1).

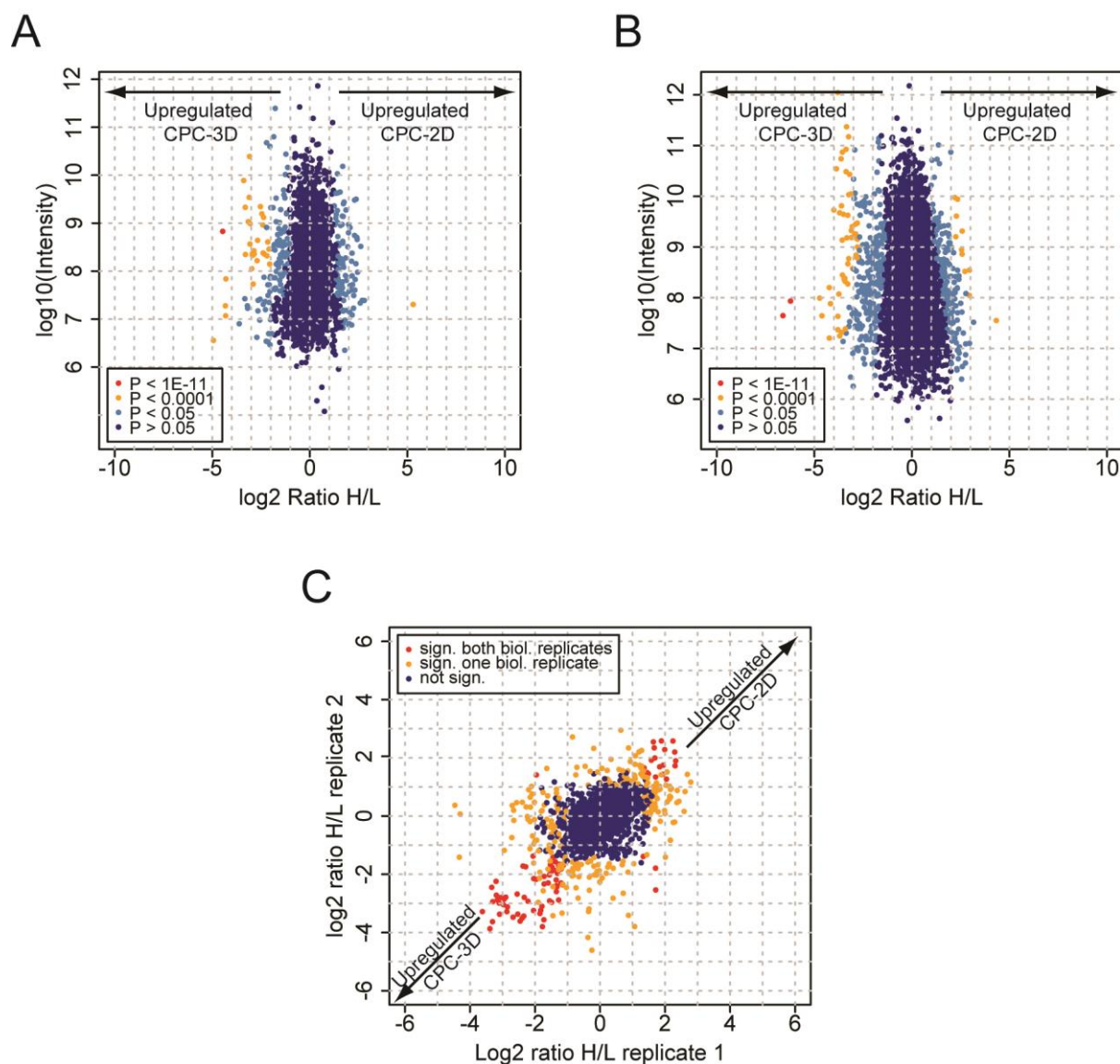


Figure 1. Quantification of proteins in CPCs grown on plastic (CPC-2D) and CPCs grown in alginate (CPC-3D).

A) First biological replicate. B) Second biological replicate. C) Comparison of protein H/L ratios in the first and second biological replicate.

The analysis of the first biological replicate resulted in the quantification of around 1900 proteins. In the second experiment around 4200 proteins were quantified. From all quantified proteins, 230 were significantly different (up or downregulation) in the first experiment (Figure 1, A) and 457 proteins in the second experiment (Figure 1, B) based on MaxQuant significance B, $p < 0.05$. Within each experiment the lowest H/L ratios that were identified to have significant differential abundance were 1 and -1.3 (log₂ scale).

The sample preparation was different for the biological replicate experiments. In both cases the cells were pooled together but lysed in different lysis conditions, either 8 M urea or 4 % SDS. Such differences can contribute to some variability in the quantification results on top of the biological variability of the CPCs. In order to increase the confidence in the protein

candidates identified as statistically significant and to narrow down the list of potential proteins showing biological significance, the intersection of protein lists showing statistical significance in both replicate experiments was selected (Figure 1, C). From the 230 and 457 significant proteins in the first and second biological replicate only 87 proteins were significant in both. Four proteins, significant in both biological replicates, showed inconsistent ratios in the separate experiments: protein EIF3G (ratio -1.95 and 1.41), protein FTL (ratio 1.33 and -1.39), protein PCOLCE (ratio 1.71 and -1.79) and protein DSP (ratio 1.71 and -2.54). The remaining 83 proteins, significant in both biological replicate showed consistent ratios (Table 1 and Table 3).

Overall, 21 proteins were found to exhibit higher abundance in CPC-2D compared to CPC-3D and were significant in both biological replicates (Table 1).

Table 1. Proteins showing higher expression in CPC-2D compared to CPC-3D, significant in both biological replicates

Uniprot ID	Gene	Protein name	Log 2 Ratio H/L		Significance p value	
			Rep 1	Rep 2	Rep 1	Rep 2
Q09666	AHNAK	Desmoyokin	1.69	1.99	0.007565	0.000104
Q8IVF2	AHNAK2	Protein AHNAK2	1.99	2.28	0.001560	0.000011
B4DTU0	ALCAM	Activated leukocyte cell adhesion molecule	1.35	1.65	0.033175	0.003068
P16070	CD44	CD44 antigen	1.41	0.98	0.025863	0.037879
P21291	CSRP1	Cysteine and glycine-rich protein 1	2.33	1.90	0.000721	0.000666
Q14247	CTTN	Amplaxin	2.04	1.27	0.015878	0.043015
P15311	EZR	Ezrin	1.58	1.92	0.030011	0.002019
Q6UWF9	FAM180A	Protein FAM180A	1.83	1.69	0.011813	0.015821
Q14315	FLNC	Filamin-C	1.44	1.49	0.034153	0.002837
Q9BQS8	FYCO1	FYVE and coiled-coil domain-containing protein 1	1.94	1.43	0.004780	0.021691
P17301	ITGA2	Integrin alpha-2	1.67	2.34	0.014938	0.000007
P26006	ITGA3	Integrin alpha-3	1.42	1.36	0.024997	0.015420
J3KN41	ITGA6	Integrin alpha-6	1.89	2.58	0.025727	0.000003
Q9UHB6	LIMA1	LIM domain and actin-binding protein 1	2.24	2.58	0.008283	0.000031
Q6UVY6	MOXD1	DBH-like monooxygenase protein 1	1.49	1.45	0.029406	0.009455
P09874	PARP1	NAD(+) ADP-ribosyltransferase 1	2.31	1.72	0.000795	0.005572
P47712	PLA2G4A	Cytosolic phospholipase A2	1.64	2.55	0.016589	0.000039
Q13637	RAB32	Ras-related protein Rab-32	1.44	1.12	0.047510	0.045829
Q9P2E9	RRBP1	Ribosome-binding protein 1	2.30	2.20	0.001626	0.000022
O60637	TSPAN3	Tetraspanin-3	1.95	1.75	0.007551	0.038913
J3QRU1	YES1	Tyrosine-protein kinase Yes	1.71	1.34	0.042850	0.032314

This list of significant proteins contained a number of prominent examples associated with cellular binding to the extracellular matrix. This included three members of the integrin alpha family (members 2,3 and 6) and CD44 – the receptor for hyaluronic acid (Isacke and Yarwood 2002). In addition, ALCAM a marker for MSCs (Bruder, Ricalton et al. 1998), was also found to have higher expression in CPC-2D compared to CPC-3D. Several of the proteins identified to have higher expression in CPC-2D are associated with the cellular cytoskeleton: LIMA1 (Chervin-Pétirot, Courçon et al. 2012), CSRP1 (Bianchi-Smiraglia, Kunnev et al. 2013), CTTN (Cosen-Binker and Kapus 2006), EZR (Tudor, te Riet et al. 2014) and FLNC (Stossel, Condeelis et al. 2001). In order to examine whether proteins associated with specific cellular functions or compartments were overrepresented in the list of proteins, upregulated in CPC-2D, functional enrichment analysis was performed. The set of 21 proteins was analyzed using DAVID bioinformatics package (Huang da 2007). The set of enriched GO categories associated with the proteins having higher expression in CPC-2D is shown in Table 2.

Table 2. Gene ontology enrichment analysis of significant proteins with higher expression in CPC-2D compared to CPC-3D.

GO Category*	GO ID	Term	Count	Enrichment p value
BP	GO:0007160	cell-matrix adhesion	4	7.51E-05
BP	GO:0031589	cell-substrate adhesion	4	1.00E-04
BP	GO:0007155	cell adhesion	6	3.32E-04
BP	GO:0022610	biological adhesion	6	3.35E-04
CC	GO:0008305	integrin complex	3	5.85E-04
CC	GO:0005912	adherens junction	4	8.72E-04
BP	GO:0009991	response to extracellular stimulus	4	0.001075694
CC	GO:0009897	external side of plasma membrane	4	0.001139546
CC	GO:0070161	anchoring junction	4	0.001178681
BP	GO:0033273	response to vitamin	3	0.001766239
CC	GO:0016323	basolateral plasma membrane	4	0.001897453
BP	GO:0007229	integrin-mediated signaling pathway	3	0.001984257
BP	GO:0060249	anatomical structure homeostasis	3	0.004484114

* BP – biological process, CC – cellular compartment.

The functional enrichment analysis of the proteins showing higher abundance in CPC-2D revealed enrichment for cell-matrix (p value 7.51E-05) and cellular (p value 3.32E-04) adhesion in addition to enrichment for integrin complex (p value 5.85E-04), Table 2. The GO

categories of the proteins having higher abundance in CPC-2D show an association with the phenotype of the CPCs when grown on plastic surface – flat cells, spread on the 2D surface and attached to adsorbed proteins on the plastic surface using integrin receptors.

From the 83 proteins showing significance in both biological replicates, 62 showed higher expression in CPC-3D compared to CPC-2D (Table 3).

Table 3. Proteins showing higher expression in CPC-3D compared to CPC-2D, significant in both biological replicates

Uniprot ID	Gene	Protein name	Log 2 Ratio H/L		Significance p value	
			Rep 1	Rep 2	Rep 1	Rep 2
Q9NP58	ABCB6	ATP-binding cassette sub-family B	-1.13	-1.67	0.032673	0.023272
P33121	ACSL1	Acyl-CoA synthetase 1	-2.06	-1.38	0.000093	0.046905
P00325	ADH1B	Alcohol dehydrogenase 1B	-2.39	-3.51	0.000005	0.000005
P51857	AKR1D1	3-oxo-5-beta-steroid 4-dehydrogenase	-1.56	-2.03	0.003199	0.004116
P04114	APOB	Apolipoprotein B-100	-2.68	-3.48	0.000000	0.000006
C9JF17	APOD	Apolipoprotein D	-3.16	-2.88	0.000000	0.000039
P07741	APRT	Adenine phosphoribosyltransf.	-1.37	-1.57	0.009293	0.023679
Q05707	COL14A1	COL14A1 protein	-1.73	-3.56	0.002561	0.000004
P08572	COL4A2	Collagen alpha-2(IV) chain	-3.33	-2.45	0.000646	0.001493
P12109	COL6A1	Collagen alpha-1(VI) chain	-1.85	-3.35	0.001272	0.000015
P12110	COL6A2	Collagen alpha-2(VI) chain	-2.05	-3.25	0.000376	0.000028
P12111	COL6A3	Collagen alpha-3(VI) chain	-1.77	-3.80	0.001993	0.000001
O00622	CYR61	CCN family member 1	-2.95	-2.85	0.000000	0.000058
P27487	DPP4	Dipeptidyl peptidase 4	-1.19	-1.81	0.035722	0.009235
Q16555	DPYSL2	Dihydropyrimidinase-related protein 2	-2.37	-1.72	0.000040	0.035261
P68104	EEF1A1	Elongation factor 1-alpha 1	-1.25	-1.66	0.027027	0.043228
Q9Y6C2	EMILIN1	Elastin microfibril interface-located protein 1	-1.82	-3.24	0.000542	0.000030
P00488	F13A1	Coagulation factor XIII A chain	-3.62	-3.28	0.000207	0.000024
P35555	FBN1	Fibrillin-1	-3.11	-2.99	0.000000	0.000123
Q92820	GGH	Gamma-glutamyl hydrolase	-1.48	-1.76	0.005188	0.017183
P36269	GGT5	Gamma-glutamyltransferase 5	-3.19	-2.72	0.001104	0.000123

Results

P07305	H1F0	Histone H1.0	-1.52	-1.93	0.019946	0.006280
O95479	H6PD	6-phosphogluconolactonase	-1.72	-2.07	0.002627	0.009602
P69905	HBA1	Hemoglobin alpha chain	-2.97	-3.04	0.000004	0.000068
Q92743	HTRA1	Serine protease HTRA1	-3.39	-3.87	0.000000	0.000000
P35475	IDUA	Alpha-L-iduronidase	-2.90	-2.79	0.000007	0.000083
C9JAF2	IGF2	Insulin-like growth factor II	-2.35	-3.44	0.000280	0.000008
P17936	IGFBP3	Insulin-like growth factor-binding protein 3	-3.07	-2.80	0.000000	0.000340
Q96CN7	ISOC1	Isochorismatase domain-containing protein 1	-1.53	-1.67	0.019017	0.023262
P04264	KRT1	Cytokeratin-1	-3.31	-3.63	0.000000	0.000002
P35908	KRT2	Cytokeratin-2e	-2.85	-3.12	0.000000	0.000058
G3XAI2	LAMB1	Laminin B1 chain	-1.97	-2.17	0.000186	0.006445
P11047	LAMC1	Laminin subunit gamma-1	-1.34	-2.39	0.018417	0.002499
P21397	MAOA	Amine oxidase [flavin-containing] A	-1.79	-3.10	0.000705	0.000010
Q9H8H3	METTL7A	Methyltransferase-like protein 7A	-1.72	-2.33	0.001115	0.000831
P55001	MFAP2	Microfibril-associated glycoprotein 1	-2.33	-2.79	0.000010	0.000065
Q08431	MFGE8	Lactadherin	-3.09	-3.37	0.000000	0.000013
P08473	MME	Neprilysin	-1.33	-2.43	0.019643	0.002090
P14543	NID1	Nidogen-1	-3.20	-2.24	0.001066	0.001507
Q14112	NID2	Nidogen-2	-2.55	-2.68	0.009375	0.000126
P40261	NNMT	Nicotinamide N-methyltransferase	-2.46	-3.61	0.000021	0.000003
Q8IXM6	NRM	Nurim	-1.45	-1.42	0.026495	0.040490
Q8IV08	PLD3	Phospholipase D3	-1.40	-1.87	0.007879	0.007380
B1ALD9	POSTN	Periostin	-2.87	-3.28	0.000000	0.000022
A8MU24	PRCP	Lysosomal Pro-X carboxypeptidase	-1.31	-1.97	0.020978	0.004710
P26022	PTX3	Pentaxin-related protein PTX3	-3.30	-2.94	0.000000	0.000159
P07093	SERPINE2	Glia-derived nexin	-1.55	-2.92	0.006598	0.000183
Q9HAT2	SIAE	Sialate O-acetyltransferase	-2.04	-2.15	0.001630	0.002336
Q8WWI5	SLC44A1	Choline transporter-like protein 1	-1.60	-2.30	0.014135	0.001164
P04179	SOD2	Superoxide dismutase	-2.22	-2.93	0.000120	0.000171

Results

P78539	SRPX	Sushi repeat-containing protein SRPX	-2.53	-3.37	0.000012	0.000013
P27105	STOM	Erythrocyte band 7 integral membrane protein	-1.34	-2.60	0.018220	0.000956
G8JLA8	TGFBI	Transforming growth factor-beta-induced protein ig-h3	-1.49	-2.96	0.009025	0.000141
P21980	TGM2	Protein-glutamine gamma-glutamyltransferase 2	-1.27	-2.88	0.015979	0.000038
P07996	THBS1	Thrombospondin-1	-1.20	-2.30	0.034426	0.001145
P35625	TIMP3	Metalloproteinase inhibitor 3	-3.06	-2.80	0.000000	0.000063
Q9C037	TRIM4	Tripartite motif-containing protein 4	-1.25	-1.85	0.018137	0.022576
Q71U36	TUBA1A	Tubulin alpha-1A chain	-1.15	-2.36	0.030065	0.002829
P07437	TUBB	Tubulin beta chain	-1.28	-1.69	0.024635	0.038484
P68371	TUBB4B	Tubulin beta-4B chain	-1.13	-1.70	0.045188	0.038068
E7EUC7	UGP2	UDP-glucose pyrophosphorylase	-2.29	-1.74	0.020006	0.013294
P41221	WNT5A	Protein Wnt-5a	-1.35	-2.19	0.039008	0.001671

The set of proteins showing higher expression in CPC-3D compared to CPC-2D contains a variety of proteins with different functions. Interestingly, there were a number of proteins associated with the chondrocyte extracellular matrix (ECM) and more specifically with the so-called pericellular matrix – the narrow ECM region that surrounds chondrocytes in articular cartilage (Poole 1997, Guilak, Alexopoulos et al. 2006). Functional enrichment analysis of the differentially abundant proteins also revealed association with the extracellular matrix (Table 4).

Table 4. Functional enrichment analysis of proteins upregulated in CPC-3D relative to CPC-2D.

GO category	GO ID	Term	Count	Enrichment p value
CC	GO:0044421	extracellular region part	25	5.37E-13
CC	GO:0031012	extracellular matrix	17	1.19E-12
CC	GO:0005578	proteinaceous extracellular matrix	16	5.86E-12
CC	GO:0044420	extracellular matrix part	11	8.41E-11
CC	GO:0005576	extracellular region	29	3.72E-09
BP	GO:0030198	extracellular matrix organization	9	5.28E-09
BP	GO:0043062	extracellular structure organization	9	1.79E-07
BP	GO:0007155	cell adhesion	15	2.45E-07
BP	GO:0022610	biological adhesion	15	2.49E-07
CC	GO:0005604	basement membrane	7	1.20E-06
MF	GO:0005201	extracellular matrix structural constituent	7	1.27E-06
BP	GO:0030155	regulation of cell adhesion	7	1.34E-05
CC	GO:0005615	extracellular space	13	3.89E-05
MF	GO:0005198	structural molecule activity	12	4.03E-05

* BP – biological process, CC – cellular compartment, MF – molecular function.

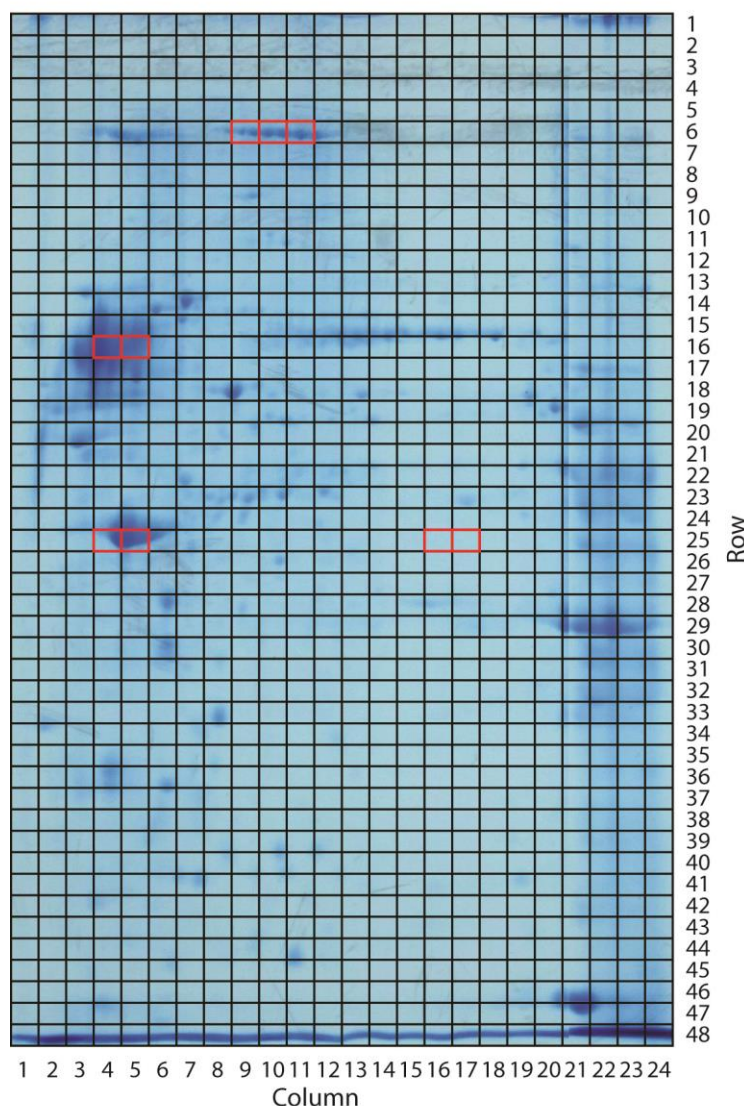
The set of 62 proteins, showing higher abundance in CPC-3D compared to CPC-2D, was enriched for the GO terms “extracellular matrix” (p value 1.19E-12) and “extracellular region” (p value 3.72E-09). In addition, the GO term “basement membrane” was also found to be enriched (p value 1.20E-06). The finding that upregulation of components of the basement membrane is associated with the chondrogenesis of CPCs is interesting since a number of studies have discovered that components of the basement membrane are present in the chondrocyte pericellular matrix *in vivo* (Kvist, Nyström et al. 2008). These observations show that when CPCs are cultured in alginate beads they upregulate genes associated with the pericellular matrix indicating ongoing chondrogenesis.

3.2 Evaluation of 2DE separation for large scale proteome profiling

In parallel to performing a large-scale analysis on the proteome differences between CPC-2D and CPC-3D, using 1D PAGE, a second proteomic experiment was performed. In this

experiment, portion of the pooled protein extract of biological replicate 1 was separated using 2DE. Studies have shown that 2DE is able to separate and visualize a large number of proteins/protein species (Hoving, Voshol et al. 2000) and it is possible to combine 2DE with metabolic labeling (Jiang and English 2002). In addition, performing 2DE on metabolically labeled samples has been shown to be very beneficial as it gives additional quantitative information about the relative abundance of proteins and protein species (Thiede, Koehler et al. 2013). In this line, we tested whether separating the metabolically labeled CPC-2D/CPC-3D sample by 2DE, fractionating the entire 2DE gel and analyzing all fractions by LC-MS/MS would result in the identification and quantification of a large number of proteins and proteins species (either isoforms or post-translationally modified proteins) and whether the 2DE separation approach would perform better than the 1DE separation approach. For this purpose, the protein extract from the first biological sample was separated in the first dimension by isoelectric focusing in the pH 4-9 ranges. Proteins were then separated based on MW using SDS-PAGE in the second dimension. The whole gel was cut into 24 columns and 48 rows resulting in 1152 gel slices (Figure 2). In comparison the whole lane of the 1DE gel was separated into 24 rows.

A



B

Row 6		Part of protein amount in fraction [%]		
Uniprot	Gene	Col 9	Col 10	Col 11
P12110	COL6A2	73.64	64.06	64.87
P12111	COL6A3	14.16	22.32	21.77
Q09666	AHNAK	3.76	6.61	5.83
P02751	FN1	3.71	1.96	1.13

Row 16		Part of protein amount in fraction [%]	
Uniprot	Gene	Col 4	Col 5
P08670	VIM	85.4	54.19
P11142	HSPA8	0.51	21.83
J9JID7	LMNB2	0.25	6.8

Row 25		Part of protein amount in fraction [%]	
Uniprot	Gene	Col 4	Col 5
J9JID7	LMNB2	30.5	8.92
P12111	COL6A3	20.36	19.97
P0CG38	POTE1	10.82	22.35

Row 25		Part of protein amount in fraction [%]	
Uniprot	Gene	Col 16	Col 17
P49411	TUFM	87.48	68.16
I3LOC1	SMG1	2.43	2.04
P08559	PDHA1	1.44	1.39

Figure 2 2DE analysis of pooled protein extracts of heavy SILAC labeled CPC-2D and light SILAC labeled CPC-3D.

Proteins were separated by IEF using pH 4 to 9 gradient followed by PAGE electrophoresis.

A) The whole gel was fractionated into 1152 slices (24 rows by 48 columns). **B)** The proteins contributing to the highest percentage of the protein amount was calculated for selected fractions (red borders). The proportion of protein amount was calculated by dividing the protein MS intensity by the sum of the MS intensity of all proteins. The three or four most abundant proteins and their proportion of total amount in the gel fraction are shown.

An examination of the protein separation in the selected 2DE configuration reveals that there are distinct protein spots within the whole separation space of the gel. Characteristic are sections of the gel with very high protein amounts (e.g. row 15-17, column 4-5), trails of proteins spots (e.g. row 6, column 9-13) and an individual protein spot (e.g. row 46-47, column 21), (Figure 2, A). It is important to consider that using the fractionation scheme employed in this study – to fractionate the whole gel into fractions that in some cases

incorporate a number of protein spots (e.g. row 6, column 10), there will be a number of proteins being identified in the same fraction. This is clearly visible when the contribution of the identified proteins to the protein amount in selected fractions is examined (Figure 2,B). The protein amount in a gel fraction can be inferred from the intensity of the signal in the mass spectrometer (de Godoy, Olsen et al. 2008). One example is the trails of proteins spots in row 6 (Figure 2, A). While there were a number of proteins identified in this fraction, the Coomassie staining pattern is mostly a result of the presence of collagen type VI alpha 2 (COL6A2), which contributed to more than 64% of the protein amount in all three selected fractions. There were also fractions in which a single protein dominated the total protein amount, e.g. row 16, column 4, in which fraction protein vimentin (VIM) contributed to more than 85% of the protein amount. In two other fractions (row 25, column 4-5), which corresponded to an intense spot with high protein amount, the protein content was comprised of several proteins with similar abundances: lamin beta 2 (LMNB2), collagen type VI alpha 3 (COL6A3) and POTE ankyrin domain family member I (POTEI). Some fractions contained no distinct Coomassie stained spots (e.g. row 25, column 16-17). Nevertheless, despite the absence of a staining pattern, the protein amount was comprised of several proteins but dominated by a single protein – Elongation factor Tu, mitochondrial (TUFM) with more than 68% of the protein amount. Evidently the different fractions of the 2DE gel can contain different proteins with varying amounts. In order to further evaluate performance of the 2DE separation approach, the numbers and the distribution of identified proteins per fraction was examined (Figure 3).

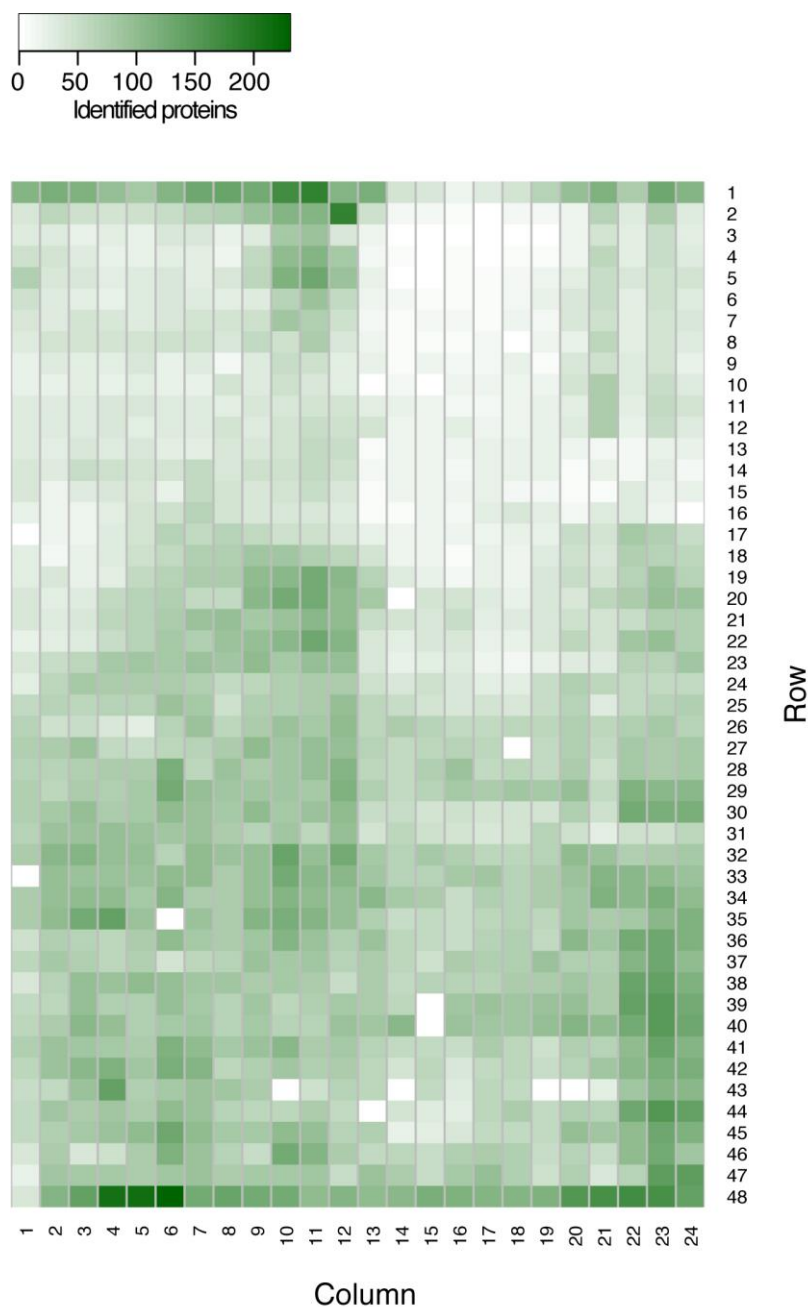


Figure 3. Number of identified proteins per 2DE gel fraction.

The numbers of identified proteins per 2DE slice were very different throughout the 2DE gel. One distinct pattern was the higher number of identified proteins in the first and the last row, i.e. the beginning and the end of the 2DE gel. The last row of the 2DE gel contained the running front of the separation in the second dimension (Figure 2, A). As the running front is the boundary of protein separation it will contain a large number of small proteins and possibly protein fragments resulting from the sample preparation. This can explain the higher number of identified proteins in the last row. The lower rows of the most basic end of the gel (columns 22-24) were also characterized by a large number of identified proteins.

On the contrary, the upper rows (2-17) of columns 14 to 19 showed lower number of identified proteins (Figure 3). Overall there was a wide range of identified proteins per gel slice. This can be seen from the distribution of the protein identifications per 2DE fraction (Figure 4).

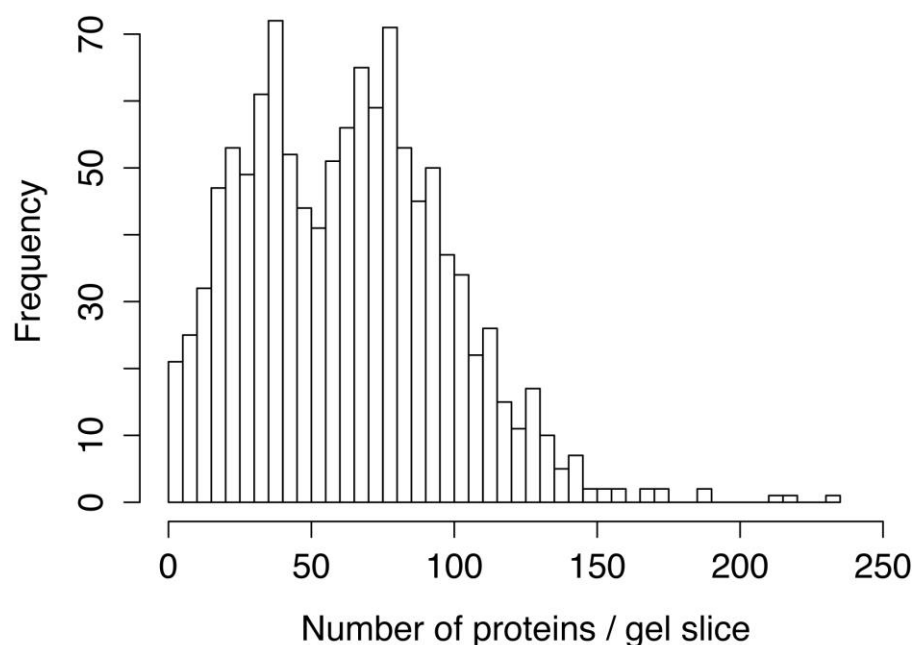


Figure 4. Histogram of the number of identified proteins per 2DE gel slice.

The histogram shown in Figure 4 reveals two peaks of the number of identified proteins per 2DE gel slice – at around 40 and 80 proteins per gel slice. Very few gel slices produced more than 150 identified proteins. When all gel slices were analyzed together, the resulting number of identified proteins was 2816. This number is much lower compared to the expected number of identified proteins from the whole 2DE gel, using simple addition of numbers of identified proteins (Figure 4) and expecting moderate degree of redundancy. This indicates a large degree of redundancy of the identified proteins in separate 2DE gel slices.

The distribution of identified proteins across the 2DE gel shown in Figure 3 depicts all proteins identified in each slice and these numbers would also include modified versions of proteins identified in other fractions and/or carryover of protein from other slices. The presence of carryover or modified proteins will result in the identification of same protein in a number of additional slices explaining the high degree of redundancy described above. An example of such observation is the distribution of the normalized intensity and ratio H/L of protein ATP synthase subunit O, mitochondrial (ATP5O) and the pattern of fractions in which the protein was identified (Figure 5).

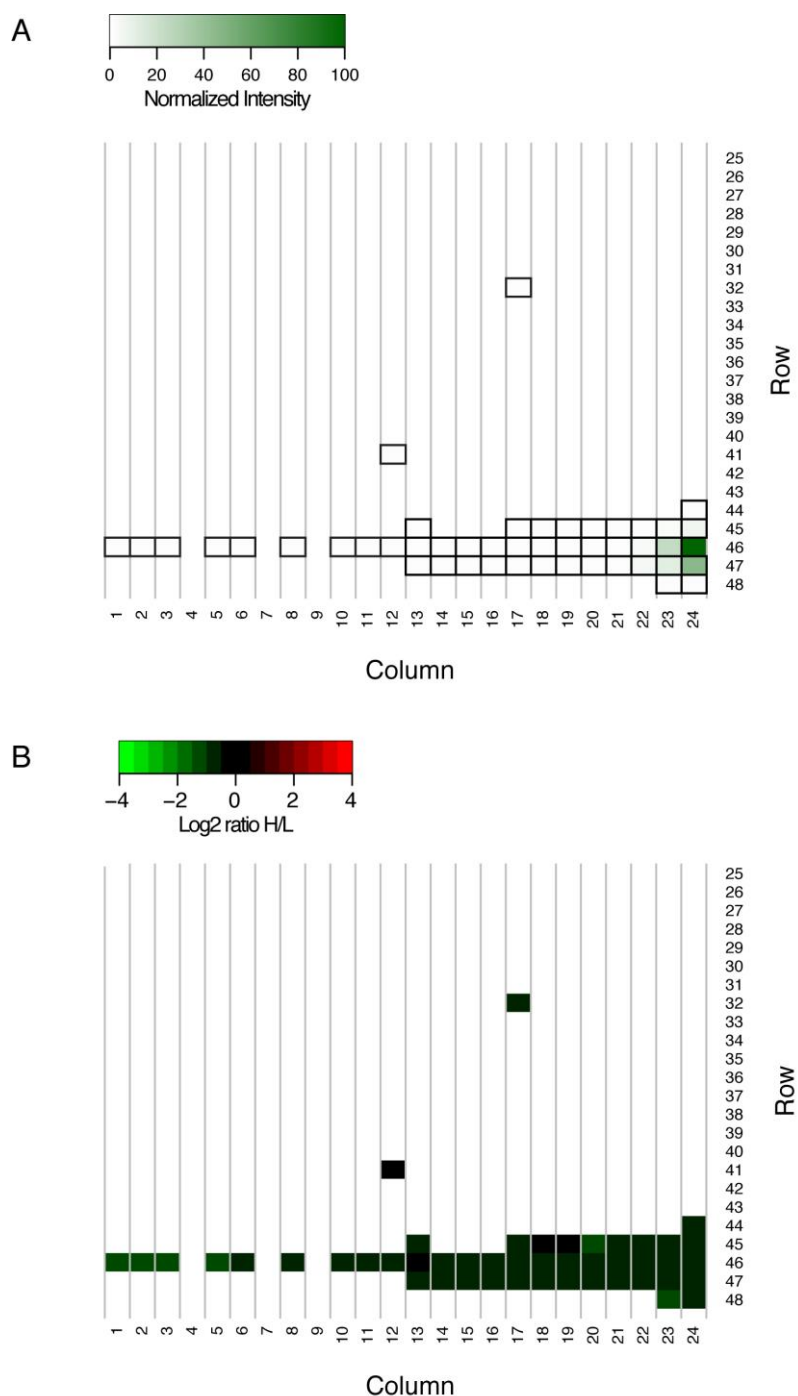


Figure 5. Heatmap of the normalized intensity (A) and ratio H/L (B) of protein ATP5O. The protein intensity in each slice was normalized to the highest value from all slices where the protein was identified. The presence of an identification of the protein in a 2DE fraction is indicated by bold borders of the fractions (A) or by presence of ratio H/L color-coding (B).

Protein ATP5O was identified with highest intensity (100%) in row 46, column 24. It had lower intensity in the bordering slices and beyond them the intensity was close to 0%, relative to the maximum. Overall, protein ATP5O was identified in 47 fractions and all of them showed very consistent log₂ ratio H/L (around -0.8) (Figure 5, B). This indicates that

the protein was either imperfectly focused in the first dimension resulting, the presence of modified versions of the protein and/or protein isoforms or a combination of both factors. Overall, the identifications of a protein in multiple fractions can be transformed into normalized intensity values relative to the highest intensity for all slices. Values in the lower percentages would indicate carryover or the presence of low abundant modified versions of the protein and/or protein isoforms. In order to examine the number of fractions in which a protein was identified, normalization to the highest intensity was carried out for all identified proteins. Overall, there were 72500 different combinations of identified proteins and gel slices. The distribution of the normalized intensities of those identifications is shown in Figure 6.

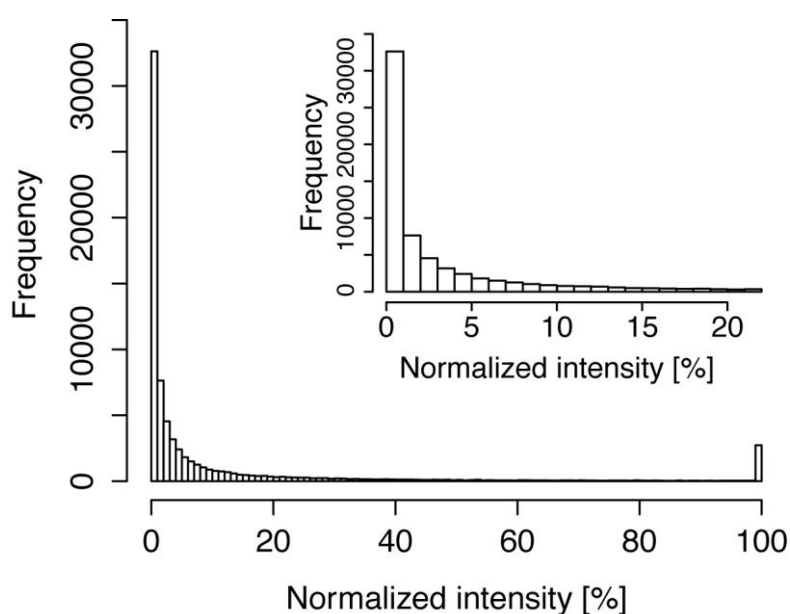


Figure 6. Distribution of normalized proteins intensities over the 2DE gel. The intensity for each identified protein in each gel slice was normalized to the highest in all slices where the protein was identified. Inset shows the same distribution but limited to intensity values between 0 and 25%.

The distribution of the normalized intensities (Figure 8) shows two distinct peaks – one at 100% and one at 0%. Out of around 72500 redundant protein identifications in different slices, around 32000 protein identifications had intensities between 0 and 1% of the highest intensity and around 59000 between 0 and 10%. These values indicate that most of the protein identifications in each slice were carryover from proteins located in other slices or modified versions of those proteins or protein isoforms. The second peak of the normalized intensity histogram, at 100%, is a direct representation of the number of proteins identified from the whole 2DE as each protein has normalized intensity equal to 100%. Filtering the identified proteins per 2DE gel slice (Figure 3 and Figure 4) based on 100% intensity can

be used to find the actual number of newly identified proteins from all fractions of the 2DE gel. The filtered results are shown in Figure 7.

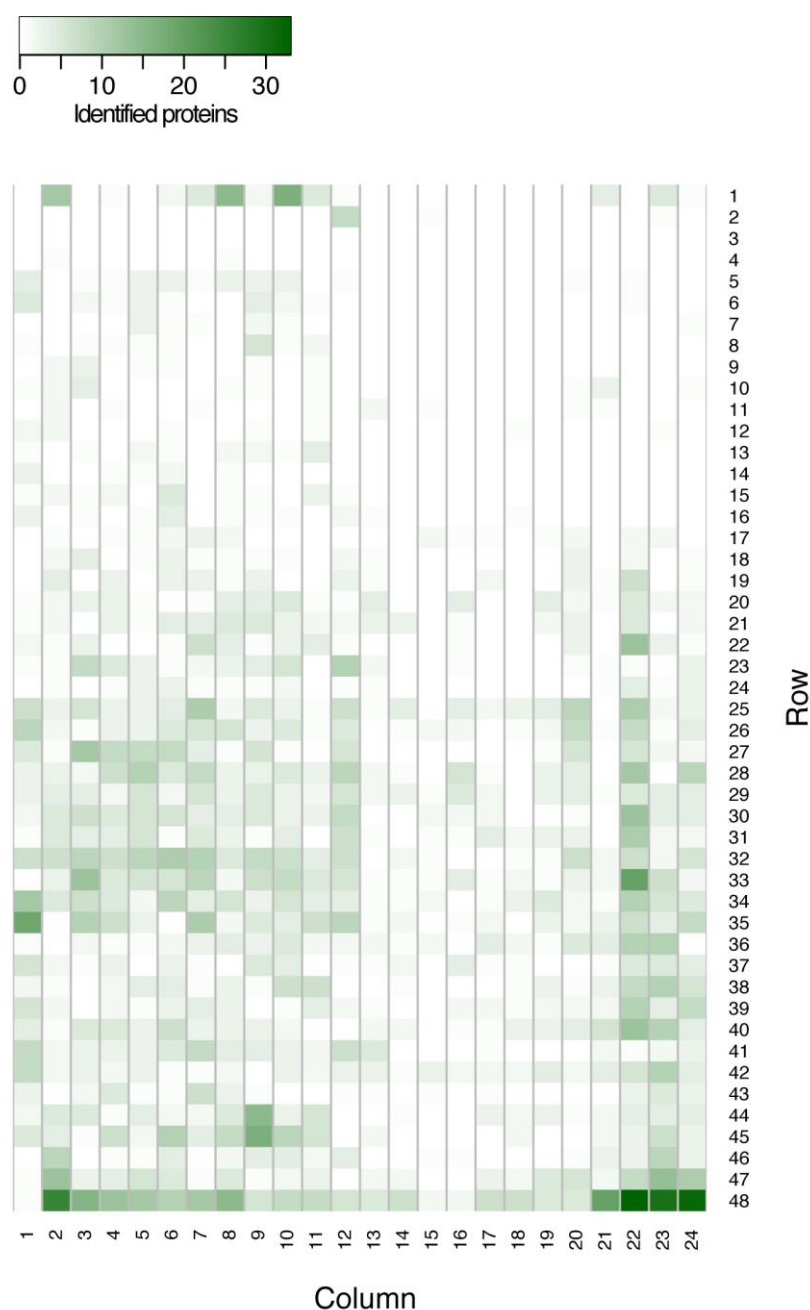


Figure 7. Identified proteins per 2DE gel slice showing maximum (100%) intensity.

The distribution of identified proteins per gel slice changed dramatically before and after proteins were filtered based on 100% intensity (compare Figure 3 and Figure 7). The filtered results showed that there are regions of the 2DE gel that are basically devoid of new proteins columns 13-20 and rows 1-15 of the 2DE gel (Figure 7). A more detailed description of the identification rate per column or row is shown in Figure 8.

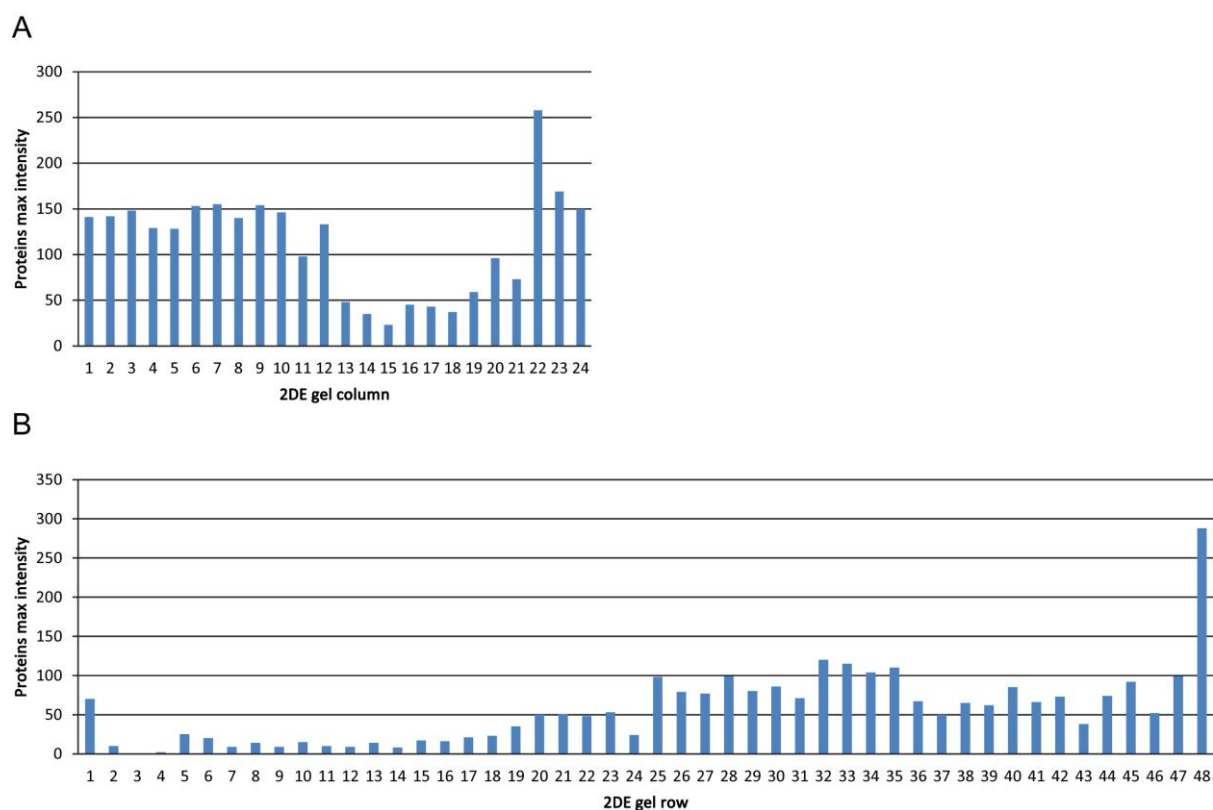


Figure 8. Numbers of proteins showing 100% intensity in each column (A) or in each row (B) of the 2DE gel.

Examining the number of identified proteins per column reveals that the rate of newly identified proteins was rather constant for columns 1 to 10, with around 150 newly identified proteins per column (Figure 8, A). Columns 13 to 18 had the lowest number of new proteins, while column 22 contained the highest number of identified proteins. The newly identified proteins per row (Figure 8, B) also showed a very distinct distribution. Proteins were practically absent from rows 2 to 4 and the rate of newly identified proteins was very low, less than 25 for rows 5 to 18. In contrast to those values, there was a very large amount of identified proteins in the last row – 288 proteins. Based on these values, it is clear that there are discrepancies between the identification rates of new proteins in the different regions of the 2DE gel. The results from Figure 8 are based on slices that contain at least one protein with 100% normalized intensity. At the same time, there were a number of fractions that contained zero proteins with 100% normalized intensity (Figure 7). In order to evaluate the rate of identification of protein showing 100% normalized intensity in a fraction, their overall distribution was plotted (Figure 9).

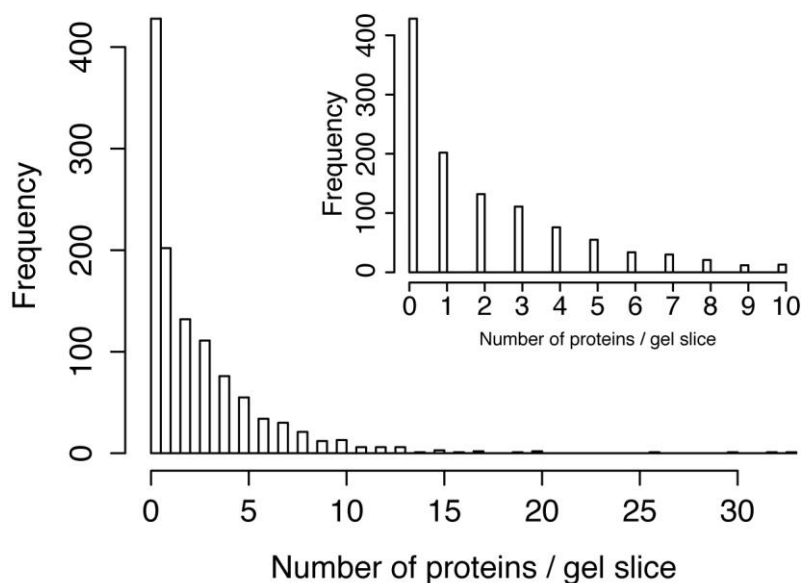


Figure 9. Distribution of identified proteins showing 100% normalized intensity per 2DE fraction.

Inset shows the same distribution but values are limited to the 0 to 10 range.

The distribution of the proteins with 100% normalized intensity was very different from the distribution of all proteins identified per gel slice (Figure 4). The analysis revealed that the distribution had a peak at 0 and skews to the right. The majority of the slices showed 0 to 5 (new) protein identifications and very few proteins had more than 15 identified proteins. Most importantly 434 (more than one third of all slices of the 2DE gel) had zero proteins showing 100% normalized intensity, showing that these fractions contained mostly protein carryover from other fractions and/or modified proteins or isoforms and did not contribute to new protein identifications. These results were consistent with the not so high number of identified proteins from the whole 2DE gel. Overall, the separation of the complex protein mixture using 2DE, fractionation of the whole 2DE gel and analyzing the protein content of each fraction did not result in a large number of identified proteins. Important factors associated with this result were the high level of redundancy of protein identification between different fractions (Figure 6) and the high number of fractions that do not contribute to new protein identifications (Figure 9).

3.3 Comparison of 1DE and 2DE approaches for proteome analysis

For both 1DE and 2DE analyses of biological replicate 1, cells were lysed and proteins were extracted in the same lysis solution prior to protein separation. This provides the unique opportunity to directly compare the physico-chemical characteristics of the proteins

identified in the two distinct separation approaches since such differences must be introduced during the protein separation process. There are two protein characteristics that have been identified to influence the identification of proteins using the 2DE approach, the protein pI and hydrophobicity (Rabilloud, Vaezzadeh et al. 2009).

3.3.1 Comparison of the pIs of proteins identified using 1DE and 2DE

In order to examine potential effect of the pI on the detectability of proteins in the 1DE and 2DE method, the pI of all identified proteins in the 1DE and 2DE experiment were calculated and their distributions were compared. The results are shown in Figure 10.

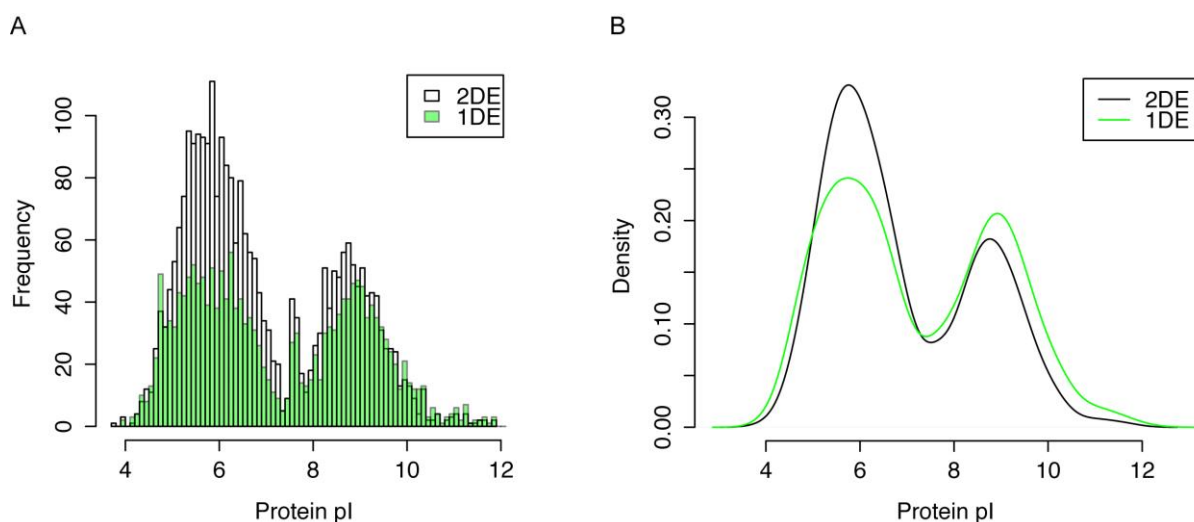


Figure 10. Distribution the pI of proteins identified using the 1DE and 2DE approaches. Histogram (A) and density plot (B) of pI of the identified proteins.

The distributions of the pIs show similar binomial shape with peaks at around pI 6 (acidic) and 8.8 to 9 (basic). However, the proportions of both peaks were different in the 1DE and 2DE experiment (Figure 10, B). The more acidic (pI 6) peak of the pIs was much higher compared to the basic peak for the 2DE analysis (Figure 10, B). In contrast, the acidic and the basic pI peaks had a similar height in the 1DE analysis (Figure 10, B). In addition, the distribution of pIs was shifted slightly to the right for the 1DE analysis compared to the 2DE analysis. These results indicate that, in the current analysis, 2DE readily identified more proteins with acidic pIs compared to proteins with basic pIs. This difference was not pronounced for the 1DE analysis.

3.3.2 Comparison of the hydrophobicity of proteins identified using 1DE and 2DE

Next, the impact of protein hydrophobicity on protein detectability was evaluated. Protein hydrophobicity was inferred from the average GRAVY score (Kyte and Doolittle 1982), calculated by dividing the sum of the hydropathy value of each residues by the protein length. Higher average GRAVY score indicates higher protein hydrophobicity (Kyte and Doolittle 1982). The comparison of the distributions of the average protein GRAVY scores of the proteins identified in the 1DE and 2DE approaches is shown in Figure 11.

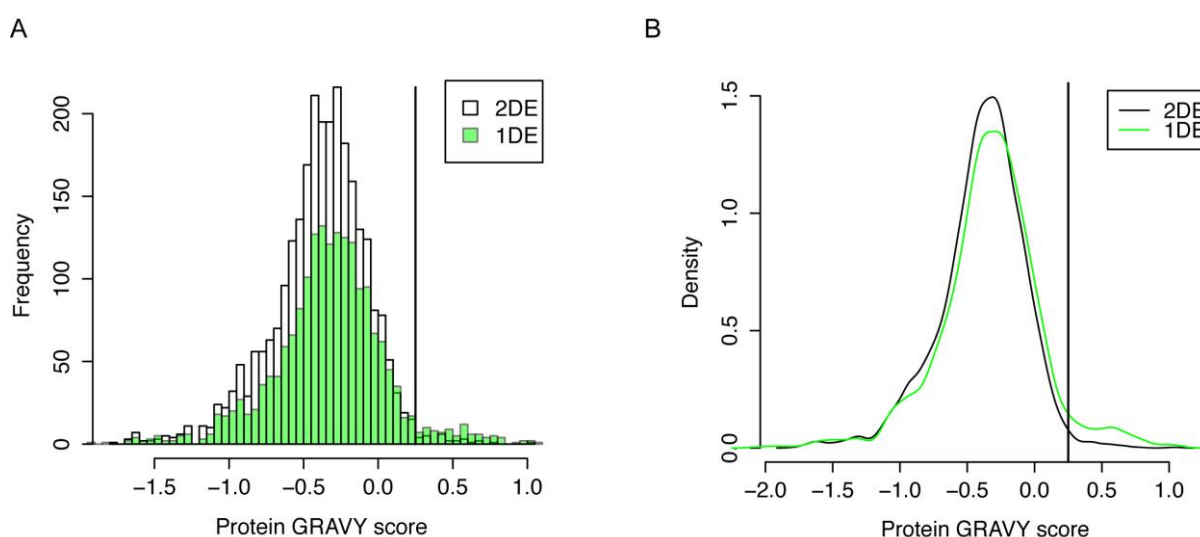


Figure 11. Distribution of GRAVY scores of all proteins identified in the 1DE and 2DE analysis.

Overall, the distributions of the average GRAVY scores of the proteins identified in the 1DE and 2DE analysis showed similar shape while the results from 1DE analysis had higher proportion of proteins with GRAVY scores above 0.25 (vertical line, Figure 11). This would indicate that there are proportionally more proteins with higher average GRAVY score identified when using 1DE compared to 2DE. The average GRAVY score can also be misleading regarding the nature of the protein (e.g. cytoplasmic vs membrane proteins) as hydrophobic transmembrane domains can also be balanced by highly polar regions, thus decreasing the average GRAVY value. In order to reveal the nature of the proteins showing higher proportion in the 1DE analysis, the list of all identified proteins having GRAVY scores above 0.25 compared between the 1DE and 2DE analyses. The comparison revealed that there were 3 proteins identified by 2DE but not with 1DE. At the same time there were 66 proteins with average GRAVY score above 0.25 that were identified by 1DE

but not with 2DE. A selected set of those proteins, confidently identified by at least 2 peptides is shown in Table 5.

Table 5. Properties of proteins identified only with 1DE and not 2DE, having GRAVY score larger than 0.25.

Uniprot ID	Gravy Score	Gene Name	Proteins structure*	Peptides
Q9Y3E0	1.086956522	GOLT1B	Multi-pass	2
Q8TBQ9	1.0375	TMEM167A	Single-pass type I	2
P61803	0.824778761	DAD1	Multi-pass	3
Q96S97	0.770186335	MYADM	Multi-pass	3
Q04941	0.765131579	PLP2	Multi-pass	2
P03905	0.727233115	MT-ND4	Multi-pass	3
P33947	0.721698113	KDEL2	Multi-pass	4
O75352	0.704453441	MPDU1	Multi-pass	2
P60033	0.68940678	CD81	Multi-pass	6
P03886	0.682075472	MT-ND1	Multi-pass	2
Q06055-2	0.668686869	ATP5G2	Multi-pass	2
D6RH22	0.639766082	OSTC	Multi-pass	2
O14735	0.612676056	CDIPT	Multi-pass	4
Q96AM1	0.605830904	MRGPRF	Multi-pass	2
Q08722	0.599380805	CD47	Multi-pass	3
Q96QD8	0.598616601	SLC38A2	Multi-pass	2
Q9P2X0-2	0.590163934	DPM3	Multi-pass	2
P03915	0.588888889	MT-ND5	Multi-pass	4
Q53TN4	0.571328671	CYBRD1	Multi-pass	2
B7ZAQ6	0.556263736	GPR89A	Multi-pass	3
O60831	0.55505618	PRAF2	Multi-pass	3
Q8IXM6	0.549236641	NRM	Multi-pass	2
Q8WWI5	0.545509893	SLC44A1	Multi-pass	4
P11166	0.534146341	SLC2A1	Multi-pass	2
Q9NRX5	0.486313466	SERINC1	Multi-pass	2
B4DK36	0.486027397	LPAR1	Membrane	3
O95197-3	0.473728814	RTN3	Multi-pass	5
Q8TB61	0.437268519	SLC35B2	Multi-pass	4
Q01628	0.381203008	IFITM3	Single-pass type II	2
Q8TCT9	0.363660477	HM13	Multi-pass	7
O43657	0.359183673	TSPAN6	Multi-pass	2
O60888-2	0.351515152	CUTA	Membrane	2
Q9HC07	0.327160494	TMEM165	Multi-pass	3

Q96N66	0.306355932	MBOAT7	Multi-pass	2
P51636	0.301234568	CAV2	Membrane	3

* single-pass: membrane protein spanning the membrane once; multi-pass: membrane protein spanning the membrane more than once; membrane: protein which is membrane-bound or membrane-associated (definitions from Uniprot).

The results from Table 5 show that the proteins identified only by 1DE, by at least 2 peptides and having GRAVY score above 0.25 are only hydrophobic single or multi-pass membrane proteins. The observation that these proteins have lower density (proportion) in the 2DE analysis shows that the 1DE analysis more readily identifies hydrophobic membrane proteins.

3.3.3 Comparison of the number of identified protein isoforms using 1DE and 2DE

In addition to comparing physico-chemical properties of the proteins such as pI and hydrophobicity, the numbers of identified isoforms using either 1DE or 2DE method were also compared. This was done because the identification of protein isoforms is a particularly difficult task in proteomics due to the subtle sequence differences between protein isoforms (Stastna and Van Eyk 2012). In this regard 2DE has been shown to be beneficial for the separation and identification of protein isoforms (Raikos, Hansen et al. 2006, Thiede, Koehler et al. 2013). For this comparison, the proteins identified in the CPC-2D/CPC-3D second biological replicate (1DE analysis) were included in order to add a reference for the comparison of the numbers of identified isoforms.

There are a number of considerations regarding the comparison of the numbers of protein isoforms. First, the possibility to identify an isoform (distinguish it from another isoform) should be evaluated. If two proteins have a very large degree of similarity, only one or a few tryptic peptides unique to each isoform will exist, making the mass spectrometric identification of both protein species in a complex protein mixture nearly impossible. One straightforward approach to evaluate how distinguishable separate isoforms are, is to perform sequence alignment of isoforms and to examine the similarity scores. Two isoforms need to differ at least by one tryptic peptide in order to be distinguishable using shotgun proteomic approaches. Second, once identified, the confidence in the identification of an isoform should be taken into account. If a protein isoform is identified using a single peptide, this identification would not be of high confidence (Carr, Aebersold et al. 2004). Third, secondary information, which can aid the (confident) identification of a protein isoforms, should be considered. Such information can be an agreement between the MW

and/or the pI of the identified protein with the region of the 1DE/2DE gel, where the protein was identified. In the comparison of the numbers of identified proteins between the 1DE and the 2DE approaches, all three points should be considered. For the comparison, all protein isoforms, identified in the two 1DE analyses of biological replicates and the 2DE analysis, were combined and sequence alignment of the isoforms was performed. Next the sum of razor peptides (Cox and Mann 2008) and unique peptides for each isoform was extracted from the MaxQuant result files. Razor peptides are peptides that are shared between two proteins while the proteins are distinguished by at least one unique peptide. Since it is not possible to determine whether the shared peptides come from one protein or another, the peptides are associated with the protein that is identified by most peptides in total. This is based on the principles of Occam's razor i.e. the presence of one of the proteins would be the simplest explanation for the existence of the observed tryptic peptides (Cox and Mann 2008). Lastly, the probability of false positive identification, reported as posterior error probability (PEP) by the MaxQuant software, was calculated. MaxQuant reports the PEP of a protein based on all identified peptides in all experiments. Since experiment specific protein PEP was needed, it was calculated as the product of the PEP of the peptides that map to it. In order to conform to the Occam's razor the protein PEP was calculated from the PEP of the razor and unique peptides.

Together in all three separate analyses, 1DE analysis of CPC-2D/CPC-3D biological replicate 1 (1DE-r1), 1DE analysis of CPC-2D/CPC-3D biological replicate 2 (1DE-r2) and the 2DE analysis of CPC-2D/CPC-3D biological replicate 1 (2DE) resulted in the identification of a total of 199 isoforms from 94 genes (Appendix 1). For one gene, PLEC, four isoforms were identified and for seven genes (TNC, POSTN, MACF1, LMNA, IMMT, FN1 and DNAJB) three isoforms per gene were identified. Overall, the sequence alignment analysis of the isoforms showed high sequence identity (Appendix 1). Based on sequence alignment with gap penalty, which includes gaps in the final calculation of the sequence identity, the isoforms from 31 genes had more than 90% identity while the isoforms of 77 genes showed more than 50% identity (Appendix 1). For the isoform sequence similarity analysis, it might be more beneficial to calculate the sequence similarity without considering gaps. Interestingly, sequence alignment without gap penalty resulted in 100% identity for 39 gene products. At the same time no proteins had 100% identity when alignment was performed using gap penalty. Additionally, the isoforms of 86 genes showed more than 90% sequence identity without gap penalty compared to 31 genes with gap penalty. When considering isoforms with more than 50% identity: 77 genes had more than 50% identity with gap penalty, compared to 102 genes when no gap penalty was applied. The differences in the alignment approaches become evident when the sequence alignment is visualized (Figure 12).

A

```

sp|P17096|HMGA1_HUMAN      1  MSESSKSSQPLASKQEKDGTEKRGRGRPRKQPPVSPGTALVGSQKEPSEVPTPKRPRGR      60
sp|P17096-2|HMGA1_HUMAN   1  MSESSKSSQPLASKQEKDGTEKRGRGRPRKQPPK-----EPSEVPTPKRPRGR      49
*****
*****

sp|P17096|HMGA1_HUMAN      61  PKGSKNKGAAKTRKTTTTPGRKPRGRPKKLEKEEEEEGISQESSEEEQ      107
sp|P17096-2|HMGA1_HUMAN   50  PKGSKNKGAAKTRKTTTTPGRKPRGRPKKLEKEEEEEGISQESSEEEQ      96
*****
*****

```

B

```

sp|Q16774-2|KGUA_HUMAN     1  MLRRPLAGLAAAALGRAPPDGMSGPRPVVLSPGSGAGKSTLLKRLQEHSGIFGFSVSH      60
sp|Q16774|KGUA_HUMAN      1  -----MSGPRPVVLSGPGAGKSTLLKRLQEHSGIFGFSVSH      39
*****
*****

sp|Q16774-2|KGUA_HUMAN     61  TRNPRPGEENGKDYYFVTREVMQRDIAAGDFIEHAEFSGNLYGTSKVAVQAVQAMNRICV      120
sp|Q16774|KGUA_HUMAN      40  TRNPRPGEENGKDYYFVTREVMQRDIAAGDFIEHAEFSGNLYGTSKVAVQAVQAMNRICV      99
*****
*****

sp|Q16774-2|KGUA_HUMAN     121 LDVDLQGVRNIKATDLRPIYISVQPPSLHVLEQRLRQNTETEESLVKRLAAQADMESS      180
sp|Q16774|KGUA_HUMAN      100 LDVDLQGVRNIKATDLRPIYISVQPPSLHVLEQRLRQNTETEESLVKRLAAQADMESS      159
*****
*****

sp|Q16774-2|KGUA_HUMAN     181 KEPGLFDVVIINDSLDQAYAELKEALSEEIKKAQRTGA      218
sp|Q16774|KGUA_HUMAN      160 KEPGLFDVVIINDSLDQAYAELKEALSEEIKKAQRTGA      197
*****
*****

```

Figure 12. Sequence alignment of identified protein isoforms from the gene HMGA1 (A) and KGUA (B).

A) For the two protein isoforms of HMGA1, the alignment identity was 88.78% with gap penalty and 98.96% without gap penalty. **B)** For the two isoforms of KGUA, the alignment identity was 90.36% with gap penalty and 100% without gap penalty. Bold sequences indicate peptides shared between both isoforms, bold red sequence indicates peptides unique to an isoform.

The alignment of the two isoforms of gene HMGA1 (Figure 12, A), show a subtle difference between the results of the two sequence similarity calculation methods (with and without gap penalty). There is a defined gap within the alignment and only peptides that span this gap in one of the isoforms can be used to infer the presence of the isoform. Peptides mapping to other regions, identical in both sequences, cannot be used to distinguish one isoform from the other. The alignment of the two isoforms of gene KGUA (Figure 12, B) shows an interesting example when the use and comparison of the identity of alignments with and without gap penalty can reveal the relationship between both isoforms. The sequence identity of the proteins is 90.36% with gap penalty while it is 100% without gap penalty. This means that the sequence of the isoform (Q16774) is completely contained within the sequence of the longer isoform (Q16774-2). In addition, the shorter isoform can only be identified if the N-terminal tryptic peptide is identified. This was the case for the two isoforms as peptides spanning the sequence corresponding to the N-terminus of the shorter isoform were identified, allowing for inference of both isoforms. Notably there were a number of other peptides, which were shared between both isoforms (Figure 12, B) and

again cannot be used to infer which isoform was expressed. In addition to sequence alignments, the number of identified isoforms and the number of identified peptides per isoform were evaluated (Table 6).

Table 6. Number of identified proteins, identified isoforms and number of peptides per isoform.

1DE-r1: 1DE analysis of first biological CPC-2D/CPC-3D replicate. 1DE-r2: 1DE analysis of second biological CPC-2D/CPC-3D replicate. Protein isoforms were ranked based on the number of razor and unique peptides mapping to each isoform (isoform with highest number of peptides having rank 1 and a decreasing rank corresponding to a decreasing number of peptides). The rank values are taken from Appendix 1. Peptide values indicate the number of peptides per protein isoform with rank >1.

	1DE-r1	1DE-r2 *	2DE
Total proteins	2176	4478	3374
Isoforms	98	159	136
Isoforms rank 1	74	93	86
Isoforms rank >1	24	66	50
1 peptide	20	52	41
2 peptides	2	9	3
3 peptides	0	1	2
>3 peptides	2	4	4

* The higher number of identified proteins in the 1DE-r2 analysis is a result of the inclusion of additional replicate analyses and the use of smaller RP-HPLC column packing material.

The highest number of isoforms was identified in the 1DE-r2 analysis, 159 isoforms. At the same time, the highest number of total identified proteins was also identified in this analysis (4478 proteins). The 2DE analysis resulted in the identification of 136 isoforms from 3374 proteins. These results indicate that the number of identified isoforms increases with an increase in the depth of the analysis. Next, the sum of razor and unique peptides per isoform was considered. Ranking isoforms based on the number of razor peptides that map to a protein allows for clear separation of the isoforms into two populations. This is because isoforms with rank 1 contain larger number of peptides than isoforms with rank >1 which contain mostly one or two peptides (Appendix 1, Razor and unique peptides, Rank). It is important to consider that the association of peptides to proteins (shared peptides designated as razor and mapped to one protein) is solely due to the protein inference and peptide to protein mapping approach implemented in MaxQuant. Peptides shared between two isoforms will be mapped to the isoform, which contains more peptides. While MaxQuant uses Occam's razor principles other approaches also exist for protein inference

and peptide – protein mapping (Li and Radivojac 2012). The results in Table 6 reveal that the identification of the majority of the isoforms that have a rank >1 is based on a single peptide. This was the case for 20 isoforms (out of 24, 1DE-r1), 52 isoforms (out of 66, 1DE-r2) and 41 (out of 50, 2DE). The finding that most of the identified isoforms were based on a single peptide shows that their acceptance as identified proteins should be done with caution. This is because single peptide identifications are generally of low confidence and have increased chances of false positive identification (Carr, Aebersold et al. 2004). In addition to numbers of peptides per protein, the probability of protein being a false positive identification (PEP value) was also calculated for all isoforms identified in each of the analyses. This was done by taking the product of the PEP of all razor and unique peptides that map to an isoform. The ranking of the isoforms based on the number of peptides that map to the protein was used to separate the isoforms into two groups (rank 1 and rank >1) and compare the distributions of their PEP values. The results of this comparison are shown in Figure 13.

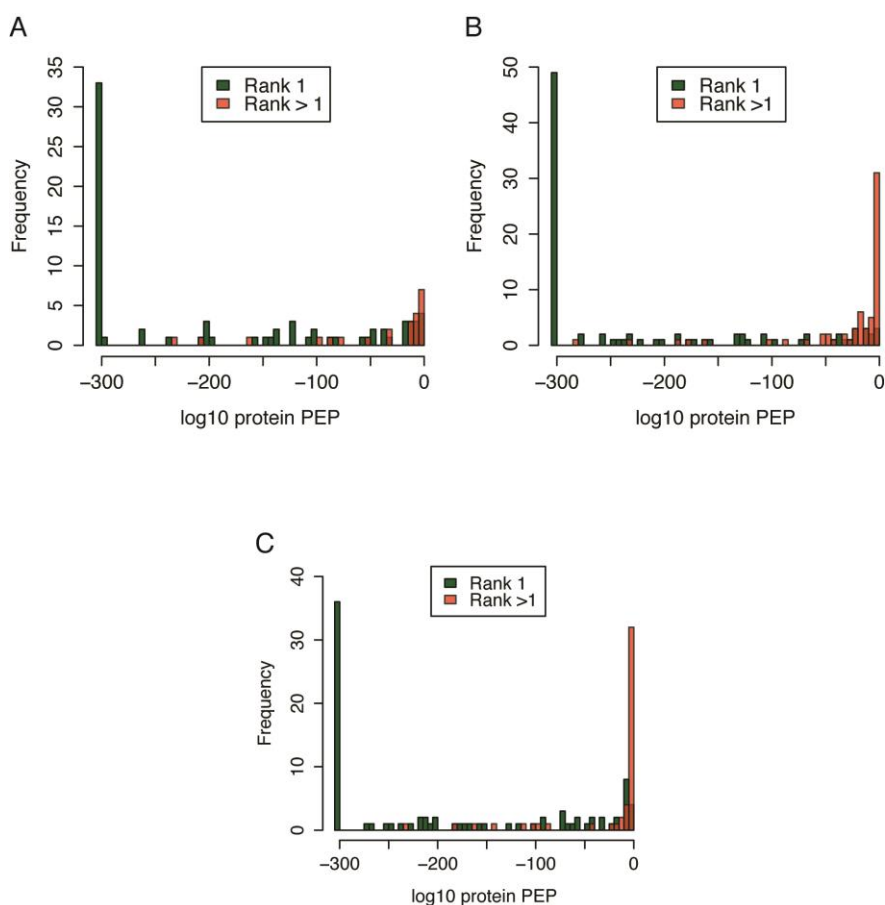


Figure 13. Distribution of protein PEP values from all isoforms identified in the 1DE analysis of first biological CPC-2D/CPC-3D replicate.

A) 1DE analysis of first biological CPC-2D/CPC-3D replicate. B) 1DE analysis of second biological CPC-2D/CPC-3D replicate (C) 2DE analysis of first biological CPC-2D/CPC-3D replicate.

A protein PEP value indicates the probability that a protein is a false positive identification (Cox and Mann 2008). A higher PEP value indicates higher probability that the protein identification is false. The distributions of the PEP values of the identified proteins show a clear trend – the protein isoforms with rank 1 (the protein isoform to which the razor peptides are mapped) have very low PEP values (< -300 on \log_{10} scale). The characteristic peak is present because values less than $1.58E-304$ are shown as zero. For this reason all PEP values that are equal to zero were replaced with the value of $1.58E-304$. On the contrary, isoforms with rank > 1 had very high PEP values (approaching 1), indicating that they are more likely to be false positives. A clear separation of the confident rank 1 and less confident rank > 1 isoforms was found in all three analyses (1DE-r1, 1DE-r2 and 2DE).

Overall, the analysis of the isoforms identified in all three experiments (1DE-r1, 1DE-r2 and 2DE) showed that protein isoforms share a large degree of similarity, which in many cases is 100 %, when gaps are not considered. The analysis also revealed that, generally one of

the two identified isoforms has a very low number of identified peptides, mostly single peptide identifications (Table 6), and a larger probability of the isoform being a false positive identification (Figure 13). Due to the lower confidence in the identification, it will be beneficial to incorporate secondary information such as the protein MW or pI. An evaluation of the identification of isoforms in the 1DE-r1 and the 2DE experiment revealed two examples in which the separation of the identified proteins increased the confidence of their identification. For this comparison, only 1DE-r1 and 2DE were considered since the same sample was used, and it was processed in the same way, decreasing the variation in identification due to sample processing. The first example is the pair of isoforms of collagen type VI alpha 2 (COL6A2) (Figure 14).

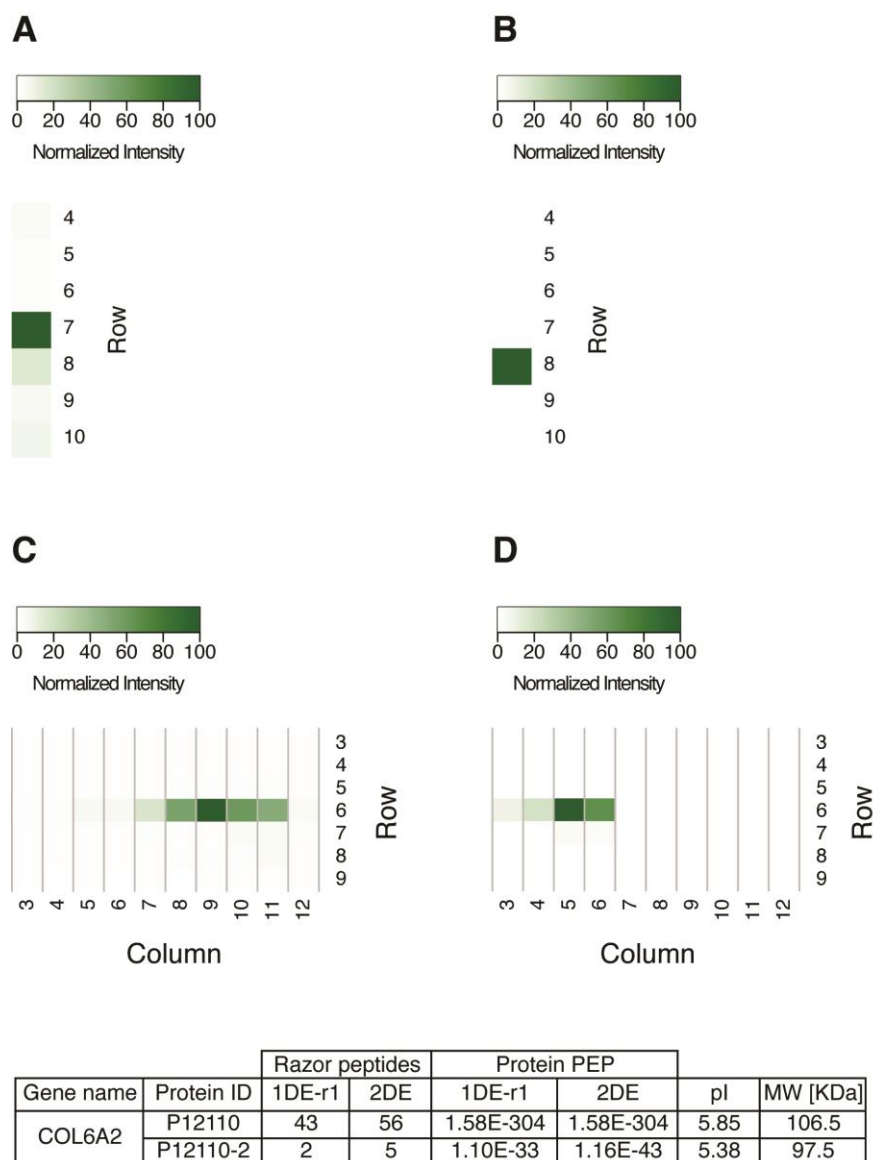


Figure 14. Intensity distribution of two isoforms of COL6A2.
Intensity distribution of isoform P12110 (A and C) and isoform P12110-2 (B and D). Isoforms were identified using 1DE (A and B) or using 2DE (C and D) separation approaches.

The two isoforms of COL6A2, (Uniprot IDs P12110 and P12110-2), differ in their predicted pI (5.85 / 5.38) and their MW (106.5 kDa / 97.5 kDa). Both isoforms were identified in both the 1DE-r1 and the 2DE analyses. The majority of the peptides, 43 (1DE-r1) and 56 (2DE) were mapped to the longer isoform (P12110) and only 2 peptides (1DE-r1) and 5 peptides (2DE) were mapped to the shorter isoform. Due to this, the protein PEP of the P12110-2 isoform was much higher (1.10E-33, 1DE-r1) compared to the PEP of the P12110 isoform (1.58E-304, 1DE-r1 and 2DE). This shows that the confidence in the identification of isoform P12110-2 is much lower compared to the one of P12110. The distribution of the intensity of isoforms P12110 in the 1DE-r1 analysis showed highest intensity in row 7 while isoform P12110-2 showed highest intensity in row 8 (Figure 14, A and B). This could be

due to the smaller MW of the P12110 isoform but the fact that both proteins are identified in bordering slices does not add significant confidence in the identification. The intensity pattern of both proteins on the 2DE gel was in agreement with differences in their pI. Isoform P12110 (predicted pI 5.85) had highest intensity in row 6 column 9 (Figure 14, C), corresponding to a pI of 6.1. Isoform P12110-2 (predicted pI 5.38) had highest intensity in the same row 6 but in column 5 (Figure 14, D), corresponding to a pI of 5.5. Both isoforms were well separated by IEF and the pI of the isoforms was in agreement with the slices in which they were identified. These results add more confidence in accepting the identification of not only the longer P12110 isoform but also of the shorter P12110-2 isoform. The evidence for the acceptance of both isoforms provided by the 1DE analysis was not as substantial as the one provided by the 2DE approach. In addition to the two isoforms of COL6A2, the two isoforms of PFN2 were examined (Figure 15).

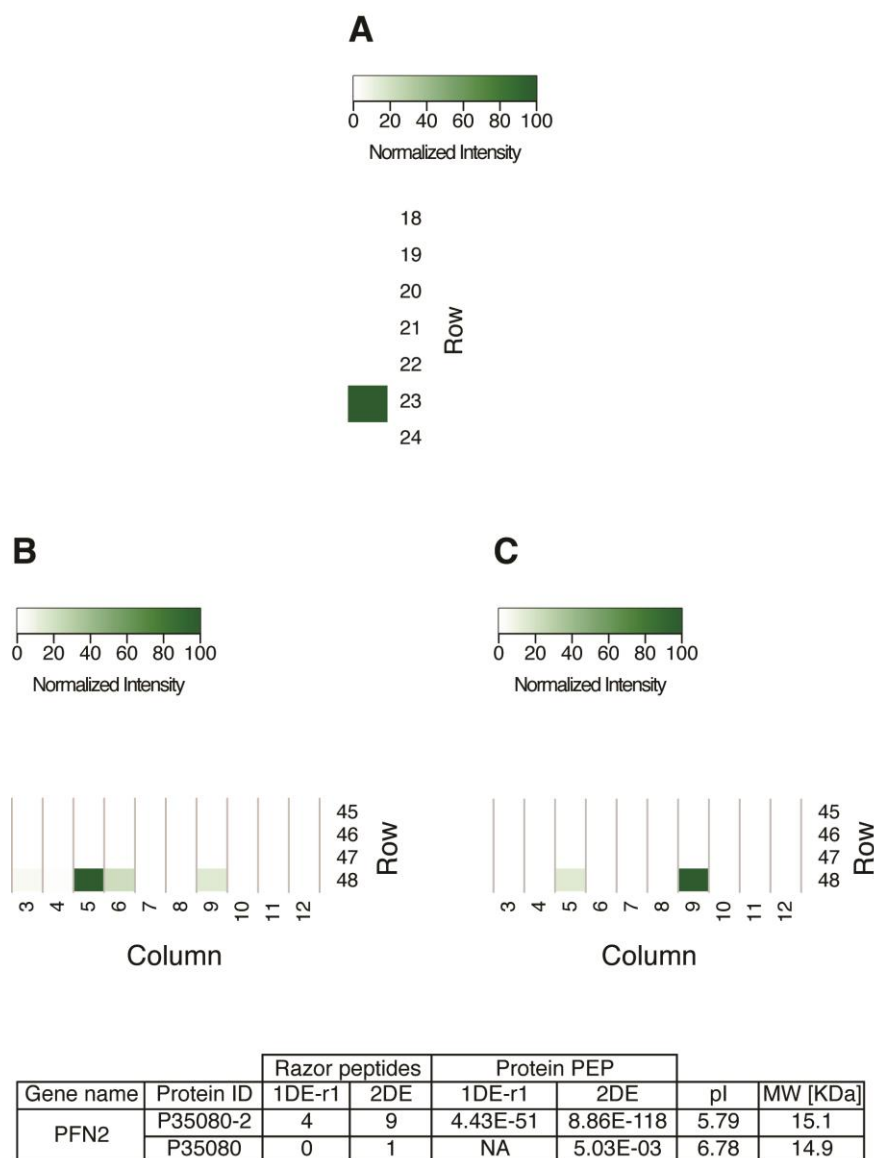


Figure 15. Intensity distribution of two isoforms of PFN2.

Intensity distribution of isoform P35080-2 of (A and B) and of isoform P35080 (C). Isoforms were identified using 1DE (A and B) or using 2DE (C) separation approaches.

The two isoforms of PFN2, (Uniprot IDs P35080 and P35080-2), differ in their predicted pI (5.79 / 6.78) and their MW (15.1 KDa /14.9 KDa). Only one isoform (P35080-2) was identified in 1DE-r1 analysis while both isoforms were identified in the 2DE analysis. The majority of the peptides, 4 (1DE-r1) and 9 (2DE) were mapped to isoform P35080-2 and only a single identified peptide was mapped to the P35080 isoform. Due to this the protein PEP of the P35080 isoform was very high (5.03E-03). This showed that the confidence in the identification of isoform P35080 is very low and points to the identification as a false positive. While the P35080 isoform was not identified in the 1DE-r1 analysis it was identified in the 2DE experiment and both isoforms had a distinct identification pattern. Isoform P35080-2 (predicted pI 5.79) had highest intensity in row 48 column 5 (Figure 15,

B), corresponding to pI of 5.5. Isoform P35080 (predicted pI 6.78) had highest intensity in the same row 48 but in column 9 (Figure 15, C), corresponding to pI of 6.1. Despite the observation that isoform P35080 had higher predicted pI than the one of the slice (6.78 vs 6.1), similarly for the 2DE analysis of the COL6A2 isoforms, both PFN2 isoforms were well separated in the pI dimension and the pI of the isoforms was in agreement with the slices in which they were identified. These results add more confidence in accepting the identification of isoform P35080, which is otherwise a very good candidate for a false positive identification.

The analysis of the COL6A2 and PFN2 indicate that the 2DE separation approach is capable of very efficient separation of protein species (e.g. isoforms), which, as exemplified above, can aid in their identification. In these terms the 2DE analysis outperforms the 1DE one.

3.3.4 Identification and quantification of protein species using 1DE and 2DE

It is interesting to examine whether performing 2DE would also provide an analytical advantage over 1DE in the respect of identification and quantification of other protein species besides. For this purpose the set of identified proteins in 1DE-r1 and 2DE was examined for the presence of several species of the same proteins. Again, only the analysis of 1DE-r1 and 2DE was considered as the protein sample and the sample processing was the same, decreasing the technical variation.

The examination of proteins identified in 1DE-r1 and 2DE revealed a number of proteins for which a number of protein species were present. In order to examine the protein species, protein coverage was shown using the position in the protein of the first amino acid of each identified peptide. The pI and the MW of each protein species were calculated using the protein sequence that was spanned by the identified peptides. The experimental pI and MW for each slice was calculated as the median of the proteins that show maximum intensity (Figure 7) in this row/column. Eight proteins that showed presence of multiple species were selected and then the separation/identification of those species was compared between the 1DE and the 2DE separation. The selected proteins were vimentin (VIM), Figure 16 and Figure 17, plastin 3 (PLS3), Figure 18 and Figure 19, T-complex protein 1 subunit alpha (TCP1), Figure 20 and Figure 21, lamin-B1 (LMNB1), Figure 22 and Figure 23, lamin A/C (LMNA), Figure 24 and Figure 25. In addition, three proteins showed characteristic horizontal spread in the 2DE approach and different ratio H/L ratios. These proteins were LMNA,

Figure 26 and Figure 27, F-actin-capping protein subunit beta (CAPZB), Figure 28 and Figure 29 and UPF0568 protein C14orf166 (C14orf166), Figure 30 and Figure 31. A comparison of the separation of the different species of each of those proteins in the 1DE and 2DE approaches was performed.

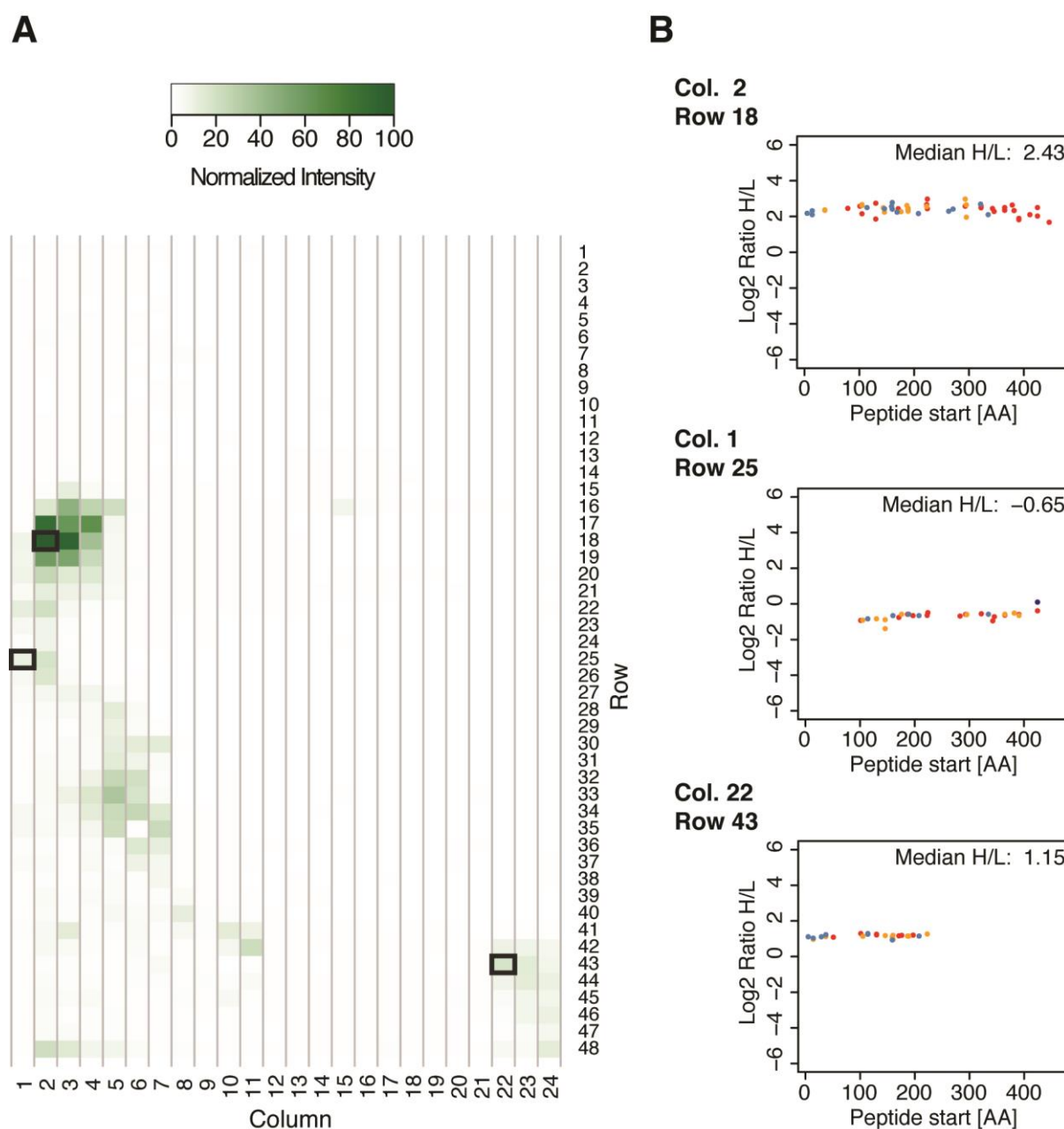


Figure 16. Identification of protein species of vimentin (VIM), 2DE experiment.

A) Distribution of protein intensity of VIM. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. **B)** Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Experimental pI and MW vs theoretical pI and MW of the protein species: column 2 / row 18: pI 5.3 / 5.05, MW 60 / 54; column 1 / row 25: pI 5.27 / 4.7, MW 49 / 43; column 22 / row 43: pI 8.41 / 6.51, MW 27 / 27. Color scheme based on peptide PEP (red: PEP $1E-11$, yellow: PEP<math><0.0001</math>, light blue: PEP <math><0.05</math>, dark blue: PEP>0.05).

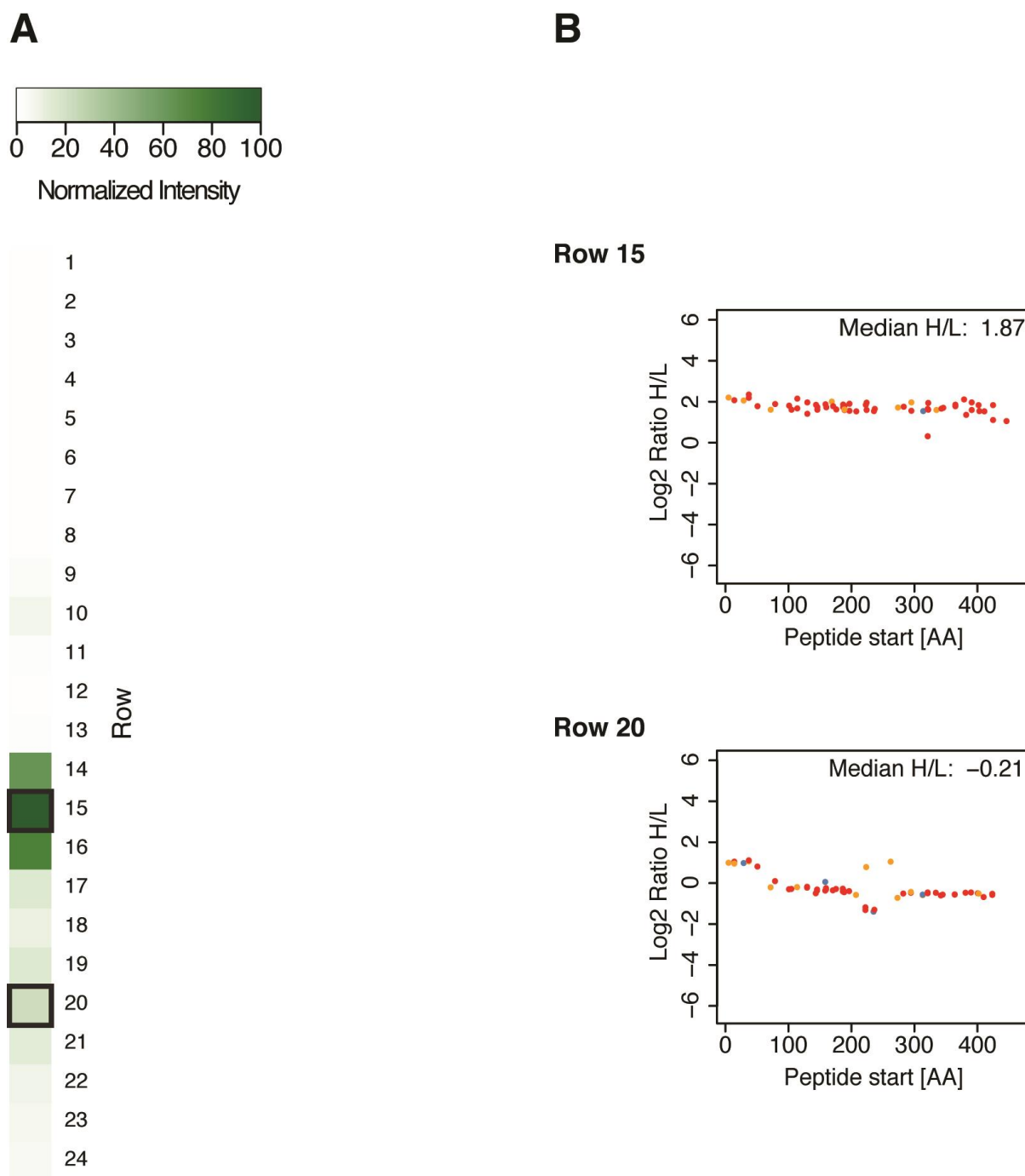


Figure 17. Identification of protein species of vimentin (VIM), 1DE experiment.

A) Distribution of protein MS intensity of VIM. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. **B)** Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP < 1E-11, yellow: PEP < 0.0001, light blue: PEP < 0.05, dark blue: PEP > 0.05).

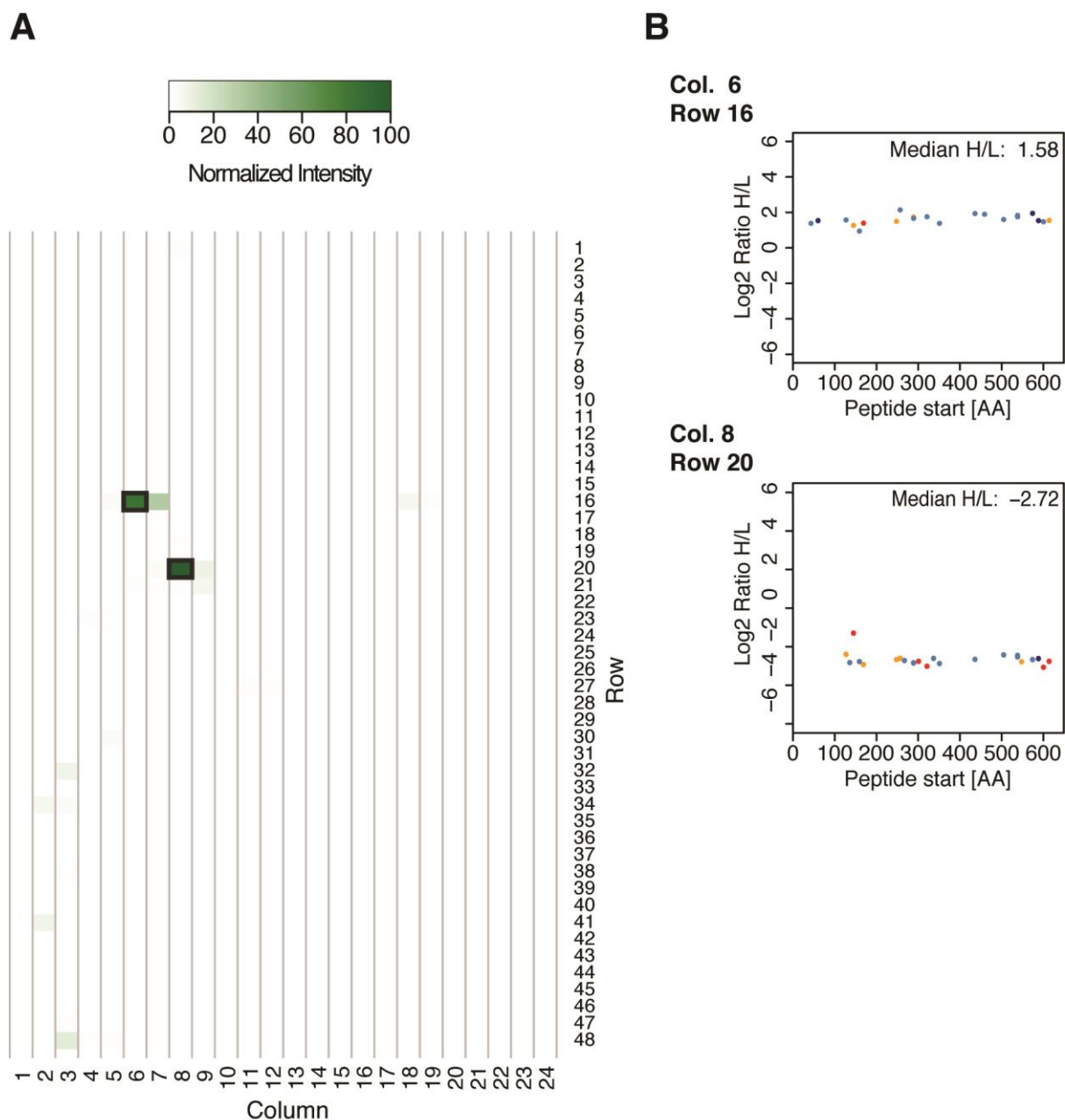


Figure 18. Identification of protein species of plastin 3 (PLS3), 2DE experiment.

A) Distribution of protein MS intensity of PLS3. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. **B)** Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Experimental pI and MW vs theoretical pI and MW of the protein species: column 6 / row 16: pI 5.62 / 5.83, MW 65 / 66; column 8 / row 20: pI 5.95 / 5.94, MW 59 / 56. Color scheme based on peptide PEP (red: PEP <math><1E-11</math>, yellow: PEP <math><0.0001</math>, light blue: PEP <math><0.05</math>, dark blue: PEP >math>>0.05</math>).

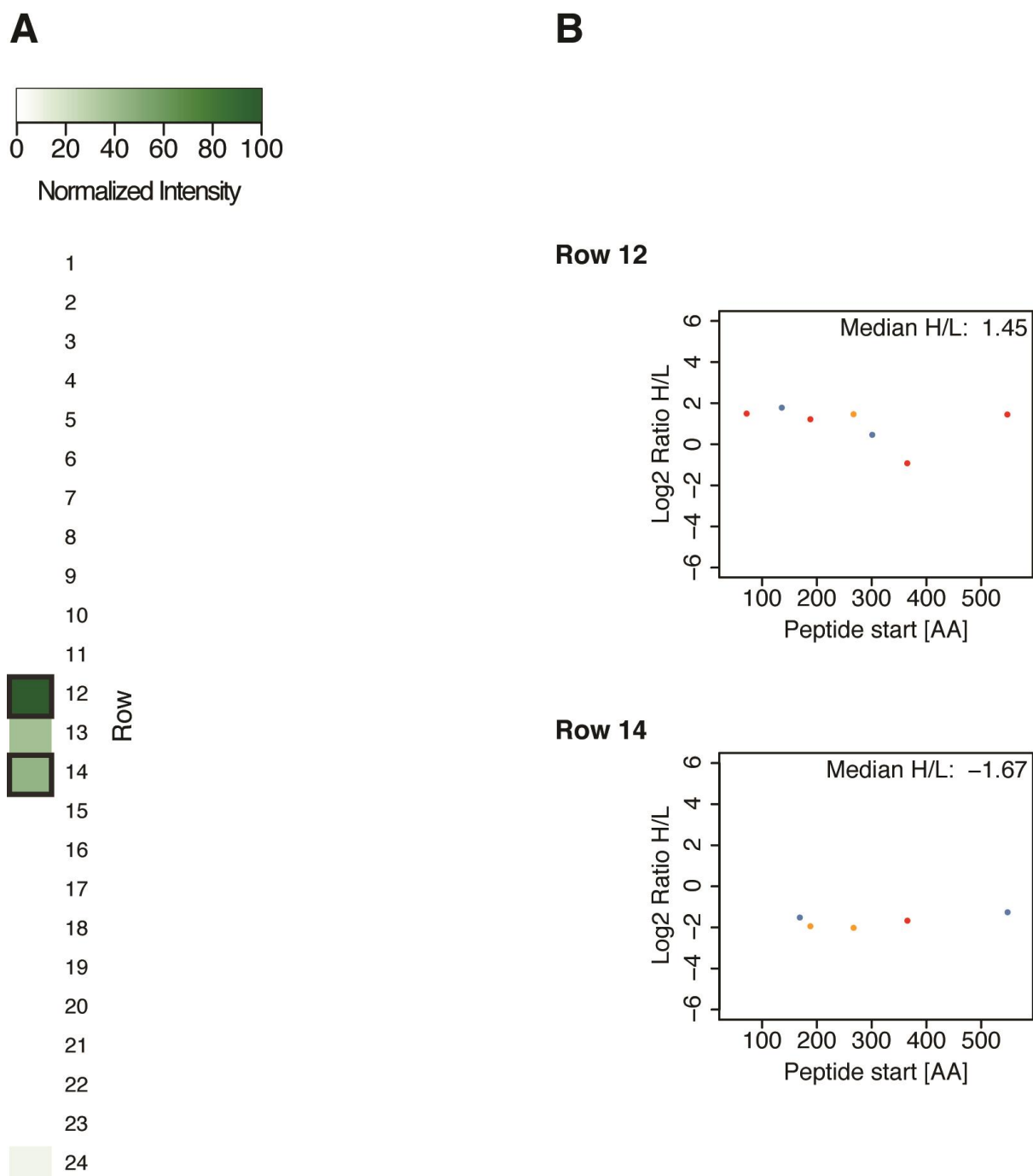


Figure 19. Identification of protein species of plastin 3 (PLS3), 1DE experiment.

A) Distribution of protein MS intensity of PLS3. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. B) Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP <math><1E-11</math>, yellow: PEP<math><0.0001</math>, light blue: PEP <math><0.05</math>, dark blue: PEP>0.05).

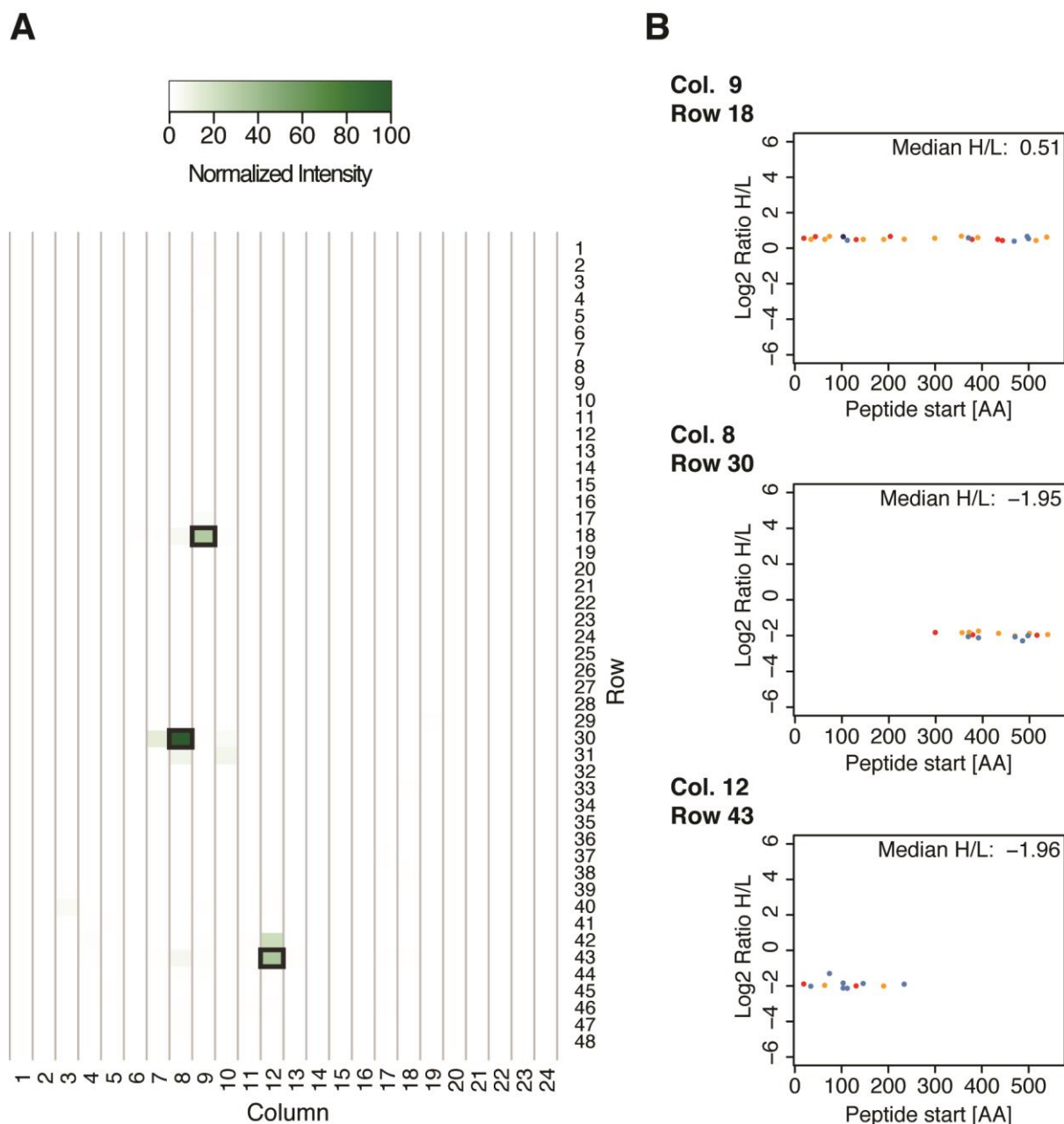


Figure 20. Identification of protein species of T-complex protein 1 subunit alpha (TCP1), 2DE experiment.

A) Distribution of protein MS intensity of TCP1. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. **B)** Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Experimental pI and MW vs theoretical pI and MW of the protein species: column 9 / row 18: pI 6.11 / 5.8, MW 60 / 60; column 8 / row 30: pI 5.9 / 5.67, MW 40 / 28; column 12 / row 43: pI 6.4 / 5.68, MW 26.8 / 26. Color scheme based on peptide PEP (red: PEP <1E-11, yellow: PEP<0.0001, light blue: PEP < 0.05, dark blue: PEP>0.05).

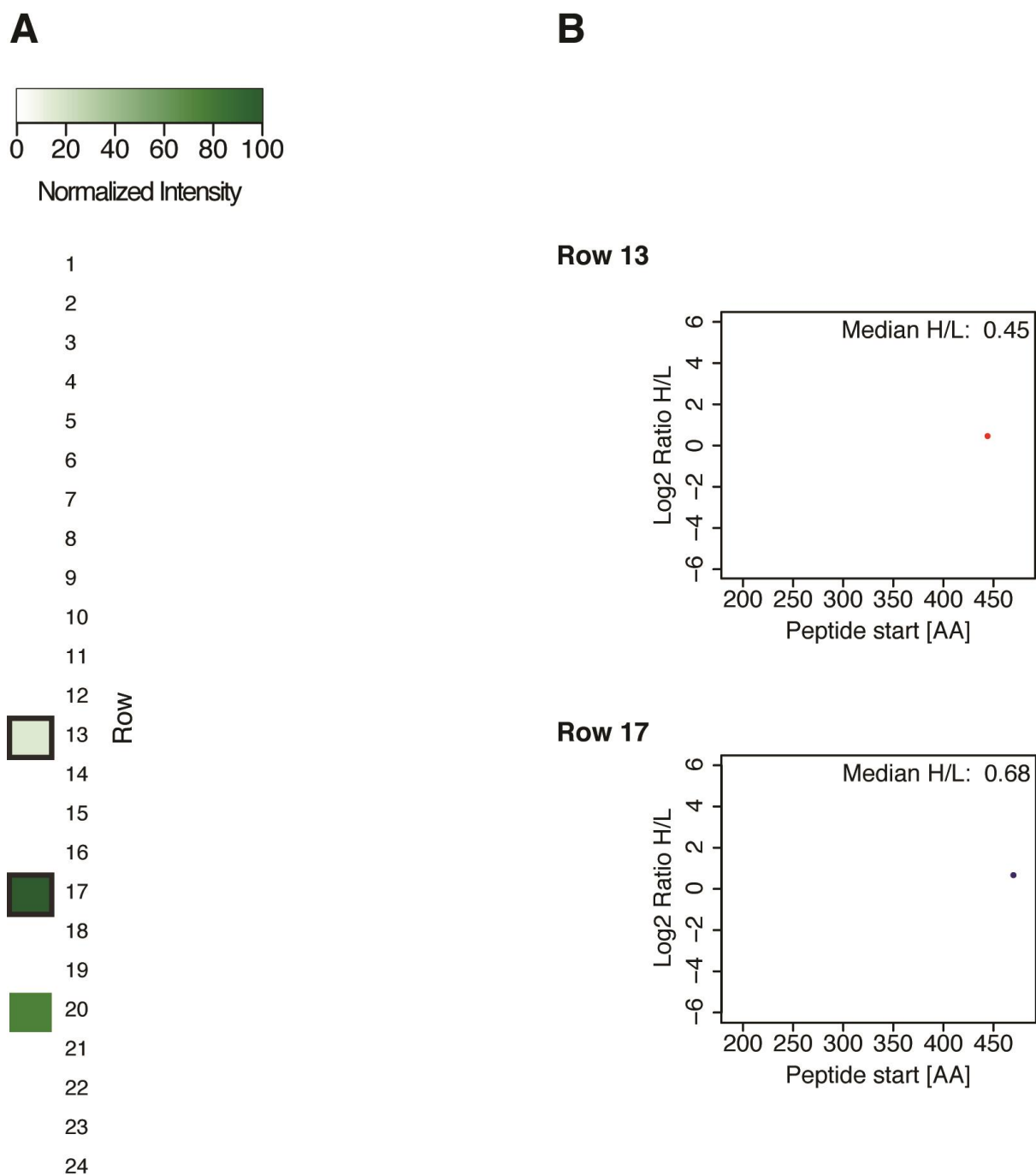


Figure 21. Identification of protein species of T-complex protein 1 subunit alpha (TCP1), 1DE experiment.

A) Distribution of protein MS intensity of TCP1. Gel slices with bold borders were selected and evaluated for ratio H/L and sequence coverage. **B)** Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP <1E-11, yellow: PEP<0.0001, light blue: PEP < 0.05, dark blue: PEP>0.05).

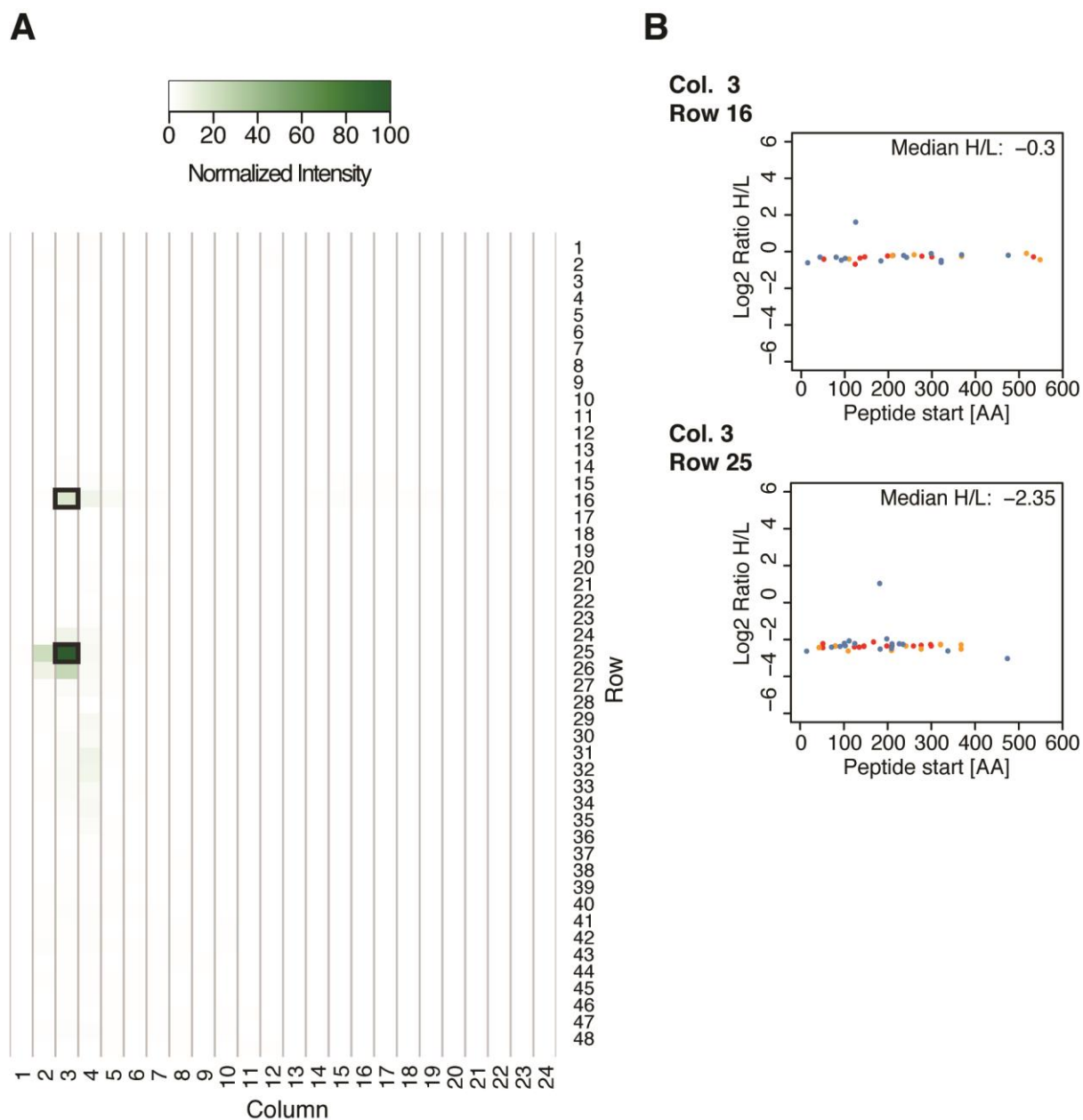


Figure 22. Identification of protein species of lamin-B1 (LMNB1), 2DE experiment.

A) Distribution of protein MS intensity of LMNB1. Gel slices with bold borders were selected and evaluated for ratio H/L and sequence coverage. B) Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Experimental pI and MW vs theoretical pI and MW of the protein species: column 3 / row 16: pI 5.4 / 5.1, MW 65 / 66; column 3 / row 25: pI 5.2 / 5.07, MW 49 / 44. Color scheme based on peptide PEP (red: PEP < 1E-11, yellow: PEP < 0.0001, light blue: PEP < 0.05, dark blue: PEP > 0.05).

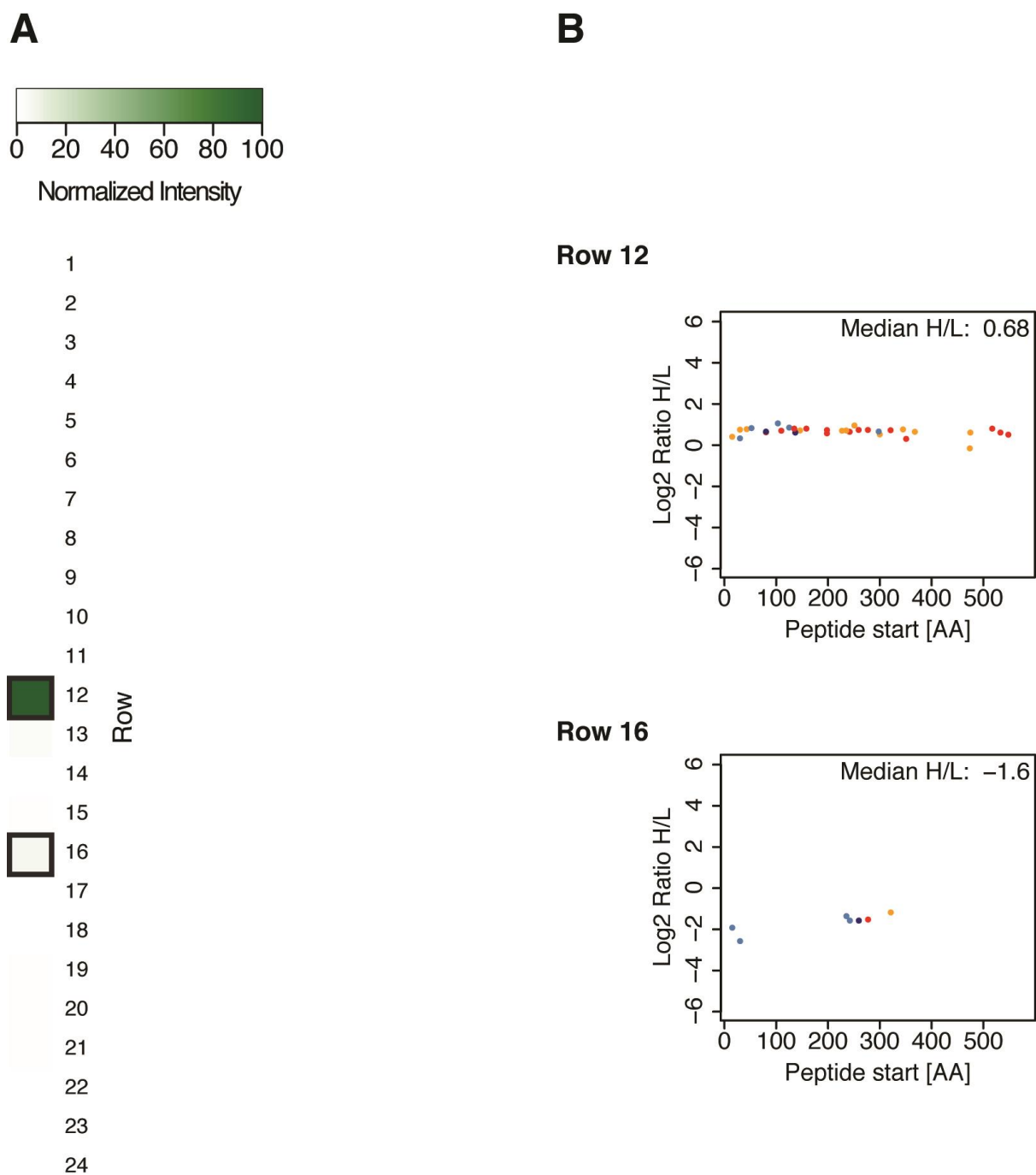


Figure 23. Identification of protein species of lamin-B1 (LMNB1), 2DE experiment.

A) Distribution of protein MS intensity of LMNB1. Gel slices with bold borders were selected and evaluated for ratio H/L and sequence coverage. **B)** Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP <math><1E-11</math>, yellow: PEP <math><0.0001</math>, light blue: PEP <math><0.05</math>, dark blue: PEP >math>>0.05</math>).

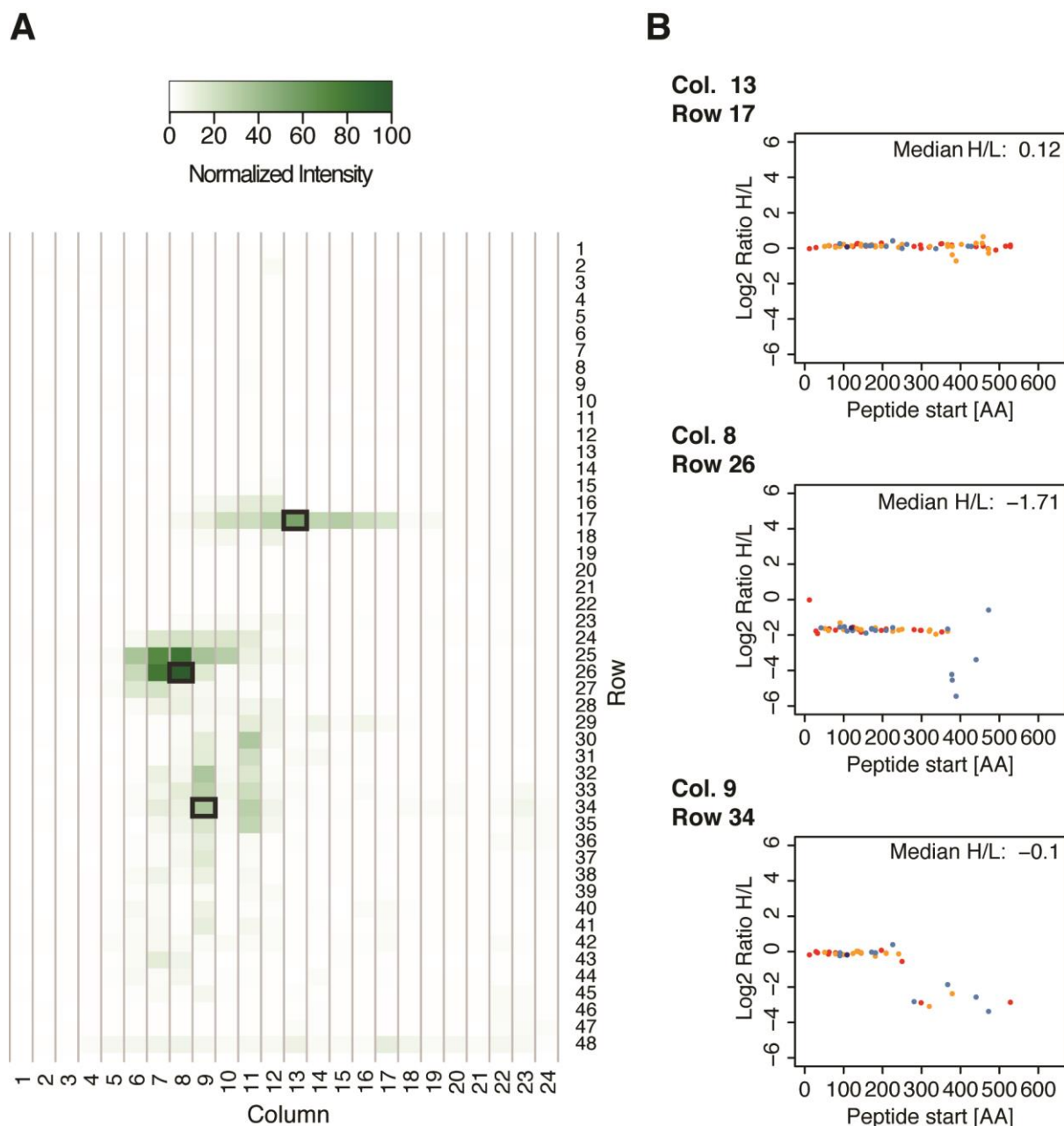


Figure 24. Identification of protein species of lamin A/C (LMNA), 2DE experiment.

A) Distribution of protein MS intensity of LMNA. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. **B)** Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Experimental pI and MW vs theoretical pI and MW of the protein species: column 13 / row 17: pI 6.3 / 8.3, MW 134 / 62; column 8 / row 26 (protein sequence: 0 – 377): pI 5.9 / 5.8, MW 49 / 43; column 9 / row 34 (protein sequence 0 – 261): pI 6.1 / 5.9, MW 40 / 30. Color scheme based on peptide PEP (red: PEP <math><1E-11</math>, yellow: PEP <math><0.0001</math>, light blue: PEP <math><0.05</math>, dark blue: PEP >0.05).

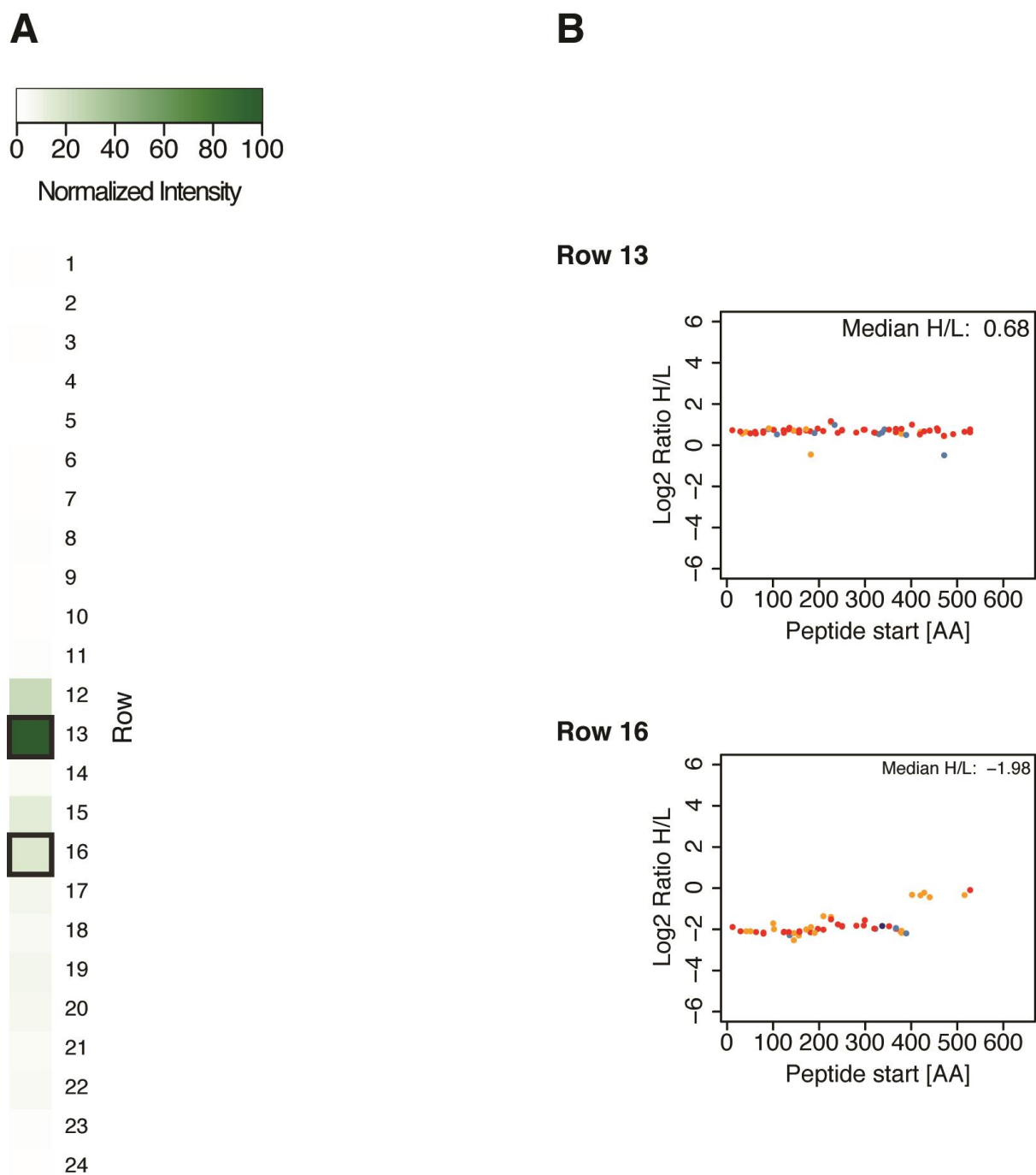


Figure 25. Identification of protein species of lamin A/C (LMNA), 1DE experiment.

A) Distribution of protein MS intensity of LMNA. Gel slices with bold borders were selected and evaluated for ratio H/L and sequence coverage. B) Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP < 1E-11, yellow: PEP < 0.0001, light blue: PEP < 0.05, dark blue: PEP > 0.05).

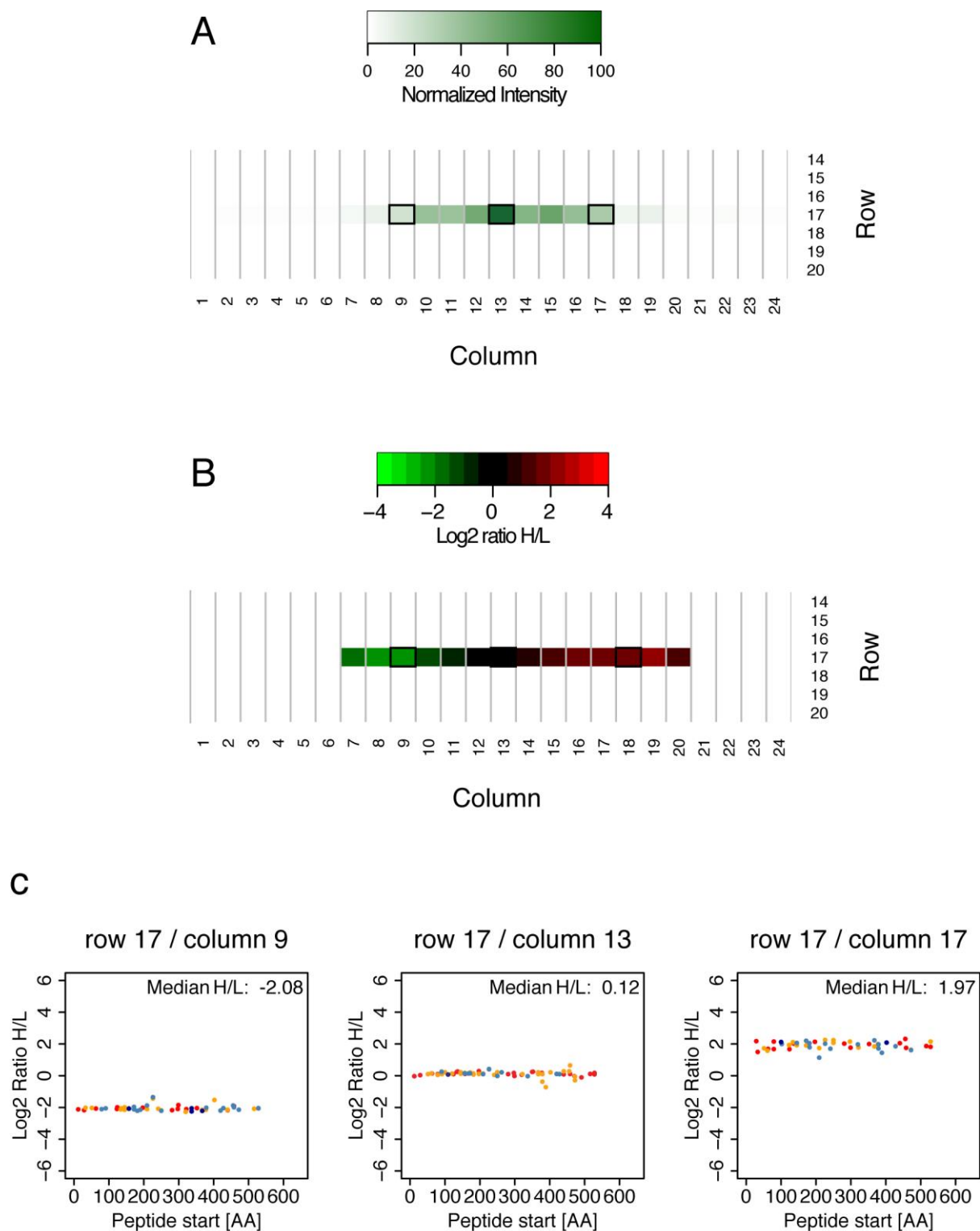


Figure 26. Identification of protein species of lamin A/C (LMNA), 2DE experiment.
A) Distribution of protein MS intensity of LMNA. **B)** Distribution of protein ratio H/L of LMNA. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. **C)** Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP <math><1E-11</math>, yellow: PEP <math><0.0001</math>, light blue: PEP <math><0.05</math>, dark blue: PEP >math>0.05</math>).

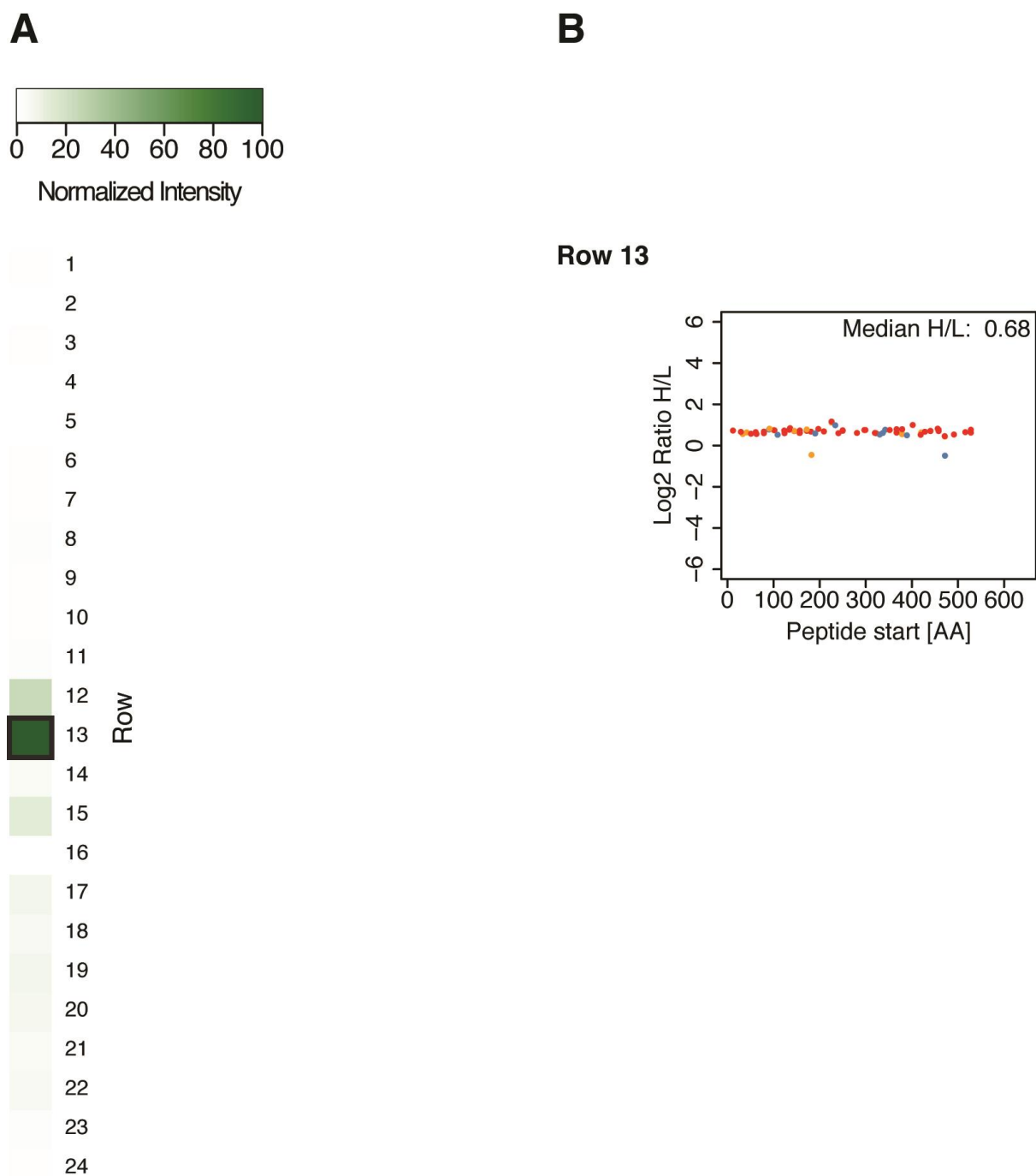


Figure 27. Identification of protein species of lamin A/C (LMNA), 1DE experiment.

A) Distribution of protein MS intensity of LMNA. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. B) Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP < 1E-11, yellow: PEP < 0.0001, light blue: PEP < 0.05, dark blue: PEP > 0.05).

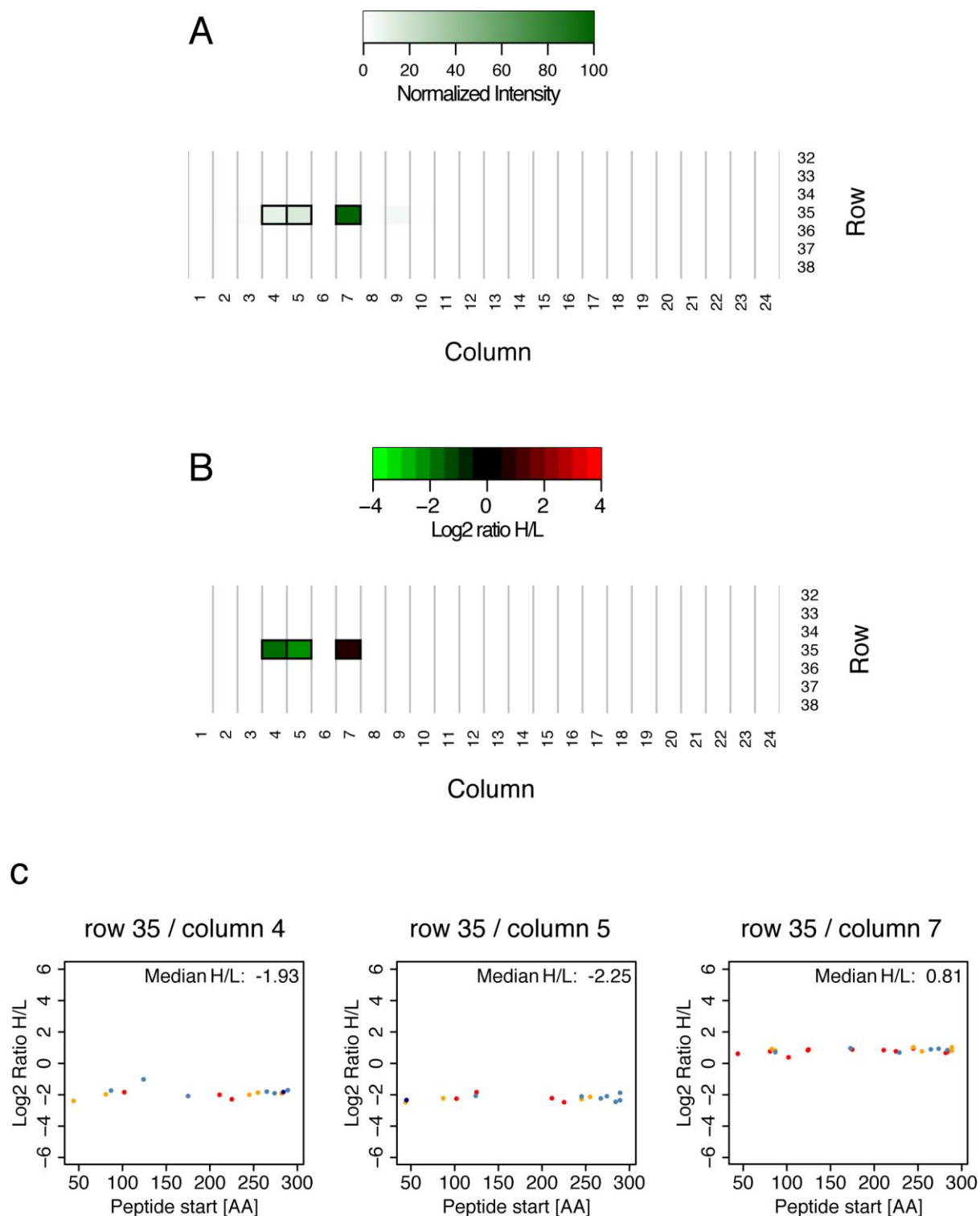


Figure 28. Identification of protein species of F-actin-capping protein subunit beta (CAPZB), 2DE experiment.

A) Distribution of protein MS intensity of CAPZB. **B)** Distribution of protein ratio H/L of CAPZB. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. **C)** Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP <math><1E-11</math>, yellow: PEP <math><0.0001</math>, light blue: PEP <math><0.05</math>, dark blue: PEP >math>0.05</math>).

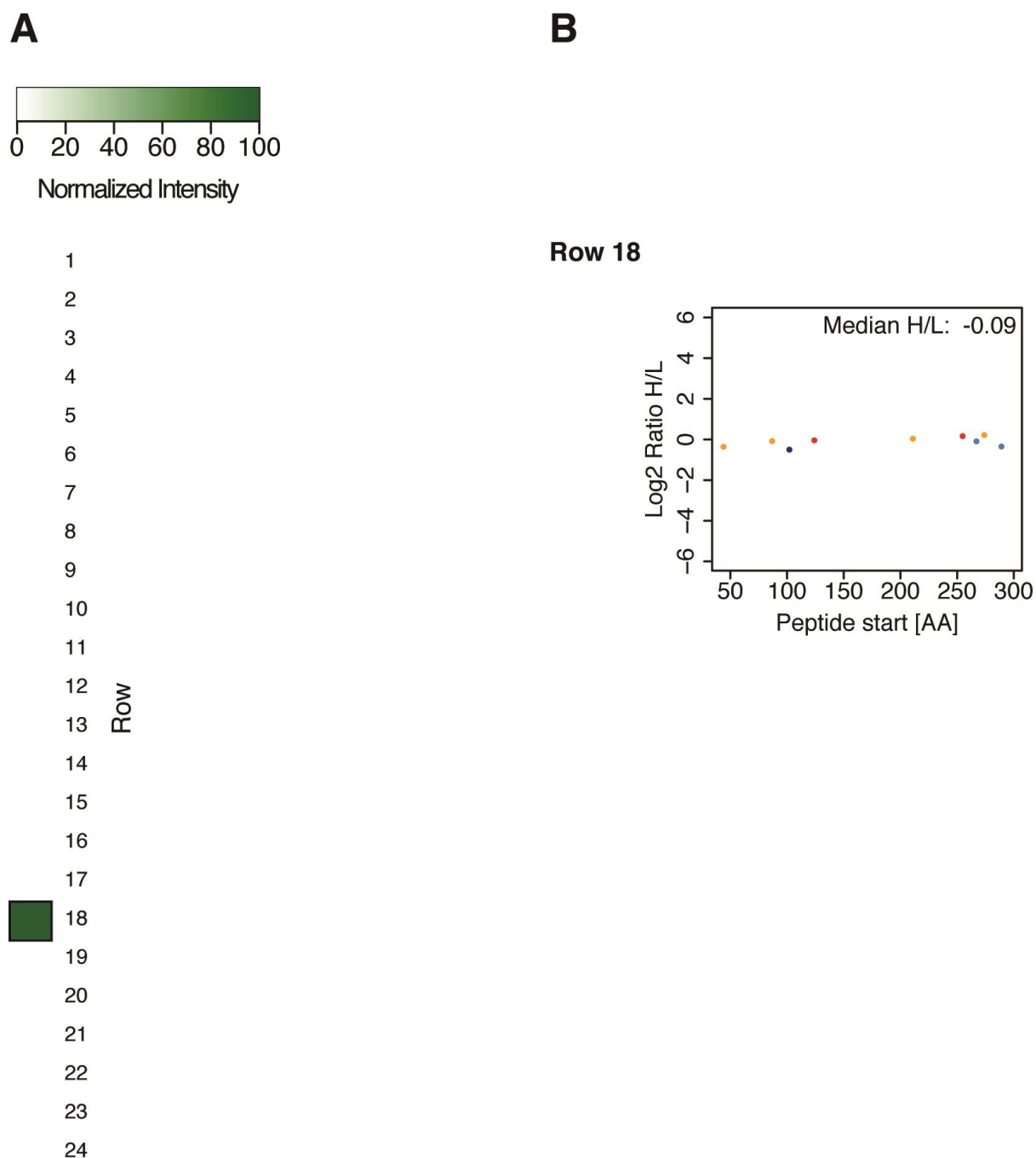


Figure 29. Identification of protein species of F-actin-capping protein subunit beta (CAPZB), 1DE experiment.

A) Distribution of protein MS intensity of CAPZB Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. B) Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP $1E-11$, yellow: PEP<math><0.0001</math>, light blue: PEP <math><0.05</math>, dark blue: PEP>0.05).

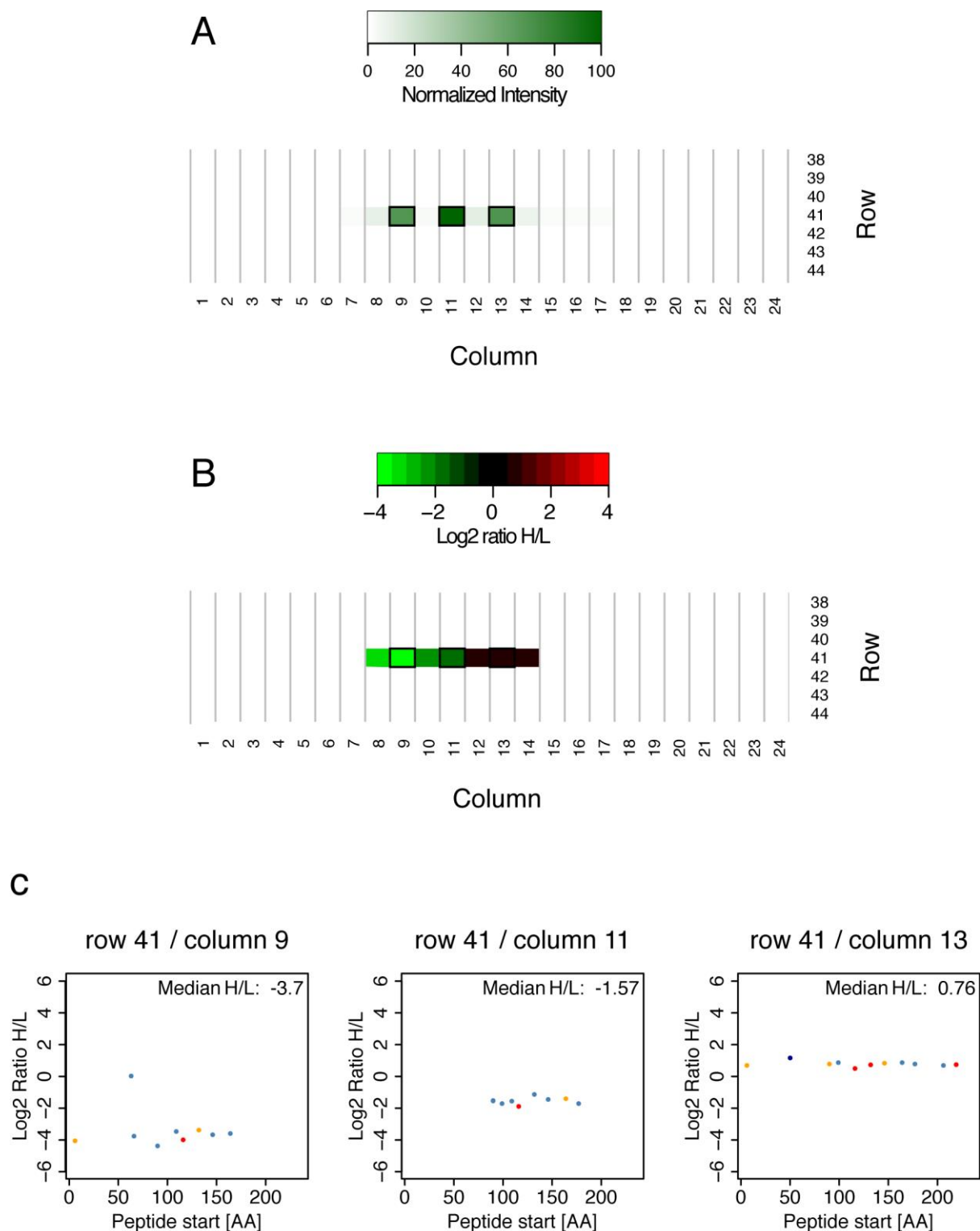


Figure 30. Identification of protein species of UPF0568 protein C14orf166 (C14orf166), 2DE experiment.

A) Distribution of protein MS intensity of C14orf166. B) Distribution of protein ratio H/L of CAPZB. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. C) Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP <math><1E-11</math>, yellow: PEP <math><0.0001</math>, light blue: PEP <math><0.05</math>, dark blue: PEP >0.05).

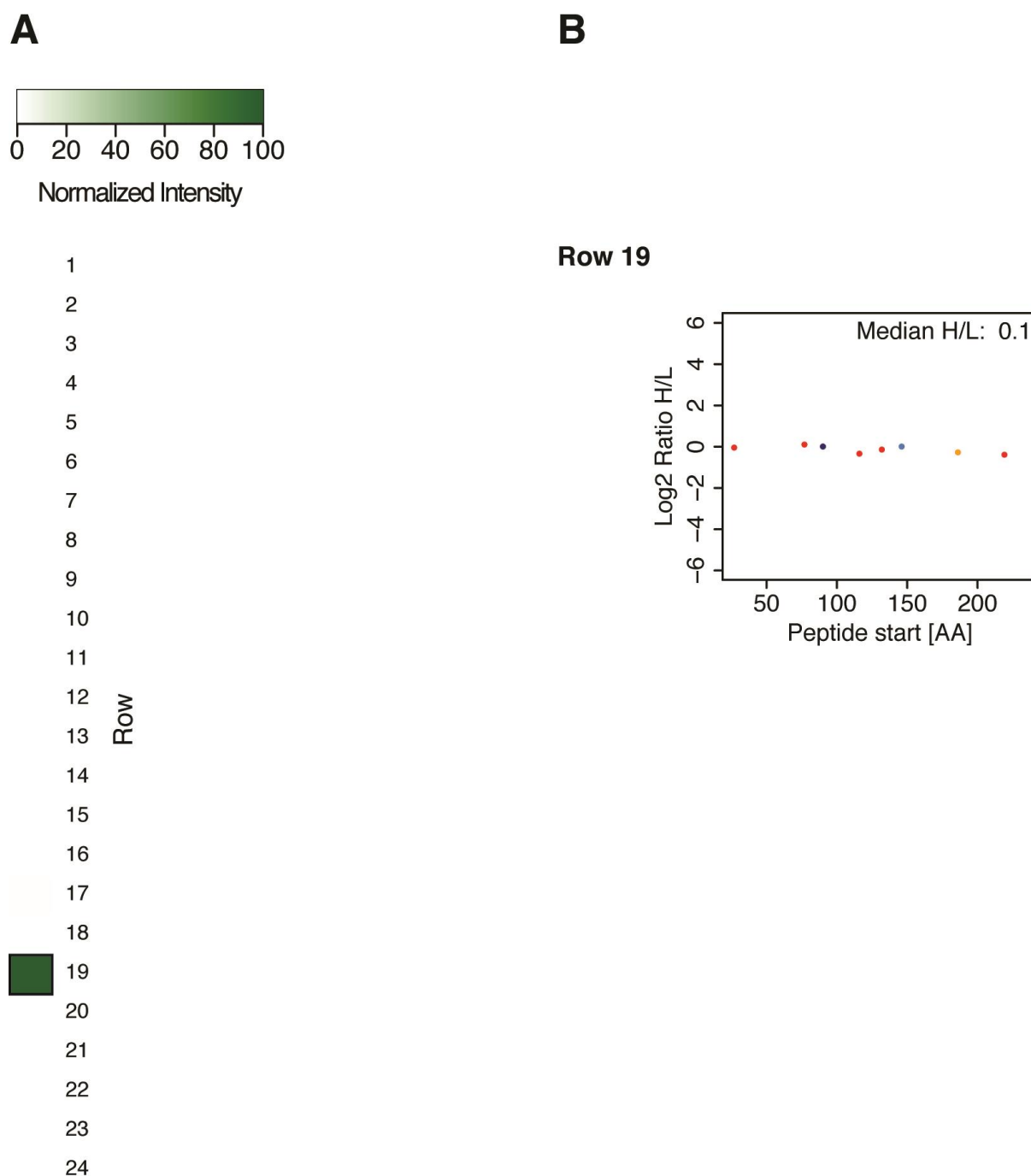


Figure 31. Identification of protein species of UPF0568 protein C14orf166 (C14orf166), 1DE experiment.

A) Distribution of protein MS intensity of C14orf166. Gel slices with bold borders were selected to display the ratio H/L of the identified peptides and the protein sequence coverage. B) Sequence coverage, presented as the position of the first amino acid of each identified peptide, versus ratio H/L of the identified peptides. Color scheme based on peptide PEP (red: PEP $1E-11$, yellow: PEP <math><0.0001</math>, light blue: PEP <math><0.05</math>, dark blue: PEP >math>0.05</math>).

An evaluation of the distribution of the intensity and the ratio H/L of protein species revealed distinct patterns of distribution over the 2DE gel. Interestingly there was an agreement between the experimental pI and MW of the 2DE slice and the theoretical pI and MW of the potential protein species. Examples of good agreement were proteins PLS3 (Figure 18) and LMNB1 (Figure 22). Protein VIM showed good agreement in the MW values while the pI values showed differences of between 0.5 and 1 (Figure 16). Protein TCP1 was also a prominent example with species showing similar experimental and theoretical pI values. Based on the sequence coverage in the different fractions, the protein seemed to be split into N and C terminal part (Figure 20) and both protein species had almost the same ratio H/L value of -1.96 on log₂ scale. These results show that multiple protein species per protein indeed exist. For several proteins, the identified species showed different H/L ratios. For example the full length VIM species had a median ratio of 2.43 while the species missing the 100 N-terminal amino acids had a median ratio of -0.65 (log₂ scale). The full length PLS3 protein species had a ratio of 1.57 while the protein species lacking the N-terminal 125 amino acids had a ratio of -2.72. The full length of LMNB1 had a ratio of -0.3 and the second protein species, which lacked approximately 200 amino acids on the C terminus, showed a ratio of -2.35. These results indicate that while multiple species of a protein are present in the protein sample, they can have different relative abundance between CPC-2D and CPC-3D. For the selected set of proteins that showed the presence of multiple species, the performance, in terms of separation and identification, of the 1DE analysis was evaluated. Interestingly the 1DE performed worse than the 2DE analysis. In several cases protein separation by 1DE did not provide enough evidence for the presence of multiple protein species per protein. In the cases where the presence of more than one protein species was indicated, there was carryover from other slices suggesting insufficient separation power. An example for this is protein VIM. For protein VIM there is evidence for the presence of three protein species (Figure 16) on the 2DE gel, one of which lacks 100 amino acids at the N-terminus. The 1DE analysis also revealed the presence of two species, with the second species possibly also lacking 100 amino acids at the N-terminus. This could be inferred from the slightly higher ratio H/L of the peptides from the N-terminus region relative to the rest of the protein. Nevertheless, the evidence for the existence of a second protein species was much less conclusive in the 1DE analysis compared to the 2DE analysis. For three other proteins: PLS3 (Figure 19), TCP1 (Figure 21) and LMNB1 (Figure 23), there was no conclusive evidence for the existence of more than one protein species. The 2DE analysis revealed the presence of at least three proteins species of LMNA (Figure 24) while the 1DE analysis indicated the existence only of two protein species (Figure 25). These results strongly suggest that 2DE

is capable of efficiently separating/isolating different species of the same protein, which aids their confident identification and characterization.

Protein separation by pI is the crucial difference between 1DE and 2DE. In order to evaluate the benefit of this additional separation dimension, three proteins that showed different horizontal ratio H/L distribution were selected and their identification in the 1DE analysis was examined. One prominent example of the presence of several protein species, which also show different relative abundance between CPC-2D and CPC-3D, was protein LMNA (

Figure 26). For LMNA a distinct horizontal spread on the 2DE gel of the protein identification and decreasing ratio H/L from right to the left was observed. An evaluation of three slices from the horizontal spread revealed that in all three slices the full-length protein was present. This indicated that the spread is most probably a result of a modification of the protein, which affected its pI and thus horizontal migration on the 2DE gel, and, furthermore, that the degree of modification of the protein was different in CPC-2D and CPC-3D resulting in different ratios across the horizontal spread. Another interesting observation related to LMNA is that the full-length protein has a theoretical pI and MW, which are very different from the experimental pI and MW of the 2DE slice (Figure 24). The theoretical pI and MW of full length LMNA is 8.3 and 62 KDa. The highest intensity of the full length LMNA protein is in column 13, row 17, which corresponds to pI of 6.3 and MW of 134 KDa. This again indicates that this is a modified version of the protein which shifts its pI to more acidic and its MW to higher values. One candidate for such a modification is phosphorylation (Zhu, Zhao et al. 2005). As expected, the 1DE analysis revealed the presence of the full length LMNA protein (Figure 27) but due to the absence of separation by pI no additional species at the same MW were revealed.

The analysis of two other proteins: CAPZB (Figure 28) and c14orf166 (Figure 30) also revealed the presence of protein species that have the same MW but are separated based on their pI. Interestingly, in both examples the ratio H/L decreased from right to left across the 2DE gel. Similarly to the example of LMNA, the 1DE analysis of both proteins (Figure 29, Figure 31) did not reveal the presence and differential quantification of separate protein species. These results indicate that due to its separation power, the 2DE approach allows for the identification and quantification of multiple protein species. In comparison, the identification rate of protein species was much lower when using the 1DE approach.

3.4 Establishment of three dimensional PAGE-pIEF-LC-MS/MS separation method

The results from the 1DE and 2DE proteomic analysis of the proteome wide changes of CPCs ongoing chondrogenesis showed that both the 2DE and 1DE analyses resulted in the identification of a large number of proteins (Table 6). For the 1DE analysis this was achieved by fractionating the whole 1D PAGE lane into 24 fractions. This result shows the utility of 1D PAGE in proteomic analysis which has already been well described (Schirle, Heurtier et al. 2003). An interesting perspective is to attempt to couple the protein separation in using 1DE with another separation dimension, e.g. at the peptide level, in order to achieve deeper coverage of the expressed proteome. In this line, the results in this chapter are dedicated to the establishment and evaluation of a novel separation method in which proteins are first separated using 1D PAGE, the whole gel lane is fractionated and subjected to in-gel digestion followed by peptide isoelectric focusing (pIEF) prior to LC-MS/MS analysis.

The complex proteome sample of HeLa cell nuclear extract (NE) was used for the method development and optimization. First, the performance of protein separation (1D PAGE) was evaluated. For this purpose 70 μ g of HeLa NE was separated using 1D PAGE, the whole gel lane was fractionated into 23 slices and all fractions that showed Coomassie staining (slices 2-21) were digested with trypsin in-gel and the resulting peptides were analyzed using LC-MS/MS (described further as GeLC-MS/MS). Next, the redundancy of protein identification between the different PAGE fractions was evaluated. Overall, 66% of all identified proteins were unique and identified in a single PAGE slice only (Figure 32, A). Furthermore, 85% of all peptides derived from trypsin digestion of the proteins were unique and identified in a single PAGE slice (Figure 32, B). These results indicate that 1D PAGE efficiently separates proteins, which, after gel fractionation and in-gel digestion, results in the identification of peptides that are non-redundant between PAGE fractions. At this point it was important to identify a method that can efficiently separate tryptic peptides from each PAGE fraction and can easily be integrated in the workflow. The method of choice was pIEF. It has been shown that pIEF can attain high resolution of separation and reproducibility using low (10 μ g) or high (100 μ g) of peptide sample (Slebos, Brock et al. 2008). While pIEF can be performed using narrow range pH gradients, e.g. pH 4 to 4.5 (Eriksson, Lengqvist et al. 2008), the pH gradient 3 to 10 was chosen as it covers most of the potential peptides generated from trypsin digestion of complex protein mixtures (Cargile, Talley et al. 2004). In order to evaluate the performance of pIEF, 70 μ g of HeLa NE were digested and separated using pIEF on an 18cm IPG strip, pH range 3-10. The

whole IPG strip was fractionated into 36 fractions, followed by peptide extraction analysis using LC-MS/MS. Next, the performance of the separation was evaluated. Since the tryptic peptides from a single protein can be identified in different fractions, only the redundancy of peptide identification in each fraction was evaluated. Overall 79 % of all identified peptides in all 36 pIEF slices were unique and found in one pIEF fraction only (Figure 32, C). This demonstrates the high resolving power of pIEF.

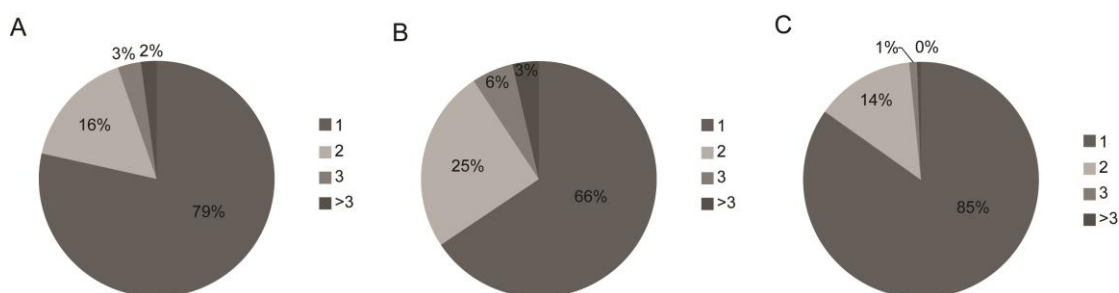


Figure 32. Separation efficiency of pIEF and 1D PAGE.

A) Proportion of peptides identified only in 1, 2, 3 or more 18-cm IPG strip slices (pIEF-LC-MS/MS). B) Proportion of proteins identified only in 1, 2, 3 or more PAGE slices (GeLC-MS/MS). C) Proportion of peptides identified only in 1, 2, 3 or more PAGE slices (GeLC-MS/MS).

Source: (Atanassov and Urlaub 2013).

Next a combined separation approach was employed in which, first, proteins were separated by PAGE and the whole gel lane was fractionated and subjected to in-gel trypsin digestion. The tryptic peptides from a PAGE fraction were directly separated by pIEF, the strip was fractionated and the peptides were extracted and analyzed using LC-MS/MS. This combined approach was named PAGE-pIEF-LC-MS/MS (Figure 33, A). In order to reduce the number of fractions (36 for the 18 cm IPG strip) for analysis, a shorter, 7 cm IPG strip was employed which reduced the number of pIEF fractions to 13.

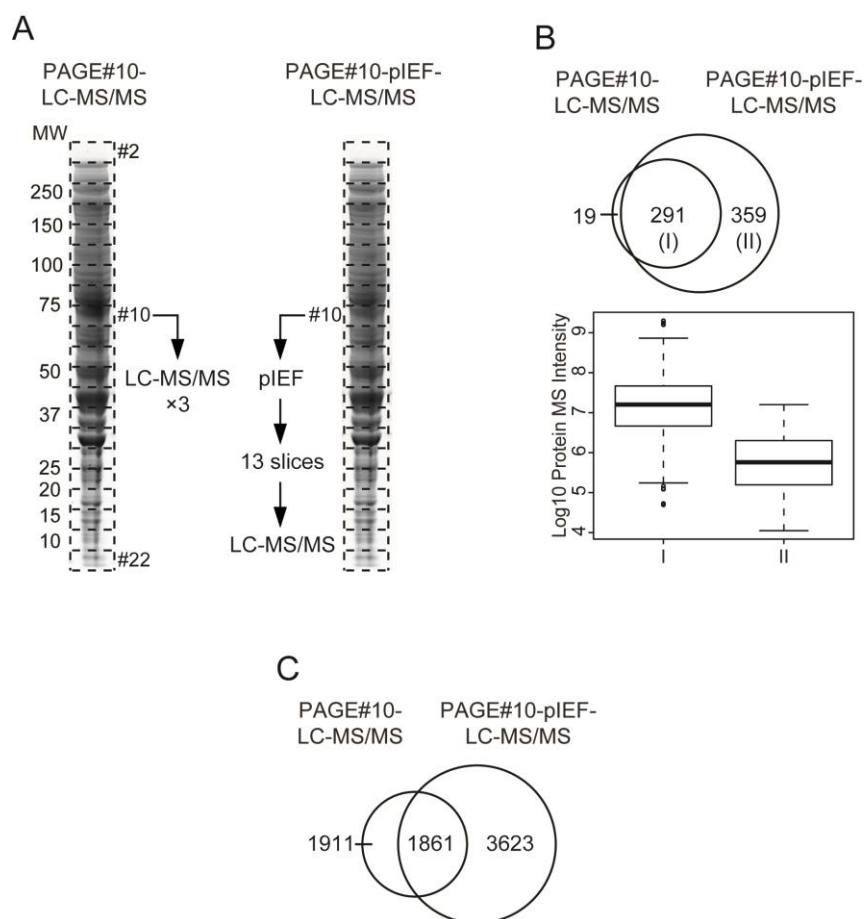


Figure 33. Proteome profiling by PAGE-pIEF-LC-MS/MS.

A) Schematic representation of the workflow. 1D PAGE gels were fractionated, and proteins from gel slice #10 were digested in-gel and then analyzed by LC-MS/MS (PAGE#10-LC-MS/MS) or, after additional separation of the extracted peptides, by pIEF prior to LC-MS/MS (PAGE#10-pIEF-LC-MS/MS). **B)** Overlap between proteins identified in PAGE#10-LC-MS/MS and PAGE#10-pIEF-LC-MS/MS and box plots showing the MS intensity distribution of "common" proteins, i.e. those identified by both PAGE#10-LC-MS/MS and PAGE#10-pIEF-LC-MS/MS (I) and of the "newly-identified" proteins, i.e. those identified only when pIEF was also performed (II). **C)** Overlap of identified peptides between PAGE#10-LC-MS/MS and PAGE#10-pIEF-LC-MS/MS.

Source (Atanassov and Urlaub 2013).

In order to evaluate the performance of the new multidimensional separation approach, the numbers of identified proteins and peptides from a PAGE slice, with and without performing pIEF, were compared. PAGE slice #10 was selected as an example. Including the pIEF step significantly increased the total number of identified proteins from the PAGE fraction. Without performing pIEF (PAGE#10-LC-MS/MS) only 310 proteins were identified. In comparison, including the pIEF separation step prior to LC-MS/MS analysis (PAGE#10-pIEF-LC-MS/MS) resulted in the identification of 650 proteins (Figure 33, B). There was a large overlap of the identified proteins using both approaches and only 19 proteins were not identified using PAGE-pIEF-LC-MS/MS. This indicates that the depth of the analysis increased, rather than identifying different subpopulations of proteins in the PAGE fraction.

Furthermore, proteins uniquely identified using the PAGE-pIEF-LC-MS/MS (359 proteins) approach had lower intensity, recorded by the mass spectrometer, as compared with those identified by both methods (Figure 33, B). This indicated that including the additional pIEF separation step achieved in the identification of lower-abundance proteins. Regarding the numbers of identified peptides, around 3700 were identified without performing pIEF prior to LC-MS/MS. In comparison, around 5400 peptides were identified when pIEF separation was performed prior to LC-MS/MS analysis (Figure 33, C). The increased number of identified peptides also resulted in an increase of the median protein sequence coverage of the 291 proteins identified in both approaches from 12 to 23%. This shows that in addition to increasing the number of identified proteins, performing pIEF also increased the number of identified peptides per proteins, which generally increases the confidence of a protein identification (Baldwin 2004). While the comparison of numbers of identified proteins and peptides shows the performance of the new method, more information is needed in order to understand the reasons for the increased rates of identification. Such information can be obtained from studying the separation patterns of the additionally identified peptides and proteins. In order to examine this, peptides identified in PAGE#10 using pIEF-LC-MS/MS were compared with those identified by LC-MS/MS. The peptides were separated into three groups, based on the proteins to which they mapped (Figure 33, B): 1) common peptides mapping to common proteins, i.e., those peptides identified by both PAGE#10-LC-MS/MS and PAGE#10-pIEF-LC-MS/MS (1861 peptides matching to the 291 common proteins); 2) newly-identified peptides, identified by PAGE#10-pIEF-LC-MS/MS only (2484 peptides) that mapped to one of the 291 common proteins; and 3) newly-identified peptides (1139) that did not map to one of 291 common proteins but match to one of the 359 proteins that were identified, by PAGE-pIEF-LC-MS/MS only. Since the peptide separation consists of fractions of the IPG strip (1 to 13) and retention on the RP-HPLC column, the distribution of the identified peptides over the IPG strip slice/RP-HPLC retention time map was examined (Figure 34).

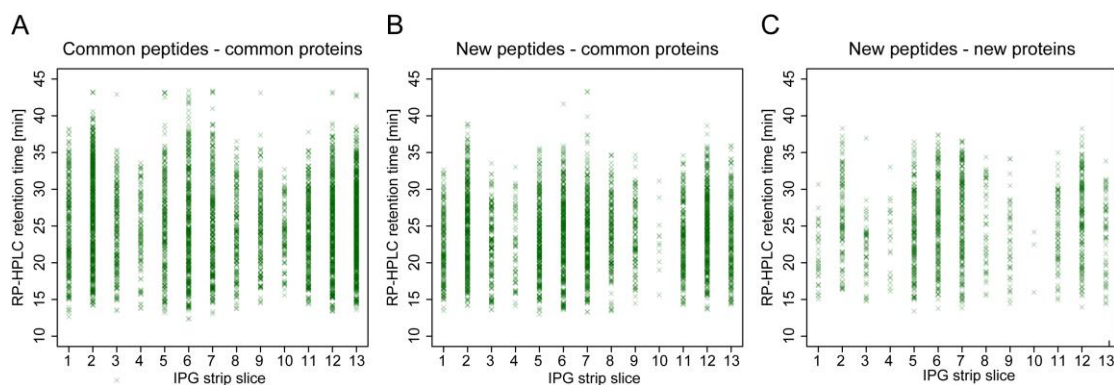


Figure 34. Orthogonality plots of RP-HPLC retention time and IPG strip slice.

(A) peptides identified in PAGE#10-LC-MS/MS and PAGE#10-pIEF-LC-MS/MS and mapped to proteins common to both analyses; **(B)** newly PAGE#10-pIEF-LC-MS/MS identified peptides from proteins common to PAGE#10-LC-MS/MS and PAGE#10-pIEF-LC-MS/MS; or **(C)** peptides from the proteins identified only by PAGE#10-pIEF-LC-MS/MS. Each cross indicates a peptide spectrum match. Numerous PSMs can map to a single peptide.

Source: (Atanassov and Urlaub 2013).

Examination of the distribution of the peptides reveals an efficient use of the separation space. Most peptides are identified with RP-HPLC retention time between 12 and 43 min and peptides are identified in all IPG fractions. The spread of the identified proteins also indicates that peptide separation by pI and hydrophobicity is orthogonal (Gilar, Olivova et al. 2005). Also, the three peptide groups, selected based on the proteins to which they mapped (Figure 34), showed similar orthogonality plots of their RP-HPLC retention time and IPG gel slice coordinates (Figure 34). These results suggest that the increased separation by applying pIEF prior to LC-MS/MS analysis provided the mass spectrometer with more time to sequence peptides coming from lower abundance proteins and that these peptides were efficiently separated from those derived from higher abundance proteins.

Next, it was tested whether performing a longer RP-HPLC gradient, a commonly applied step when separating complex mixtures (Köcher, Swart et al. 2011), would deliver a larger number of protein/peptide identifications when compared to pIEF separation prior to regular (33 min) LC-MS/MS analysis. The longer gradient was set to 221 minutes, a gradient length which has been shown to be a good balance between number of identified proteins and analytical time (Xu, Duong et al. 2009). Performing an LC-MS/MS analysis of the peptides from PAGE fraction 10 using the long RP-HPLC gradient resulted in the identification of 425 proteins and 3792 peptides. In comparison 238 proteins and 2141 peptides were identified using the short RP-HPLC gradient. Importantly, performing the additional pIEF separation resulted in the identification of a larger number of proteins (741) and peptides (8606) (Figure 35).

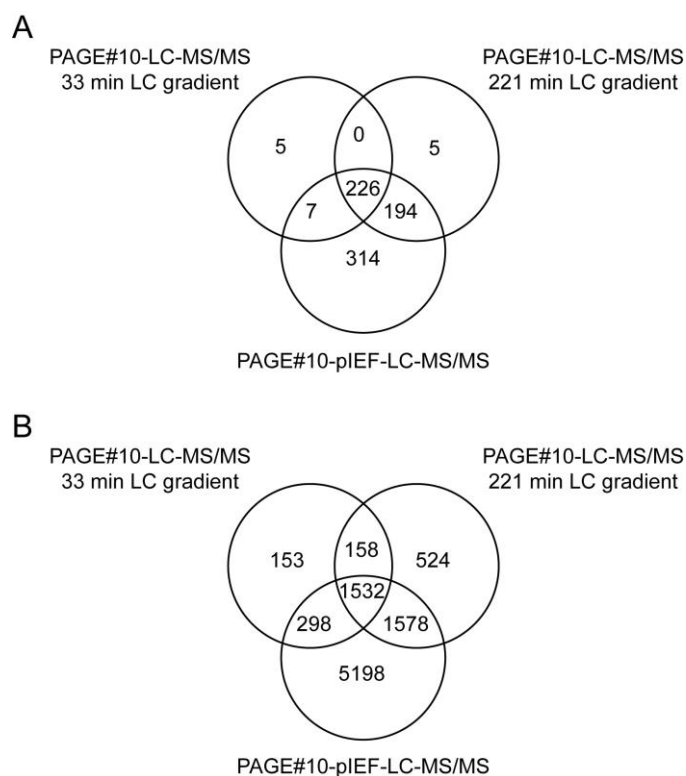


Figure 35. Comparison of numbers of identified proteins using short LC gradient, long LC gradient and pIEF.

A) Overlap of proteins identified from PAGE#10 using 33-min LC separation gradient (PAGE#10-LC-MS/MS 33 min LC gradient), 221-min LC separation gradient (PAGE#10-LC-MS/MS 221 min LC gradient), and pIEF fractionation prior to LC-MS/MS (PAGE#10-pIEF-LC-MS/MS). B) Overlap of peptides identified from PAGE#10 using 33-min LC separation gradient (PAGE#10-LC-MS/MS 33 min LC gradient), 221-min LC separation gradient (PAGE#10-LC-MS/MS 221 min LC gradient), and pIEF fractionation prior to LC-MS/MS (PAGE#10-pIEF-LC-MS/MS).

Source: (Atanassov and Urlaub 2013).

When pIEF was performed on all PAGE fractions, all three proteomics workflows – 1D PAGE-LC-MS/MS (GeLC-MS/MS), 18 cm pIEF-LC-MS/MS, and the novel PAGE-pIEF-LC-MS/MS were compared in terms of numbers of identified proteins and peptides. While the analysis using 18 cm pIEF strip separation is different from the GeLC-MS/MS and PAGE-pIEF-LC-MS/MS approaches, in terms of sample preparation, it was included in the comparison as the method is commonly used as the sole separation method prior to LC-MS/MS analysis (Krijgsveld, Gauci et al. 2006, Hubner, Ren et al. 2008). The comparison between the three approaches revealed that the application of the novel PAGE-pIEF-LC-MS/MS method resulted in increased proteome coverage, as inferred from the numbers of identified proteins and peptides (Figure 36, A).

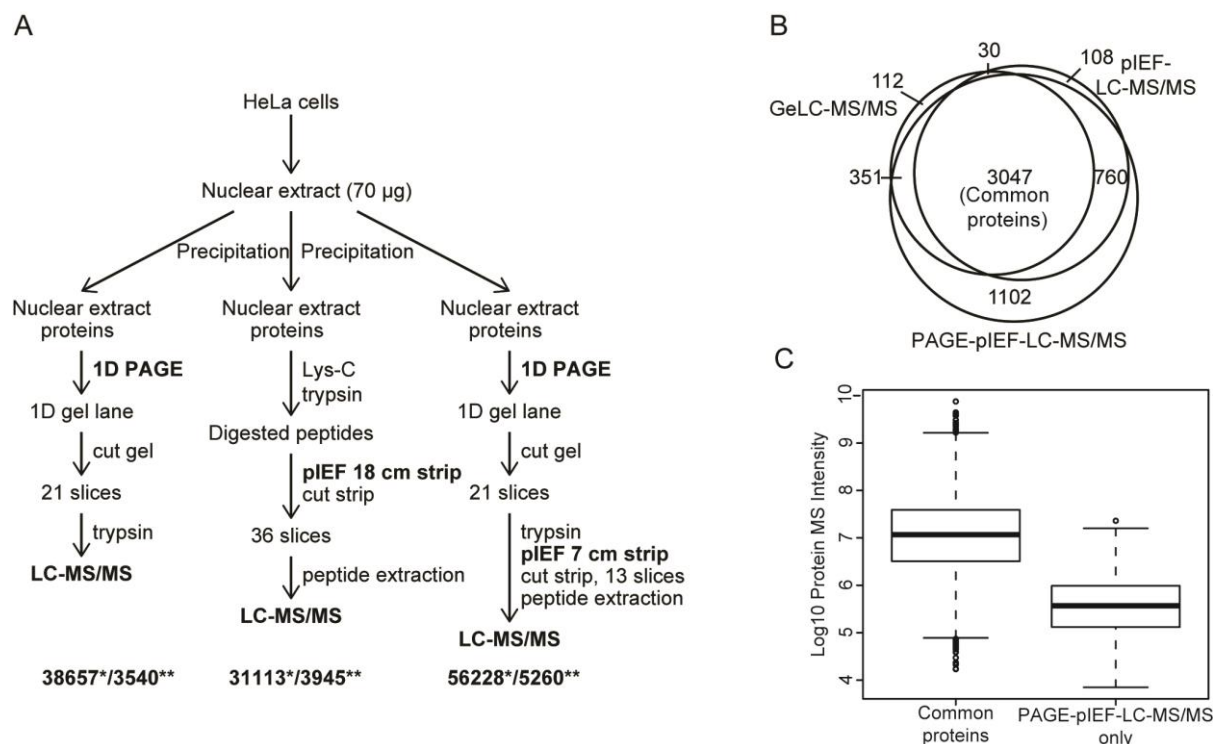


Figure 36. Comparison of proteome coverage using GeLC-MS/MS, pIEF-LC-MS/MS and PAGE-pIEF-LC-MS/MS.

A) Schematic presentation of the 3 workflows employed in the study, showing the number of identified peptides (*) and proteins (**). **B)** Overlap of the proteins identified by GeLC-MS/MS, pIEF-LC-MS/MS, or PAGE-pIEF-LC-MS/MS. **C)** Intensity distributions of proteins identified in all three workflows (Common proteins) and those identified only by PAGE-pIEF-LC-MS/MS. Source: (Atanassov and Urlaub 2013).

Using the GeLC-MS/MS approach, 3540 proteins were identified based on 38657 peptides while using pIEF-LC-MS/MS, 3945 proteins and 31113 peptides were identified. At the same time 5260 proteins and 56228 peptides were identified using the PAGE-pIEF-LC-MS/MS method. The overlap of identified proteins was large with around 3050 proteins being identified using all three approaches. Since the GeLC-MS/MS and the PAGE-pIEF-LC-MS/MS approaches differ only by the additional separation by pIEF, the sequence coverage of the proteins identified using both approaches was compared. The median sequence coverage of all proteins identified in both methods increased from 17% (GeLC-MS/MS) to 26% for PAGE-pIEF-LC-MS/MS. This was due to the higher number of identified peptides using the latter approach. The large number of proteins identified using all three approaches (3047 proteins) indicated that similar sets of proteins are identified using all three methods. At the same time more than 1100 proteins were identified only after performing PAGE-pIEF-LC-MS/MS (Figure 36, B). These 1100 proteins had a much lower MS intensity distribution when compared with those identified in all three workflows (Figure 36, C), indicating that the proteins identified by PAGE-pIEF-LC-MS/MS only were of lower abundance. While the use of the PAGE-pIEF-LC-MS/MS approach resulted in the

identification of more than 5200 proteins, there were a number of proteins that were only identified using the other two approaches. An examination of the identity of those proteins revealed that more than 90% of the 112 proteins identified only by GeLC-MS/MS and of the 108 proteins identified only by pIEF-LC-MS/MS were based on a single peptide. It has been recognized that acceptance of protein identification based on a single peptide has to be done with caution due to the higher risk of protein false positive identification (Nesvizhskii and Aebersold 2004). In comparison to those numbers, 41% of the 1100 proteins identified only by PAGE-pIEF-LC-MS/MS were based on one peptide and the remaining proteins were identified by at least 2 peptides. This showed that most of the additionally identified lower abundance proteins were of high confidence.

The observation that the application of PAGE-pIEF-LC-MS/MS resulted in an increase in the proteome coverage (in comparison to GeLC-MS/MS) shows that the additional pIEF separation step has efficiently separated peptides prior to LC-MS/MS analysis. It was already established that the first dimension, 1D PAGE, efficiently separates proteins and results in identifications that are not redundant between PAGE fractions (Figure 32, A). In order to evaluate the efficiency of pIEF separations, the redundancy of peptide identifications in all pIEF fractions was calculated. In all except one of the pIEF separations, more than 70% of the identified peptides were found only in a single IPG gel fraction. The low redundancy of peptide identification between fractions, indicate that the separation was very efficient.

Besides the total number of identified proteins using PAGE-pIEF-LC-MS/MS it was important to identify the rate of protein identifications per PAGE fraction. The numbers of identified proteins were also compared with the results from the GeLC-MS/MS method. Overall, there was a significant increase in the number of identified proteins per PAGE fraction, after applying pIEF (Figure 37).

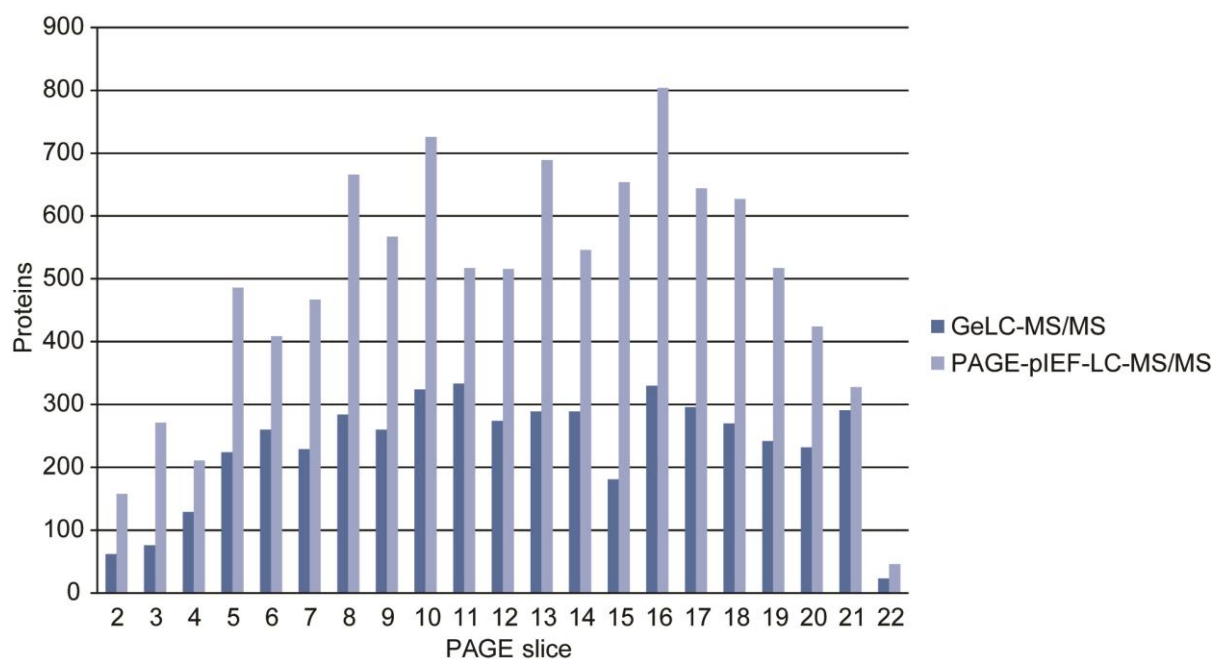


Figure 37. Proteins identified for each PAGE slice from the GeLC-MS/MS and PAGE-pIEF-LC-MS/MS approaches.

Source: (Atanassov and Urlaub 2013).

An exception was the very low MW region (PAGE fractions 21 and 22), where a few additional identifications were obtained after performing pIEF prior to LC-MS/MS analysis. While an increase in the depth of proteome coverage is the primary goal of a separation technique, its reproducibility is an essential factor. As the PAGE-pIEF-LC-MS/MS approach relies on two different separation techniques prior to LC-MS/MS analysis, the technical reproducibility of pIEF and PAGE-pIEF was examined separately. In order to evaluate the reproducibility of pIEF, three 1D PAGE fractions, corresponding to the high, middle and low protein MW region (PAGE#05, MW ~ 150 kDa, PAGE#10, MW ~ 70 kDa, and PAGE#15, MW ~ 40 kDa) were selected. The tryptic peptides from each fraction were split into two and analyzed using pIEF separation prior to LC-MS/MS. The selected criteria for evaluation of the reproducibility of the method were consistency in the number of protein and peptide identifications, overlap of the identified proteins in the replicate pIEF analyses and, finally, the distribution of identified peptides across the IPG strip. The numbers of identified proteins and peptides in the pIEF replicates are shown in Table 7.

Table 7. Numbers of identified peptides and proteins in pIEF replicates.
Source: (Atanassov and Urlaub 2013)

PAGE slice	PAGE#05		PAGE#10		PAGE#15	
pIEF replicate	pIEF-r1	pIEF-r2	pIEF-r1	pIEF-r2	pIEF-r1	pIEF-r2
Identified peptides	4803	5011	4505	4467	3237	3238
Identified proteins	376	368	533	529	528	516

Parallel to the consistent identification rates (Table 7), there was a very large overlap of the identified proteins in each replicate analysis (Figure 38, A-C).

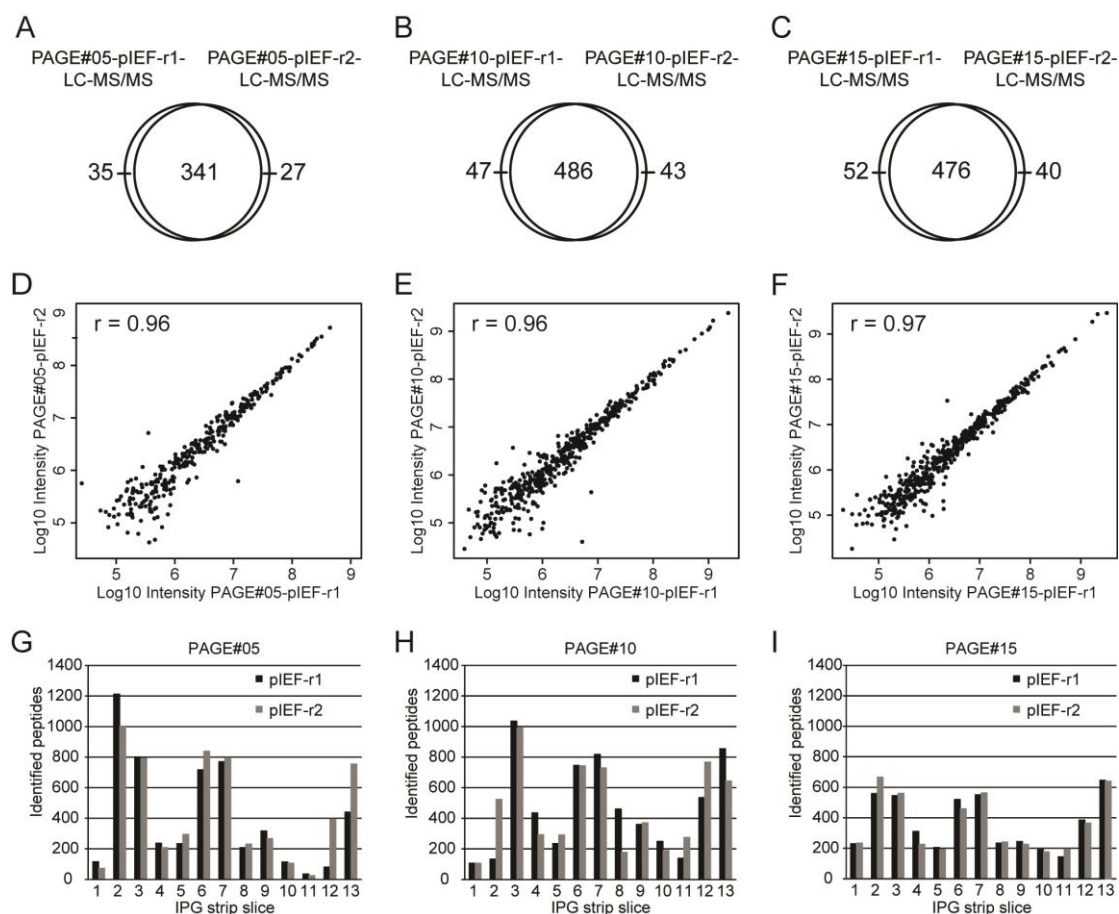


Figure 38. Reproducibility of pIEF.

Overlap of identified proteins across replicate pIEF analyses of tryptic peptides from the same PAGE subfraction (A – C), correlation of intensity of proteins identified in both pIEF replicate analyses (D – F) and distribution of identified peptides on the IPG strips (G - I). Source: (Atanassov and Urlaub 2013).

Only around 10% of the proteins identified in each replicate were unique to it and not identified in the other. Perfect reproducibility would be characterized by a complete overlap

of the identified proteins but this is rarely observed due to the semi-random nature of the peptide sequencing/identification using mass spectrometric analysis (Liu, Sadygov et al. 2004). The identified proteins in each PAGE fraction that were identified in both pIEF replicate analyses showed very high correlation of their MS intensity (Figure 38, D-F). The Pearson's correlation coefficient (r) values for each pair of pIEF replicates was in the range 0.96 – 0.97, indicating very high reproducibility (Ly, Ahmad et al. 2014). Furthermore, a very similar distribution of the identified peptides across the IPG strips was observed (Figure 38, G-I). Differences in the pIEF fraction in which a peptide is identified and the overall distribution of identified peptides can arise as a result of slight differences in the manual fractionation of the IPG strip. Nevertheless, when the sample process is reproducible, peptides should be identified in the same or at least bordering fractions. In this context, the pIEF analysis performed very well as in all three replicate experiments more than 99% of the peptides identified in both pIEF replicates were found in the same or within ± 1 IPG gel slice. Overall, the comparison of replicate pIEF analyses of PAGE fractions showed that the separation is very reproducible.

In the newly established PAGE-pIEF-LC-MS/MS method pIEF comprises the second dimension of separation prior to LC-MS/MS analysis. For this reason it is important to document the overall reproducibility of the 3D separation method. As the PAGE-pIEF-LC-MS/MS method is based on 1D PAGE separation in the first dimension, two separate experiments were performed. In each experiment, the reproducibility of the protein and peptide identifications from the same fractions from three parallel PAGE lanes was examined. The analysis of the parallel fractions was done either using pIEF separation (PAGE-pIEF-LC-MS/MS workflow) or by directly carrying out LC-MS/MS analysis (GeLC-MS/MS workflow). In this manner, the overall reproducibility of PAGE-pIEF-LC-MS/MS can be evaluated and compared to the GeLC-MS/MS approach. For the analysis, the same set of PAGE fractions (PAGE#05, PAGE#10, and PAGE#15) was selected as for the pIEF reproducibility experiment described in the previous section. Each parallel PAGE fraction was assigned as a PAGE replicate (e.g. PAGE#05-r1, PAGE#05-r2 and PAGE#05-r3, etc.). First, the distribution of the identified peptides across the IPG strip in the PAGE-pIEF-LC-MS/MS replicates was evaluated (Figure 39).

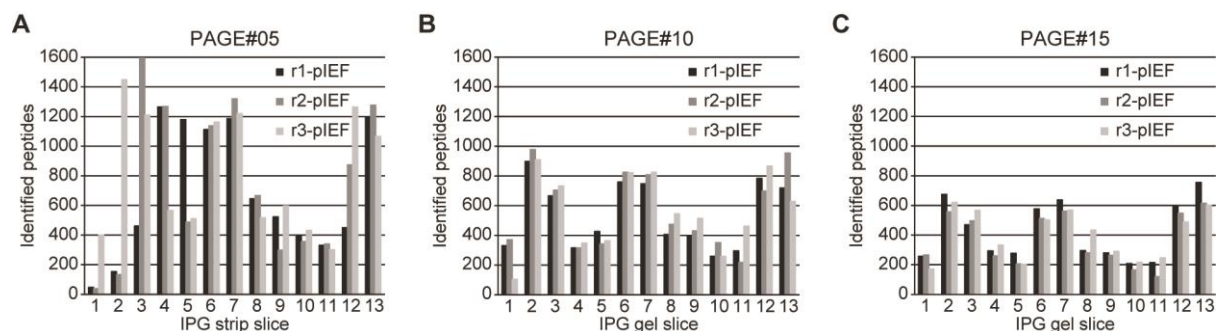


Figure 39. Reproducibility of the distribution of identified peptides across the IPG strip. Replicate pIEF analyses of tryptic peptides from the same PAGE subfraction of three replicate 1D PAGE lanes were performed. Source: (Atanassov and Urlaub 2013).

The peptide distributions from the pIEF analyses of the three replicate PAGE fractions were very similar, except for a difference in the numbers of identified peptides in the acidic end (slices 1–5) of the IPG strip from the pIEF analyses of PAGE#05-r1 and PAGE#05-r2 (Figure 39, A). The remaining slices (6 to 13) showed rather similar numbers of identified peptides. One reason for obtaining results in which the identification rate is consistent in one part of the IPG strip and inconsistent in the other is an improper rehydration at the acidic end of the IPG strip. This would result in underperformance of the IEF separation and differences in the identification rate.

Next, the overall reproducibility of GeLC-MS/MS and PAGE-pIEF-LC-MS/MS was examined by looking at the numbers of identified proteins and peptides in a PAGE lane replicate. The results of the comparison are shown in Table 8.

Table 8. Reproducibility of GeLC-MS/MS and PAGE-pIEF-LC-MS/MS. Numbers of identified peptides and proteins, correlation of protein intensity and percentage of total proteins that are identified in all PAGE slice replicates using either GeLC-MS/MS or PAGE-pIEF-LC-MS/MS. Show correlation is between PAGE replicates 1 and 2 (r1), 2 and 3 (r2), and 3 and 1 (r3). Source: (Atanassov and Urlaub 2013).

	Experiment	GeLC-MS/MS			PAGE-pIEF-LC-MS/MS		
		r1	r2	r3	r1	r2	r3
PAGE #05	Peptides	3188	3241	3120	6557	7513	8569
	Proteins	256	244	250	439	437	473
	Correlation protein intensity	0.89	0.95	0.89	0.67	0.68	0.97
	Total Proteins	303			571		
	In all replicates [%]	67.00			58.14		
PAGE	Peptides	2443	2417	2408	5111	5629	5402

#10	Proteins	315	303	311	551	597	585
	Correlation protein intensity	0.84	0.98	0.83	0.80	0.70	0.95
	Total Proteins	379			774		
	In all replicates [%]	65.17			52.71		
PAGE #15	Peptides	1538	1705	1723	3583	3470	3512
	Proteins	218	233	237	585	593	581
	Correlation protein intensity	0.75	0.98	0.79	0.93	0.94	0.84
	Total Proteins	271			716		
	In all replicates [%]	66.42			63.27		

In both the GeLC-MS/MS and the PAGE-pIEF-LC-MS/MS approaches, there was a high consistency in the numbers of identified peptides and proteins (Table 8, Peptides, Proteins). A single exception was the pIEF analysis of the three PAGE#05 replicates in which 6557, 7513 and 8569 peptides were identified. These differences were most probably a consequence of the dissimilar numbers of identified peptides across the IPG strip (Figure 39, A). Next, the overlap of the proteins, identified in all three PAGE replicates, was examined. As expected, in both the GeLC-MS/MS and PAGE-pIEF-LC-MS/MS analyses, the number of total proteins, i.e. those identified together in all three 1D PAGE replicates, (Table 8, Total proteins) was higher compared with the number of proteins identified in a single replicate analysis (Table 8, Proteins). Overall, the number of proteins identified together in all three PAGE replicates was higher for the PAGE-pIEF-LC-MS/MS analysis compared to the GeLC-MS/MS analysis. At the same time, the percentage of the proteins identified together in all three PAGE replicates that were also identified in all three replicates, was higher for GeLC-MS/MS (65-67%) than for PAGE-pIEF-LC-MS/MS (52-63%). These results indicate that using PAGE-pIEF-LC-MS/MS more proteins are identified from a PAGE fraction while those identifications were slightly less reproducible compared to the GeLC-MS/MS approach. In addition to the protein identification rate, the correlation of protein intensity across PAGE replicates was also examined (Table 8, Correlation protein intensity). Overall the proteins identified using the GeLC-MS/MS approach showed higher Pearson's correlation values, ranging from 0.75 to 0.98, compared with the PAGE-pIEF-LC-MS/MS approach, where the values were between 0.67 and 0.97. Interestingly, in both approaches, two out of the three PAGE replicates showed very high correlation coefficients of the protein intensity. This was the case for all three fractions from replicates PAGE-r2 and PAGE-r3 in the GeLC-MS/MS analysis. In those

fractions the correlation of protein intensity was in the range 0.95 - 0.98. For the PAGE-PIEF-LC-MS/MS analysis, the correlation of protein intensity for fractions #05 and #10 from replicates PAGE-r3 and PAGE-r1 was in the range 0.95 - 0.97. Furthermore the correlation of protein intensity for fraction 15 from replicates 1) PAGE-r1 and PAGE-r2 and 2) PAGE-r2 and PAGE-r3 was 0.93 and 0.94 respectively. These results indicate that besides reproducibility in the numbers of identified peptides and proteins, very high reproducibility of PAGE-PIEF-LC-MS/MS could be achieved.

3.5 Identification of secreted signaling proteins from CPC cell culture medium

Up to this point, a novel 3D separation approach was established and evaluated, a large-scale protein quantification of protein abundance from CPC-2D and CPC-3D was performed and a number of protein species were identified and quantified using a large-scale 2DE analysis. The latter experiments were performed with whole cells, which were pooled together and lysed, resulting in the identification and the quantification of a large number of cellular and extracellular proteins (e.g. Table 1 and Table 3). Nevertheless, the preparation of cells for proteomic sample processing usually involves detachment of cells using trypsin and successive washing prior to cell lysis. Such treatment can remove small signaling molecules and prevent them from being identified in the standard mass spectrometry based proteomic experiments. In addition, secreted signaling proteins have unique half-lives and are also able to trigger feedback mechanisms controlling their own production and secretion making them hard to identify when working with whole cells alone. For this reason it is very beneficial to study the so-called “secretome” of a cell (Hathout 2007). One source of secreted proteins is the cell culture conditioned medium. By starving the cells for a limited period (12-48h), a balance can be achieved between removing the major contaminating proteins from the serum additive and giving the cells enough time to secrete signaling factors without significantly affecting the cell homeostasis (Stastna and Van Eyk 2012). In addition, the use of metabolically labeled cells will allow to distinguish between a contaminant originating from the serum additive and a truly secreted protein as only proteins secreted by the metabolically labeled cell will show the corresponding signal in the mass spectrometer (Colzani, Waridel et al. 2009).

In addition to the whole cell based experiments described in the previous sections, a proteomic study dedicated to the identification of proteins, secreted by CPCs *in vitro*, was performed. In order to facilitate the identification of secreted proteins, conditioned cell culture medium of heavy SILAC labeled CPCs, grown in 2D culture was collected 24h after

serum removal. A large amount of culture medium (around 15ml) was concentrated using TCA precipitation (Chen, Lin et al. 2005) and separated into 10 parallel 1D PAGE lanes. Since many secreted signaling proteins have low MW (Frith, Forrest et al. 2006), after fractionation of the 1D PAGE lanes into 23 slices each, only slices 16-23 from all 10 lanes were digested in-gel and used for further analysis. The PAGE fractions having the same number in each lane were combined (e.g. all ten PAGE#16 fractions), and subjected to in-gel trypsin digestion. The resulting tryptic peptides were separated using pIEF prior to LC-MS/MS analysis. The PAGE-pIEF-LC-MS/MS approach was applied as the method results in deeper proteome coverage and an increase in the numbers of identified proteins and peptides is characteristic of all regions of a 1DE lane (Figure 37). The PAGE-pIEF-LC-MS/MS analysis of the low MW 1DE regions resulted in the identification of more than 1280 proteins, including 146 contaminants (MaxQuant contaminants database). Functional enrichment analysis of the proteins revealed that 22.4% of the proteins belonged to the group of extracellular region (GO:0005576), and 21.1% to the group of cytosol (GO:0005829). It is possible that a large number of extracellular proteins are contaminants originating from the cell culture serum additive. Although the serum is removed after extensive washing, proteins can still stick to the plastic of the flask or attach to cells and later be identified in parallel to the cell secreted proteins (Bunkenborg, Garcia et al. 2010). Nevertheless, only proteins that are secreted by the metabolically labeled cells will show the corresponding heavy signal in the mass spectrometric analysis. Unless there is a source of the light state of the protein, secreted proteins should have very high H/L ratios. When quantification is performed an important parameter that can influence the results is a re-quantification step that is optional in the quantification software, MaxQuant (Cox and Mann 2008). This parameter performs a second round of quantification when the program could detect only one peptide (light or heavy) but not its respective pair. In this manner larger H/L ratios can be recorded. At the same time the increased quantitation rate can result in erroneous quantifications when a peptide is of lower abundance and the missing pair is noise. Figure 40 shows the effect of the re-quantify option the results of protein quantification.

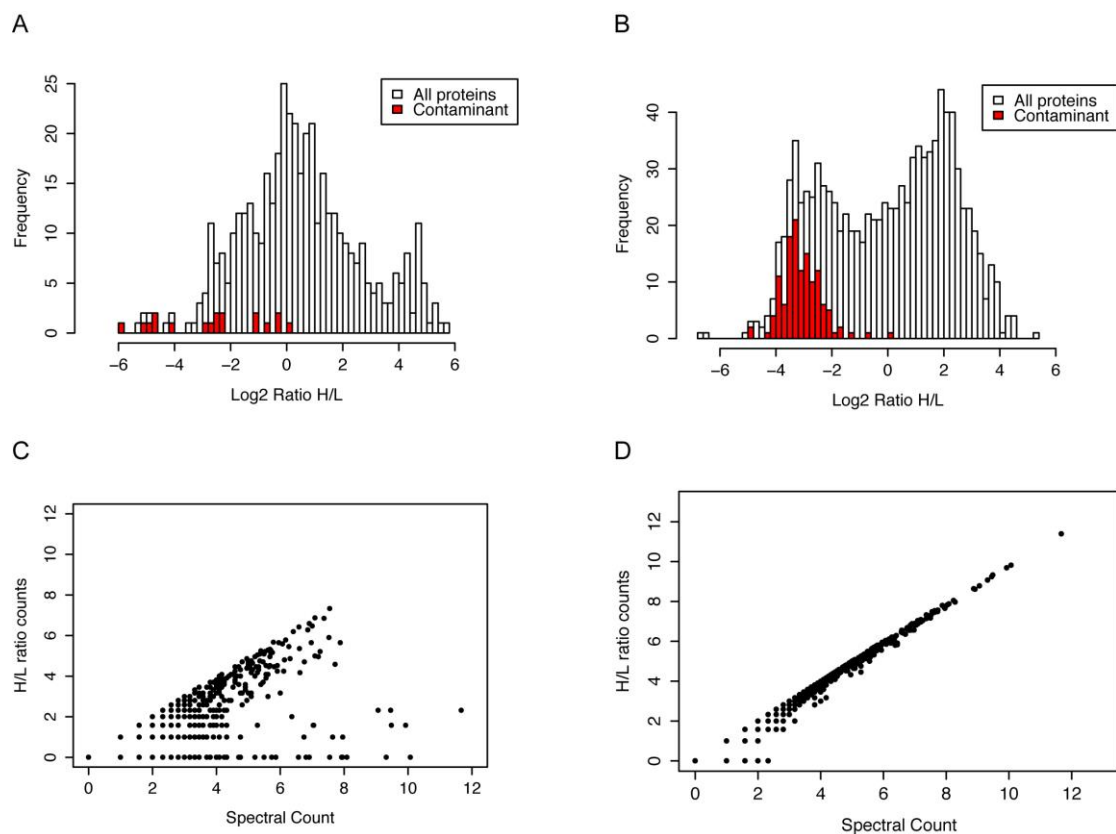


Figure 40. Quantification of proteins from the CPC culture medium. The distribution of protein ratios change depending on whether the re-quantify option of MaxQuant was turned off (A) or on (B). The number of ratio H/L counts (log₂ scale) relative to the spectral count for each protein also changes depending on quantification without (C) or with (D) using the re-quantify option.

When the option is turned off (Figure 40, A), the distribution of protein ratios has its peak at 0. This indicates that (in this quantification manner) the majority of the proteins identified in the culture medium are close to ratios of 1:1. Besides the peak at 0, there is also a second peak in the high (>4) ratio H/L range. GO analysis of the proteins in this region showed enrichment for the terms “extracellular region”, p value 9.8E-7, and “extracellular matrix”, p value 3.2E-6. When the re-quantify option was turned on, the distribution of protein ratios changed drastically (Figure 40, B). The previously characteristic peak at 0 disappeared and instead, two peaks, at log₂ ratios between -3 and -4 and around 2, become apparent. Proteins showing ratio H/L higher than 2 were enriched for the GO terms “extracellular region”, p value 1.4E-10, and “extracellular matrix”, p value 6.3E-10. There were other differences between the quantification results shown in Figure 40, A and B. First, more proteins were quantified using the re-quantify option – 1024 compared to 446 without re-quantify. In addition more contaminant proteins (MaxQuant database) were quantified using re-quantify (129) compared to 18 when the option was off. One of the major differences was the relationship between the spectral counts (number of MS/MS

sequencing events) and ratio H/L counts per protein (Figure 40, C and D). When the re-quantify option was turned off, there were a number of proteins that were identified by a large number of MS/MS spectra but had very few H/L ratios or were not quantified at all. On the contrary when the option was turned on, the number of ratio counts was proportional to the number of MS/MS spectra. This shows that overall more quantification events per protein were recorded when the re-quantify option was used.

While there are differences in the quantification and identification of proteins in the culture medium of SILAC labeled cells depending on the use of the re-quantify option of MaxQuant, proteins termed as secreted (characterized by GO terms related to the extracellular space) have high H/L ratios. Two interesting questions are what is the benefit of performing PAGE-pIEF-LC-MS/MS analysis on the low MW regions and what is the most reliable way to identify/quantify secreted proteins. In order to examine the novel protein identifications gained in this experiment, all identified genes which encoded for proteins and were identified in any of the whole cell based experiments in the study were combined, namely the 1DE analyses of CPC-2D/CPC-3D biological replicates 1 and 2 and the 2DE analysis of CPC-2D/CPC-3D biological replicate 1. This combined set of identifications resulted in a total of 4900 genes and their protein products. For comparison with this list, all proteins identified from the CPCs condition medium and quantified (using the re-quantify option) with log₂ ratio H/L above -1 (Figure 40, B) were selected. When a log₂ ratio H/L value of -1 was used as a cutoff, only two contaminants (MaxQuant contaminant database) were included in the list of proteins. The total selected proteins with ratio above -1 were 641 proteins. Overall 51 from the selected 641 proteins were not identified in any of the whole cell based experiments. These 51 proteins were used for GO analysis. All proteins that were characterized by the GO term “extracellular region” were selected. Finally, for each of those proteins, the following identification/quantification characteristics were extracted: log₂ ratio (with and without re-quantify), ratio count (with and without re-quantify) and number of peptides by which the protein was identified. The set of proteins that were not identified in the whole cell based studies, showing log₂ ratio H/L of -1 or higher, and being classified as extracellular can be found in Table 9.

Table 9. List of secreted proteins identified from CPC cell culture medium and not in any of the whole cell based analyses in this study.

Re-quantify option (MaxQuant)		Log 2 Ratio H/L		Ratio Count		Pep- tides
		off	on	off	on	
Gene ID	Protein name					
IGFBP6	insulin-like growth factor binding protein 6	5.29	4.30	12	35	9
IGFBP7	insulin-like growth factor binding protein 7	-2.94	1.43	12	30	8
METRNL	meteorin, glial cell differentiation regulator-like	3.18	2.49	3	17	8
PDGFC	platelet derived growth factor C	NaN	2.25	0	9	7
VEGFC	vascular endothelial growth factor C	NaN	2.15	0	14	7
STC1	stanniocalcin 1	4.79	3.27	2	10	6
BDNF	brain-derived neurotrophic factor	NaN	3.61	0	5	4
PENK	proenkephalin	NaN	3.36	0	4	4
STC2	stanniocalcin 2	5.11	3.38	3	7	4
EFEMP2	EGF-containing fibulin-like extracellular matrix protein 2	NaN	1.28	0	3	3
TGFB2	transforming growth factor, beta 2	2.77	3.08	2	4	3
ADAMTS 5	ADAM metallopeptidase with thrombospondin type 1 motif, 5	NaN	1.87	1	2	2
ULBP2	UL16 binding protein 2	NaN	1.95	0	4	2
ANG	angiogenin, ribonuclease, RNase A family, 5	NaN	2.34	0	5	2
IFI30	interferon, gamma-inducible protein 30	4.10	3.73	2	9	2
LY96	lymphocyte antigen 96	NaN	2.56	0	5	2
NBL1	neuroblastoma, suppression of tumorigenicity 1	2.10	1.88	3	4	2
PLAC9	placenta-specific 9	5.48	4.44	2	4	2
SECTM1	secreted and transmembrane 1	NaN	1.96	0	6	2
SEMA3B	sema domain, immunoglobulin domain (Ig),	NaN	2.71	0	2	2
SLIT1	slit homolog 1 (Drosophila)	NaN	1.36	0	6	2
VEGFA	vascular endothelial growth factor A	NaN	0.42	0	4	2
ADAMTS 2	ADAM metallopeptidase with thrombospondin type 1 motif, 2	NaN	2.05	0	2	1

COLEC10	collectin sub-family member 10 (C-type lectin)	NaN	-0.90	0	2	1
INHBA	inhibin, beta A	NaN	1.59	0	2	1
SERPINI1	serpin peptidase inhibitor, clade I (neuroserpin), member 1	NaN	2.37	0	2	1

Overall, all except three of the identified proteins were identified by more than one peptide. Only six proteins had calculated ratio H/L when the re-quantify option was turned off. Protein H/L ratio values in both quantification approaches were in agreement with each other except for IGFBP7 for which the values were opposite (-2.94 without re-quantify and 1.43 with re-quantify). Overall, the ratio H/L values were lower compared to when the re-quantify option was turned off.

The application of PAGE-pIEF-LC-MS/MS to the identification of secreted proteins from the culture medium enabled the identification of lower abundance proteins and proteins that could not be detected from whole cell analyses. The analysis revealed the presence of a number of signaling proteins in the CPC-2D culture medium. These included two insulin-like growth factor binding proteins, three mitogens (substances that promote cell division), namely PDGFC, VEGFC and VEGFA, and TGFB2, a signaling factor known to be involved in *in vivo* chondrogenesis (Kulyk, Rodgers et al. 1989).

3.6 Optimization of the pIEF fractionation scheme

The results presented so far show that pIEF is a very efficient method for peptide separation and when combined with other separation techniques (e.g. as part of PAGE-pIEF-LC-MS/MS) achieves deeper proteome coverage and the identification of low abundant proteins (e.g. secreted proteins from cell conditioned medium). From a methodological point of view it is interesting to examine whether the fractionation scheme can be optimized in order to achieve similar numbers of identified peptides/proteins but at a reduced time.

The separation of peptides using pIEF and the distribution of identified peptides over the IPG strip is governed by differences in the pI of the peptide. Upon fractionation of the IPG strip, a single peptide can be identified in one or multiple fractions. When the MS intensity of a peptide is taken into consideration, the slice in which the peptide intensity is at the maximum is also the slice in which the peptide is focused due to its pI. If the same peptide is identified in other slices, this can be considered as carryover. Moreover, taking into account the peptide identifications in each pIEF fraction that have only the maximum intensity relative to other fractions, reveals the number of peptides identified in the

respective fraction that are not a carryover from other fractions. Considering a typical pIEF analysis of a complex protein mixture using a 18cm IPG strip, there are two regions on the IPG strip that contain low number of peptides: slices 10-12 and 25-28 (Figure 41, A). In addition to the lower number of identified peptides, closer inspection of the identity of those peptide shows that most of them are not unique to those IPG strip slices (Figure 41, A, Maximum intensity), as such peptides would have maximum intensity in those slices.

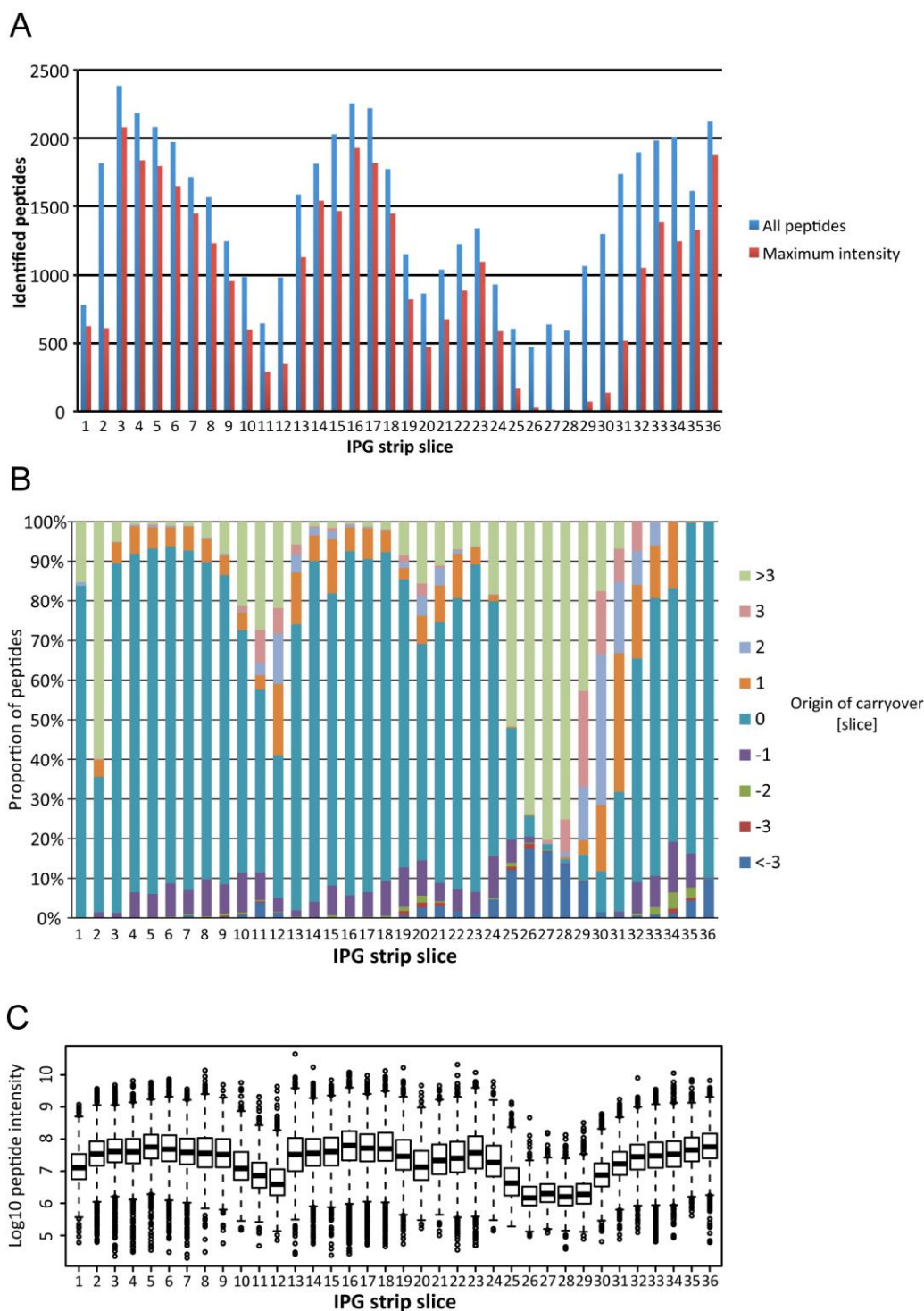


Figure 41. Analysis of the distribution of peptides across an 18cm IPG strip.

A) Distribution of all identified peptides and peptides with maximum intensity. B) Source of peptide carryover. Value indicates the distance, in slices, from the slice where the maximum intensity of the peptide was identified. Positive values indicate higher slice number, negative values indicate lower slice number, 0 indicates same slice. C) Boxplot of the intensity of all identified peptides in the respective IPG strip slice.

Overall, the proportion of the peptides having maximum intensity, compared to all peptides identified in the slice was different in the acidic portion of the strip (slices 3 to 24),

compared to the basic one. In the acidic regions (except slices 1 and 2) the proportion of peptides having maximum intensity was constant and around 80% (Figure 41, A). In contrast, slices 25 to 30 contain very few peptides with maximum intensity. This shows that in this region almost all identified peptides were a carryover from other slices. In the following slices (31 to 34), the proportion of peptides showing maximum intensity was lower compared to the acid end, while those slices contained still a large number of peptides with maximum intensity. An optimization of the fractionation scheme would involve either cutting the IPG strip in slices of different length or combining fractions in specific manner. The proper combining scheme would pool together slices that contain the most maximum intensity peptides with those slices that contain carryover of those peptides. The second important criteria are the maximum number of peptides that can be identified from an IPG strip slice with the chosen LC-MS/MS parameters (30 min 3-33% ACN gradient, Orbitrap XL mass spectrometer). In the experiments described in this work this value was up to around 2500 peptides per IPG slice (Figure 41, A).

Identifying carryover of peptides can be carried in a straightforward manner when the peptide intensity is taken into consideration. A peptide identified in more than one slice will have maximum intensity in one slice and lower than maximum intensity in the other slices. Based on this, two things can be discovered: 1) whether the peptide identification is a carryover (peptide intensity in IPG strip fraction is not the highest from all fractions where the peptide was identified) and 2) where does this carry over come from. The latter can be inferred from the difference between the number of the fraction where the peptide was identified and the fraction in which the peptide shows maximum intensity. Figure 41, A indicates how many peptides are carryover (e.g. almost all peptides in fractions 27 and 28 are carryover from other slices). Figure 41, B shows where this carryover comes from. The overall results show that in the acidic and middle portions of the IPG strip the majority peptides are not carryover (0 values). Those peptides that are carryover from other slices are either coming from previous slice (value of -1) or from the next slice (value of +1). The proportions of those values numbers are similar for the acidic and middle part of the IPG strip. Slice 12 is a good example of how fractions should be combined for optimization (Figure 41, B). In this fraction approximately 45% of the identified peptides are focused in this slice (maximum intensity) and the remaining peptides are carryover. Around 5% of the peptides come from the previous slice and the remaining 50% of the identified peptides are carryover from peptides from the next slices (1, 2, 3 or more slices after this one). Slice 12 can be combined with slice 13 because in this manner carryover would be diminished and projected number of identified peptides would be less than 2500. In the basic portion of the IPG strip (slices 30-36) the majority of the carryover comes from following slices (higher percentage for positive values, Figure 41, B.). This indicates that optimization should be

carried out by combining a slice with the next ones. The final aspect of the optimization is whether some slices can be skipped and not analyzed altogether. As pointed above, fractions 26 to 27 contain only carryover peptides. If they are to be combined, this will result to several slices being combined into one fraction. This might have drawbacks for the downstream analysis as there are contaminants coming from the IPG gel. The peptide abundance (inferred from the peptide MS signal) in these fractions shows what amount of peptide material would be lost if these fractions (e.g. 26 to 29) are not analyzed (Figure 41, C). The distribution of peptide intensities in each slice shows that slices 26 to 29 contain peptides with intensity that is much lower (median log₁₀ intensity of 6.3), compared to slices from the more basic slices (median log₁₀ intensity of 7.5). This indicates that the carryover peptides are more than 10 times less abundant compared to the slices where the carryover comes from. This also indicates that skipping these fractions in the analysis should not be detrimental to the peptide identification rate. Based on the following criteria: proportion of peptides showing maximum intensity, proportion of carryover peptides, direction of the carryover, the distribution of intensities of the identified peptides in a fraction, and a peptide target of no more than 2500 peptides the following fractionation scheme was proposed – Table 10.

Table 10. Optimization of the fractionation scheme for 18 cm pIEF. Strip is cut into 36 slices that are directly analyzed (Standard fractions) or combined in selected manner (Optimized fractions) prior to LC-MS/MS analysis. In both fractionation schemes samples 1 to 8 are analyzes separately.

Optimized fractions	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Standard fractions	9 10	11 12 13	14	15	16	17	18 19	20 21 22	23 24 25	30 31 32	33 34	35	36	

Based on this scheme described in Table 10, the total number of fractions for analysis was significantly reduced from 36 fractions to 22. Slices 26-29 from strip were not analyzed or combined with other fractions in the optimized scheme. The results of the direct comparison between both approaches are shown in Figure 42.

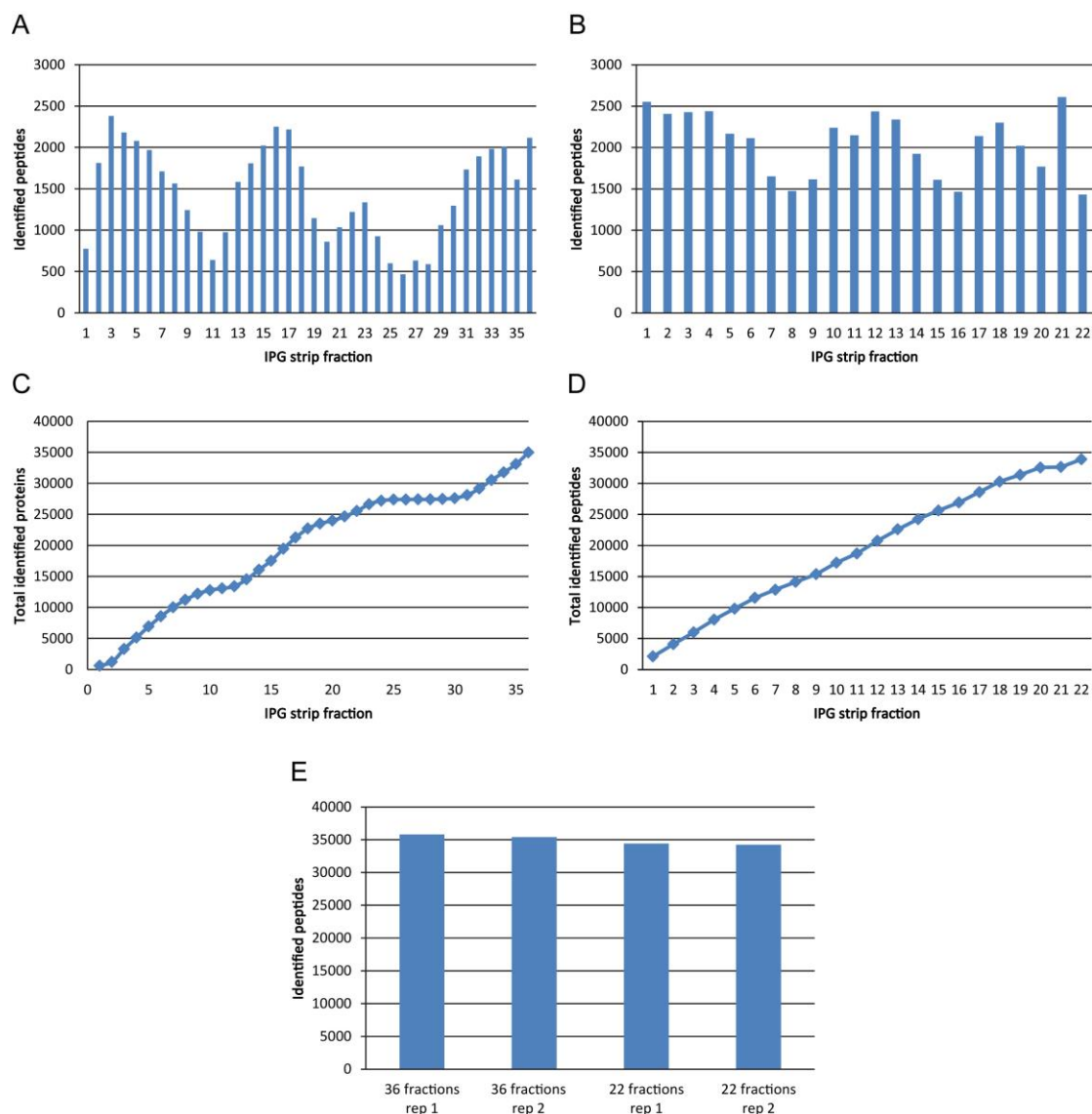


Figure 42. Reproducibility of the standard and optimized 18cm pIEF fractionation scheme. Direct comparison of peptide distribution (A and B) and total identified peptides (C and D) using 36 or 22 fractions of 18 cm IPG strip. E) Reproducibility of the numbers of identified peptides from the standard and optimized fractionation scheme.

Compared to the 36 fractions analysis, the optimized fractionation method resulted in much more consistent number of identified peptides per fraction (Figure 42, A and B). In the 36 fractions approach there were a number of slices showing as low as 500 identified peptides, while in the 22 fraction approach no fraction produced less than 1500 peptide identifications (Figure 42, A and B). The number of identified peptides per slice is not a direct measure of increased peptide identification rate since they can be a carryover from other slices. This is most clearly visible when the total number of identified peptides per slice is taken into consideration (Figure 42, C and D). For the analysis with 36 slices, there were slices whose analysis contributed significantly to the total number of peptide identifications (e.g. slices 3-8, Figure 42, C). At the same time, other regions of the IPG

strip (e.g. fractions 10-12 and 26-30) practically did not contribute to any new/additional peptide identifications. Using the optimized fractionation scheme those peptide poor regions were virtually eliminated and almost all IPG strip fractions provided large numbers of new peptide identifications resulting in a consistent increase in the total number of identified peptides (Figure 42, D). Overall, the total number of identified peptides 36 and 22 fraction experiments was reproducible. Figure 42, E shows the total number of identified peptides identified in repeated analysis using either the 36 or 22 fractions approach. The former approach resulted in around 35000 peptides while the latter in around 34000. This slight decrease can be explained by the lower number of fractions used for analysis. Nevertheless the disadvantage of this slightly lower number of identified peptides is compensated by the significant gains in peptide identifications per total analysis time.

4 Discussion

4.1 Quantitative proteome differences between CPCs grown on 2D plastic and in 3D alginate

One of the main aims of this work was to quantify the relative differences of protein abundance between CPCs grown in monolayers, in 2D culture flasks, and in 3D alginate beads. The transfer of cells from a 2D to 3D culture can have an immense effects on the phenotype of the cells (Baker and Chen 2012). An example for this is the initiation of chondrogenic differentiation upon the encapsulation of CPCs and MSCs in general, into alginate beads (Ma, Hung et al. 2003, Koelling, Kruegel et al. 2009). One of the main differences between 2D and 3D culture is the adhesive nature of the environment. Treated polystyrene efficiently binds proteins from the cell culture media serum additive (Ryan 2008). This creates a flat and very adhesive surface to which cells bind and spread. On the contrary, alginate – a polysaccharide produced by brown algae (Lee and Mooney 2012), has very poor binding to serum proteins (Smetana 1993). This prevents the cells from engaging into adhesion with the 3D culture surface which results in a round cellular shape. Since cells can feel the stiffness of their extracellular substrate (Discher, Janmey et al. 2005), the spread phenotype of stiff plastic surface is usually associated with the presence of focal adhesions (Burrige, Fath et al. 1988) and stress fibers. These traits are not characteristics of the cellular phenotype in a 3D cell culture (Cukierman, Pankov et al. 2001). The process of changes of the cellular phenotype or behavior in response to such external physical stimuli (e.g. the effect of 2D vs 3D cell culture) is called mechanotransduction (Guilak, Cohen et al. 2009). In this respect, there were a number of proteins that were downregulated upon culturing of CPCs in 3D alginate culture. Overall, downregulated proteins were associated with cell-matrix adhesion, which was in line with the changes induced by detaching the cells from the stiff 2D surface and transferring them to the 3D alginate culture. Three of the proteins that were downregulated upon culturing of CPCs in alginate were integrins. Integrins act as linkers between the cytoskeleton and the ECM (Hynes 2002) and in this work it was shown that integrins A2, A3 and A6 were downregulated upon culturing of CPCs in alginate. Interestingly, it has been shown that integrin internalization is enhanced when cells are transferred from stiff to soft substrate (Du, Chen et al. 2011). While in the study by Du, Chen et al. integrin trafficking was examined after 2 h of transfer to soft substrate, in this study, quantitative analysis was performed after 3 weeks of cell culturing in alginate beads. Alginate is between 10^5 and 10^7 times softer than polystyrene plastic (Higuchi, Ling et al. 2013). It is possible that the short

term (2h) internalization of integrins and altered signaling (Du, Chen et al. 2011) on soft substrata is translated into changes of integrin gene expression in long term culturing like the one used in this study. In addition to changes of integrin gene expression, there were a number of proteins associated with actin cytoskeleton, which were also downregulated upon CPC culturing in alginate. These included filamin-C, a bridging protein that can interact with both integrins and actin (Loo, Kanner et al. 1998), protein LIMA1 (eplin), an actin associated protein which inhibits actin filament depolymerization and cross-links filaments in bundles (Maul and Chang 1999) and EZR (ezrin), a protein which is associated with cell adhesion and interacts with the focal adhesion kinase (Pouillet, Gautreau et al. 2001). Changes in the relative abundance of these proteins between CPC-2D and CPC-3D reflect changes in the cellular phenotype resulting from cell culturing on a stiff 2D surface compared to a 3D soft matrix.

Parallel to these changes, two MSCs markers were downregulated upon culture in 3D alginate: ALCAM (CD166) and hyaluronan receptor (CD44). The downregulation of those proteins presents evidence for the loss of mesenchymal phenotype by the CPCs and switch to a chondrogenic phenotype. Decrease in the expression of both proteins has been observed during the chondrogenic differentiation of MSCs (Lee, Choi et al. 2009).

The induction of chondrogenesis of CPCs upon culturing in alginate beads has so far only been shown by an increase in the expression of a collagen type II (Koelling, Kruegel et al. 2009). This work provides further evidence for ongoing chondrogenesis upon encapsulation of CPCs in alginate. This evidence comes from the set of proteins that were upregulated in CPC-3D compared to CPC-2D. Multiple proteins associated, *in vivo*, with the chondrocyte PCM, a narrow region of ECM that surrounds the chondrocyte (Poole 1997), were upregulated. One prominent example of a protein associated with the PCM is collagen type VI (Poole, Ayad et al. 1988). This protein is a heterotrimer of proteins COL6A1, COL6A2 and COL6A3; in this study, all three proteins were found to have higher abundance in CPC-3D compared to CPC-2D. In addition to collagen type VI, the following proteins, associated with the PCM were upregulated: laminins (Durr, Lammi et al. 1996, Kvist, Nyström et al. 2008), nidogens (Kruegel, Sadowski et al. 2008, Kvist, Nyström et al. 2008) and collagen type IV (Kvist, Nyström et al. 2008). In addition to these proteins, transforming growth factor-beta-induced protein ig-h3 (TGFB1) was also found to be upregulated in CPC-3D. This protein was identified as part of the chondrocyte PCM in a proteomic study focused on PCM components (Zhang, Jin et al. 2011). These results show that when CPCs are embedded in alginate beads they upregulate the protein components of a functional PCM. It has been shown that the function of the PCM is indispensable to that of the chondrocyte in articular cartilage (Poole 1997). The PCM is essential for transduction of biochemical and biomechanical signals (Guilak, Alexopoulos et al. 2006). In

addition, the development of a PCM during the chondrogenic differentiation of MSCs positively affects the cellular resistance to deformation (Twomey, Thakore et al. 2014) which is essential in the context of articular cartilage (Guilak, Alexopoulos et al. 2006). Although CPCs have been associated with the presence of regenerative tissue *in vivo* (Tesche and Miosge 2005), it remains to be shown that CPCs can generate an ECM with high physical resistance (Koelling, Kruegel et al. 2009). While testing the ability of CPCs to produce such matrix would involve a number of experiments, the work presented here shows that CPCs induced towards chondrogenic differentiation upregulate protein components of the PCM, which can further influence the development and the response of the cells to mechanical tension (Steward, Wagner et al. 2013).

4.2 Protein identification and quantification using 2DE

In addition to the standard 1DE based analysis on proteome changes between CPC-2D and CPC-3D, a 2DE separation method was employed. The performance of the 2DE approach was evaluated in comparison to that of the 1DE separation approach. One of the surprising results from this experiment was the total number of identified proteins from the whole 2DE gel – around 2800. This value was very low compared to what is expected from such a large scale separation platform, large number of fractions and the use of a modern mass spectrometer for peptide/protein identification. Closer examination of the results showed that a large number of 2DE gel fractions (more than 400) did not contain new proteins, but either carryover of proteins identified in other 2DE slices or modified versions/isoforms of already identified proteins. This indicates that analyzing the protein content of a whole 2DE gel is not beneficial for the identification of a high number of proteins. The number of identified proteins from the whole gel was in contrast to the expected number of identifiable proteins from related 2DE studies. In some case more than 5000 different protein spots have been visualized (Hoving, Voshol et al. 2000). However, it is not clear how many of those proteins spots correspond to protein species or isoforms. In addition, when only protein spots are selected for LC-MS/MS analysis, the depth of proteome coverage is not significant (Gygi, Corthals et al. 2000). An approach to circumvent this limitation is to select for LC-MS/MS analysis, regions of the 2DE gel without protein spots, in which case low abundant proteins can be identified (Fey and Larsen 2001). In the experiment presented here, the whole 2DE gel was cut and analyzed, thus proteins from both categories of proteins (with or without characteristic spot) should have been identified. One of the proposed reasons for the low number of identified proteins from the whole 2DE gel is the spread of proteins across the 2DE gel, leading to redundancy in the identified proteins across several 2DE gel fractions. The redundancy of protein

identifications can be very high and the presence of a large number of proteins in a fraction can mask lower abundance proteins and prevent their identification. Nevertheless, due to the high separation power of the technique, 2DE is able to separate and isolate multiple protein species helping their identification (Rabilloud, Vaezzadeh et al. 2009). Similar results were observed in this study. Using 2DE separation, multiple protein species could be detected in this study. In comparison, the evidence for the existence of those species was inconclusive when using the data from the 1DE analysis alone. One example of a protein for which multiple protein species were present was vimentin. Interestingly, post-translational processing (cleavage) of vimentin has been observed in a number of studies. For example, appearance of cleavage products has been observed in high density cell cultures (Ben-Ze'ev, Babiss et al. 1986). Vimentin products have also been observed in experiments related to cellular resistance to apoptosis and sheer stress (Buchmaier, Bibi et al. 2013) and in healthy and osteoarthritic chondrocytes (Lambrecht, Verbruggen et al. 2008). These reports indicate that the cleavage of VIM is a process which is characteristic to a number of cells and conditions rather than an experimental artifact. Cleavage of the protein has also observed in this study and might be associated with the drastic changes of the CPC shape from ellipsoid (CPC-2D) to round (CPC-3D) and the corresponding effects on the cellular cytoskeleton.

In addition to the identification of multiple protein species, due to its separation power, 2DE can achieve higher confidence in the identification of protein isoforms. As shown in this study, the identification of isoforms is especially challenging due to the high sequence similarity of protein isoforms, the low number of peptides that can distinguish two isoforms and the low confidence of the achieved identifications. Using positional information from the 2DE gel and the predicted pI and MW, more information can be obtained regarding the presence / absence of a protein isoform.

A comparison between the characteristics of the proteins identified using 2DE and 1DE revealed that, the 2DE approach identified less basic proteins as proportion of all identified proteins, compared to the 1DE approach. Second, it appears that 2DE has a bias against very hydrophobic membrane proteins which is in agreement with previous reports (Rabilloud 2009). Proteins with high GRAVY scores that were identified with 1DE but not identified using 2DE were exclusively hydrophobic membrane proteins.

Despite these drawbacks, the use of 2DE resulted in the identification and quantification of a number of protein species. These findings, in addition to other reports (Thiede, Koehler et al. 2013) show that 2DE indeed has a special niche in proteomic experiments (Rogowska-Wrzesinska, Le Bihan et al. 2013) which has not been taken over by separation platforms with less separation dimensions (e.g. 1DE).

4.3 Establishment and evaluation of PAGE-pIEF-LC-MS/MS

In this work, in addition to the large-scale analyses based on 1DE and 2DE a novel 3D separation approach, namely PAGE-pIEF-LC-MS/MS was established and evaluated. The novel method utilizes a large separation space based on protein separation by MW and peptide separation by pI prior to LC-MS/MS analysis. Due to the separation power, a deeper coverage of the complex proteome sample is achieved than either 2D separation methods, used for comparison: GeLC-MS/MS (1DE followed by LC-MS/MS analysis) and 18 cm pIEF-LC-MS/MS. The method benefits from the power of 1D PAGE to efficiently separate proteins and concentrate them in a single PAGE fraction. The proteins are then trypsin digested and the resulting peptides, efficiently separated using pIEF. One of the main advantages of this method is that it combines two orthogonal peptide separation techniques (pIEF and RP-HPLC). The orthogonality of the separation was inferred from the lack of correlation between the LC retention time and the peptide pI (position on the IPG strip) as seen in this and other studies (Cargile, Talley et al. 2004). Yet, this orthogonal separation is on a small to medium-scale and can easily be applied to several PAGE fractions of interest. From a technical perspective, the separation by 1DE and pIEF is easily interfaced without the need to exchange buffers. Nevertheless, caution is needed when several PAGE fractions are combined in which case salts that are present in the sample, might compromise the IEF process. In this regard pIEF will not tolerate high concentration of Ca^{2+} , commonly added to aid trypsin digestion. The results of this work show that pIEF separation is very reproducible, despite relying on several steps that can induce variability. These steps include the IPG strip overnight rehydration, IEF, manual fractionation, extraction and desalting. Nevertheless, upon proper handling the separation process, the number of identified proteins and peptides can be very reproducible. Importantly, performing pIEF as a second separation dimension also could significantly increase the protein sequence coverage and the number of identified proteins from all PAGE subfractions. This shows that the PAGE-pIEF-LC-MS/MS approach can also be applied, very successfully, to the analysis of specific PAGE regions. This advantage was further employed in the identification of low abundant proteins, secreted by CPCs.

4.4 Identification of secreted proteins from CPC culture medium

Secreted signaling proteins are important components of the cell-cell communication. They can act in an autocrine or paracrine manner and influence the cell fate and differentiation (Scheel, Eaton et al. 2011). Signaling proteins have unique half-lives outside the cell which can facilitate their identification. In addition their concentration increases with the increase

in cell numbers in the culture and in time. When these factors are taken into consideration it seems advantageous to study the protein content of the conditioned cell culture medium (Hathout 2007). For such studies to be feasible, the cell culture medium should not contain serum additive or the serum concentration should be reduced significantly (Meissner, Scheltema et al. 2013). Removing or reducing the serum additive amount might affect the cell viability and behavior, but cell medium collection time up to 24 or 48 hours can be a tradeoff between acceptable cellular phenotype and time scale of protein secretion (Stastna and Van Eyk 2012). For the experiments presented in this work, the protein content of the culture media of CPCs grown for 24h without serum additive, was extracted and separated using 1D PAGE. Low MW regions of the PAGE gel were selected for analysis since secreted signaling proteins have low MW (Frith, Forrest et al. 2006). By combining medium concentration using TCA precipitation, 1DE analysis in multiple lanes and pooling of parallel PAGE slices it is possible to significantly increase the starting material, which in this case was close to 15 ml of culture medium. The identification of a number of secreted proteins was an indication of the power of the PAGE-pIEF-LC-MS/MS approach. A number of secreted proteins, identified from the cell culture medium using the PAGE-pIEF-LC-MS/MS approach, were not identified from the whole cell based experiments, in which close to 5000 proteins were identified altogether. This indicates that the signaling proteins are either removed from the cells during sample preparation and/or that they are of very low abundance. Quantification of proteins, identified in the cell culture medium, proved to be powerful step in discriminating between as secreted proteins and proteins originating from the serum additive. Secreted proteins could be easily pinpointed using high H/L ratios and GO analysis. The software used in this study, MaxQuant, allowed for two types of quantification to be performed – with or without the re-quantify option. The results of both analyses (with and without re-quantify) were very different so the results must be interpreted with caution. If other quantification software is used, the quantitation results should be inspected before proteins are chosen as candidates for confirmative analyses. One way to test the quantification results used here was to look at the proportion of quantification events in comparison to spectral counts. If this proportion is low, this means that the software is able to identify the proteins present in the sample but is not able to quantify them. In this study, using the re-quantify option resulted in the identification of more secreted protein candidates compared to the case when the option was turned off. Interestingly, in a study focused on the quantification of cell secreted proteins in blood plasma, protein quantification was also carried using MaxQuant but the use of the re-quantify option was discouraged (Boersema, Geiger et al. 2013). Notably, in that study, the relationship between numbers of identified and sequenced peptides was not taken into consideration.

The application of PAGE-pIEF-LC-MS/MS approach to the low MW regions resulted in the identification of a number of secreted proteins which are generally accessible only through dedicated enrichment techniques such as antibody arrays (Ranganath, Levy et al. 2012). In addition some of the identified secreted proteins can act in a paracrine / autocrine fashion. For example VEGF can enhance the CPC migration rate (Gerter, Kruegel et al. 2012) while TGFB2 is a inducer of chondrogenesis (Kawakami, Rodriguez-León et al. 2006).

4.5 Optimization of pIEF fractionation

In addition to the evaluation of the PAGE-pIEF-LC-MS/MS approach and its application to the discovery of low abundant proteins, effort was made to optimize the fractionation scheme of the IPG strip, in order to reduce the analytical time. In this line, the fractionation scheme for the 18 cm pIEF strip was optimized in order to avoid peptide poor regions and to combine fractions that contain large number of carryover peptides. In this manner a comparable number of peptides can be identified in an optimized fractionation scheme comprised of 22 fractions, instead of 36, resulting in a significant reduction in the analytical time. This optimization of the fractionation scheme remains to be applied to the 7 cm IPG strips, which should decrease the analytical time of the PAGE-pIEF-LC-MS/MS approach while preserving its power for deep proteome profiling.

5 References

- Abian, J., A. J. Oosterkamp and E. Gelpí (1999). "Comparison of conventional, narrow-bore and capillary liquid chromatography/mass spectrometry for electrospray ionization mass spectrometry: practical considerations." Journal of mass spectrometry **34**(4): 244-254.
- Aebersold, R. and M. Mann (2003). "Mass spectrometry-based proteomics." Nature **422**(6928): 198-207.
- Akiyama, H., M.-C. Chaboissier, J. F. Martin, A. Schedl and B. de Crombrughe (2002). "The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6." Genes & development **16**(21): 2813-2828.
- Atanassov, I. and H. Urlaub (2013). "Increased proteome coverage by combining PAGE and peptide isoelectric focusing: Comparative study of gel-based separation approaches." Proteomics **13**(20): 2947-2955.
- Aulin, C., K. Bergman, M. Jensen-Waern, P. Hedenqvist, J. Hilborn and T. Engstrand (2011). "In situ cross-linkable hyaluronan hydrogel enhances chondrogenesis." J Tissue Eng Regen Med **5**(8): e188-196.
- Baker, B. M. and C. S. Chen (2012). "Deconstructing the third dimension: how 3D culture microenvironments alter cellular cues." J Cell Sci **125**(Pt 13): 3015-3024.
- Baker, B. M. and C. S. Chen (2012). "Deconstructing the third dimension: how 3D culture microenvironments alter cellular cues." Journal of cell science **125**: 3015-3024.
- Baldwin, M. A. (2004). "Protein Identification by Mass Spectrometry: Issues to be Considered." Molecular & Cellular Proteomics **3**(1): 1-9.
- Bantscheff, M., M. Schirle, G. Sweetman, J. Rick and B. Kuster (2007). "Quantitative mass spectrometry in proteomics: a critical review." Anal Bioanal Chem **389**(4): 1017-1031.
- Barrabes, S., A. Sarrats, E. Fort, R. De Llorens, P. M. Rudd and R. Peracaula (2010). "Effect of sialic acid content on glycoprotein pI analyzed by two-dimensional electrophoresis." Electrophoresis **31**(17): 2903-2912.
- Ben-Ze'ev, A., L. E. Babiss and P. B. Fisher (1986). "Cleavage of vimentin in dense cell cultures. Inhibition upon transformation by type 5 adenovirus." Exp Cell Res **166**(1): 47-62.
- Benya, P. D. and J. D. Shaffer (1982). "Dedifferentiated chondrocytes reexpress the differentiated collagen phenotype when cultured in agarose gels." Cell **30**: 215-224.
- Bi, W., J. M. Deng, Z. Zhang, R. R. Behringer and B. de Crombrughe (1999). "Sox9 is required for cartilage formation." Nat Genet **22**(1): 85-89.
- Bianchi-Smiraglia, A., D. Kunnev, M. Limoge, A. Lee, M. C. Beckerle and A. V. Bakin (2013). "Integrin- β 5 and zyxin mediate formation of ventral stress fibers in response to transforming growth factor β ." Cell Cycle **12**(21): 3377-3389.
- Biemann, K. (1988). "Contributions of mass spectrometry to peptide and protein structure." Biomed Environ Mass Spectrom **16**(1-12): 99-111.
- Boersema, P. J., T. Geiger, J. R. Wisniewski and M. Mann (2013). "Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples." Mol Cell Proteomics **12**(1): 158-171.
- Branca, R. M., L. M. Orre, H. J. Johansson, V. Granholm, M. Huss, A. Perez-Bercoff, J. Forshed, L. Kall and J. Lehtio (2014). "HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics." Nat Methods **11**(1): 59-62.

- Bruder, S. P., N. S. Ricalton, R. E. Boynton, T. J. Connolly, N. Jaiswal, J. Zaia and F. P. Barry (1998). "Mesenchymal Stem Cell Surface Antigen SB-10 Corresponds to Activated Leukocyte Cell Adhesion Molecule and Is Involved in Osteogenic Differentiation." Journal of Bone and Mineral Research **13**(4): 655-663.
- Buchmaier, B. S., A. Bibi, G. A. Muller, G. H. Dihazi, M. Eltoweissy, J. Kruegel and H. Dihazi (2013). "Renal cells express different forms of vimentin: the independent expression alteration of these forms is important in cell resistance to osmotic stress and apoptosis." PLoS one **8**(7): e68301.
- Buljan, M., G. Chalancon, S. Eustermann, G. P. Wagner, M. Fuxreiter, A. Bateman and M. M. Babu (2012). "Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks." Molecular cell **46**: 871-883.
- Bunkenborg, J., G. E. Garcia, M. I. Paz, J. S. Andersen and H. Molina (2010). "The minotaur proteome: avoiding cross-species identifications deriving from bovine serum in cell culture models." Proteomics **10**(16): 3040-3044.
- Burrige, K. and M. Chrzanowska-Wodnicka (1996). "Focal adhesions, contractility, and signaling." Annual review of cell and developmental biology **12**: 463-518.
- Burrige, K., K. Fath, T. Kelly, G. Nuckolls and C. Turner (1988). "Focal adhesions: transmembrane junctions between the extracellular matrix and the cytoskeleton." Annual review of cell biology **4**: 487-525.
- Burrige, K. and K. Wennerberg (2004). "Rho and Rac take center stage." Cell **116**(2): 167-179.
- Cargile, B. J., D. L. Talley and J. L. Stephenson (2004). "Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides." Electrophoresis **25**(6): 936-945.
- Carr, S., R. Aebersold, M. Baldwin, A. Burlingame, K. Clauser and A. Nesvizhskii (2004). "The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data." Mol Cell Proteomics **3**(6): 531-533.
- Chen, Y. Y., S. Y. Lin, Y. Y. Yeh, H. H. Hsiao, C. Y. Wu, S. T. Chen and A. H. Wang (2005). "A modified protein precipitation procedure for efficient removal of albumin from serum." Electrophoresis **26**(11): 2117-2127.
- Chen, Y. Y., S. Y. Lin, Y. Y. Yeh, H. H. Hsiao, C. Y. Wu, S. T. Chen and A. H. J. Wang (2005). "A modified protein precipitation procedure for efficient removal of albumin from serum." Electrophoresis **26**(11): 2117-2127.
- Chervin-Pétirot, A., M. Courçon, S. Almagro, A. Nicolas, A. Grichine, D. Grunwald, M.-H. Prandini, P. Huber and D. Gulino-Debrac (2012). "Epithelial Protein Lost In Neoplasm (EPLIN) Interacts with α -Catenin and Actin Filaments in Endothelial Cells and Stabilizes Vascular Capillary Network in Vitro." Journal of Biological Chemistry **287**(10): 7556-7572.
- Colinge, J., J. Magnin, T. Dessingy, M. Giron and A. Masselot (2003). "Improved peptide charge state assignment." Proteomics **3**(8): 1434-1440.
- Colnot, C. (2005). "Cellular and molecular interactions regulating skeletogenesis." J Cell Biochem **95**(4): 688-697.
- Colzani, M., P. Waridel and J. Laurent (2009). "Metabolic labeling and protein linearization technology allow the study of proteins secreted by cultured cells in serum-containing media." Journal of proteome ...: 4779-4788.
- Cosen-Binker, L. I. and A. Kapus (2006). Cortactin: The Gray Eminence of the Cytoskeleton.

- Cox, J. and M. Mann (2008). "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification." *26*: 1367-1372.
- Cox, J. and M. Mann (2012). "1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data." *BMC Bioinformatics* **13 Suppl 16**: S12.
- Cox, J., N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen and M. Mann (2011). "Andromeda: a peptide search engine integrated into the MaxQuant environment." *J Proteome Res* **10**(4): 1794-1805.
- Cukierman, E., R. Pankov, D. R. Stevens and K. M. Yamada (2001). "Taking cell-matrix adhesions to the third dimension." *Science (New York, N.Y.)* **294**: 1708-1712.
- Curtis, A. S. G., J. V. Forrester, C. McInnes and F. Lawrie (1983). "ADHESION OF CELLS TO POLYSTYRENE SURFACES." *Journal of Cell Biology* **97**(5): 1500-1506.
- Daniels, K. and M. Solursh (1991). "Modulation of chondrogenesis by the cytoskeleton and extracellular matrix." *Journal of cell science* **100 (Pt 2)**: 249-254.
- de Crombrughe, B., V. Lefebvre, R. R. Behringer, W. Bi, S. Murakami and W. Huang (2000). "Transcriptional mechanisms of chondrocyte differentiation." *Matrix Biol* **19**(5): 389-394.
- de Godoy, L. M. F., J. V. Olsen, J. Cox, M. L. Nielsen, N. C. Hubner, F. Frohlich, T. C. Walther and M. Mann (2008). "Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast." *Nature* **455**(7217): 1251-1254.
- Dessau, W., H. von der Mark, K. von der Mark and S. Fischer (1980). "Changes in the patterns of collagens and fibronectin during limb-bud chondrogenesis." *J Embryol Exp Morphol* **57**: 51-60.
- Di Palma, S., M. L. Hennrich, A. J. Heck and S. Mohammed (2012). "Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis." *J Proteomics* **75**(13): 3791-3813.
- Dignam, J. D., R. M. Lebovitz and R. G. Roeder (1983). "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei." *Nucleic Acids Res* **11**(5): 1475-1489.
- Discher, D. E., P. Janmey and Y.-L. Wang (2005). "Tissue cells feel and respond to the stiffness of their substrate." *Science (New York, N.Y.)* **310**: 1139-1143.
- Discher, D. E., P. Janmey and Y. L. Wang (2005). "Tissue cells feel and respond to the stiffness of their substrate." *Science* **310**(5751): 1139-1143.
- Docheva, D., D. Padula, C. Popov, P. Weishaupt, M. Pragert, N. Miosge, R. Hickel, W. Bocker, H. Clausen-Schaumann and M. Schieker (2010). "Establishment of immortalized periodontal ligament progenitor cell line and its behavioural analysis on smooth and rough titanium surface." *Eur Cell Mater* **19**: 228-241.
- Du, J., X. Chen, X. Liang, G. Zhang, J. Xu, L. He, Q. Zhan, X.-Q. Feng, S. Chien and C. Yang (2011). "Integrin activation and internalization on soft ECM as a mechanism of induction of stem cell differentiation by ECM elasticity." *Proceedings of the National Academy of Sciences of the United States of America* **108**: 9466-9471.
- Ducret, A., I. Van Oostveen, J. K. Eng, J. R. Yates, 3rd and R. Aebersold (1998). "High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry." *Protein Sci* **7**(3): 706-719.
- Durr, J., P. Lammi, S. L. Goodman, T. Aigner and K. von der Mark (1996). "Identification and immunolocalization of laminin in cartilage." *Exp Cell Res* **222**(1): 225-233.

- Eng, J. K., B. C. Searle, K. R. Clauser and D. L. Tabb (2011). "A Face in the Crowd: Recognizing Peptides Through Database Search." Molecular & Cellular Proteomics **10**(11).
- Eriksson, H., J. Lengqvist, J. Hedlund, K. Uhlén, L. M. Orre, B. Bjellqvist, B. Persson, J. Lehtiö and P.-J. Jakobsson (2008). "Quantitative membrane proteomics applying narrow range peptide isoelectric focusing for studies of small cell lung cancer resistance mechanisms." Proteomics **8**: 3008-3018.
- Fell, H. B. (1925). "The histogenesis of cartilage and bone in the long bones of the embryonic fowl." Journal of Morphology **40**(3): 417-459.
- Fenn, J. B., M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse (1989). "ELECTROSPRAY IONIZATION FOR MASS-SPECTROMETRY OF LARGE BIOMOLECULES." Science **246**(4926): 64-71.
- Fey, S. J. and P. M. Larsen (2001). "2D or not 2D." Current Opinion in Chemical Biology **5**(1): 26-33.
- Folkman, J. and A. Moscona (1978). "Role of cell shape in growth control." Nature **273**(5661): 345-349.
- Frantz, C., K. M. Stewart and V. M. Weaver (2010). "The extracellular matrix at a glance." Journal of Cell Science **123**(24): 4195-4200.
- Frith, M. C., A. R. Forrest, E. Nourbakhsh, K. C. Pang, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki, T. L. Bailey and S. M. Grimmond (2006). "The abundance of short proteins in the mammalian proteome." PLoS Genet **2**(4): e52.
- Gerber, H. P., T. H. Vu, A. M. Ryan, J. Kowalski, Z. Werb and N. Ferrara (1999). "VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation." Nat Med **5**(6): 623-628.
- Gerter, R., J. Kruegel and N. Miosge (2012). "New insights into cartilage repair — The role of migratory progenitor cells in osteoarthritis." Matrix Biology **31**(3): 206-213.
- Gevaert, K., F. Impens, B. Ghesquière, P. Van Damme, A. Lambrechts and J. Vandekerckhove (2008). "Stable isotopic labeling in proteomics." Proteomics **8**(23-24): 4873-4885.
- Giancotti, F. G. (1999). "Integrin Signaling." Science **285**(5430): 1028-1033.
- Gilar, M., P. Olivova, A. E. Daly and J. C. Gebler (2005). "Orthogonality of separation in two-dimensional liquid chromatography." Analytical chemistry **77**: 6426-6434.
- Girardet, J. M., L. Miclo, S. Florent, D. Molle and J. L. Gaillard (2006). "Determination of the phosphorylation level and deamidation susceptibility of equine beta-casein." Proteomics **6**(12): 3707-3717.
- Goldring, M. B. (2012). "Chondrogenesis, chondrocyte differentiation, and articular cartilage metabolism in health and osteoarthritis." Therapeutic advances in musculoskeletal disease **4**: 269-285.
- Görg, A., W. Weiss and M. J. Dunn (2004). "Current two-dimensional electrophoresis technology for proteomics." Proteomics **4**(12): 3665-3685.
- Graumann, J., N. C. Hubner, J. B. Kim, K. Ko, M. Moser, C. Kumar, J. Cox, H. Scholer and M. Mann (2008). "Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins." Mol Cell Proteomics **7**(4): 672-683.
- Grinnell, F. and M. Feld (1982). "Fibronectin adsorption on hydrophilic and hydrophobic surfaces detected by antibody binding and analyzed during cell adhesion in serum-containing medium." Journal of Biological Chemistry **257**(9): 4888-4893.
- Guilak, F., L. G. Alexopoulos, M. L. Upton, I. Youn, J. B. Choi, L. I. Cao, L. A. Setton and M. A. Haider (2006). "The Pericellular Matrix as a Transducer of Biomechanical and

- Biochemical Signals in Articular Cartilage." Annals of the New York Academy of Sciences **1068**(1): 498-512.
- Guilak, F., D. M. Cohen, B. T. Estes, J. M. Gimble, W. Liedtke and C. S. Chen (2009). "Control of Stem Cell Fate by Physical Interactions with the Extracellular Matrix." Cell Stem Cell **5**(1): 17-26.
- Gygi, S. P., G. L. Corthals, Y. Zhang, Y. Rochon and R. Aebersold (2000). "Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology." Proc Natl Acad Sci U S A **97**(17): 9390-9395.
- Hall, B. K. and T. Miyake (2000). "All for one and one for all: condensations and the initiation of skeletal development." BioEssays **22**(2): 138-147.
- Halligan, B. (2009). ProMoST: A Tool for Calculating the pI and Molecular Mass of Phosphorylated and Modified Proteins on Two-Dimensional Gels. Phospho-Proteomics. M. Graauw, Humana Press. **527**: 283-298.
- Hathout, Y. (2007). "Approaches to the study of the cell secretome." Expert Rev Proteomics **4**(2): 239-248.
- Heck, A. J. R. (2008). "Native mass spectrometry: a bridge between interactomics and structural biology." Nature Methods **5**(11): 927-933.
- Henzel, W. J., T. M. Billeci, J. T. Stults, S. C. Wong, C. Grimley and C. Watanabe (1993). "Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases." Proc Natl Acad Sci U S A **90**(11): 5011-5015.
- Higuchi, A., Q.-D. Ling, Y. Chang, S.-T. Hsu and A. Umezawa (2013). "Physical cues of biomaterials guide stem cell differentiation fate." Chemical reviews **113**: 3297-3328.
- Hoving, S., B. Gerrits, H. Voshol, D. Muller, R. C. Roberts and J. van Oostrum (2002). "Preparative two-dimensional gel electrophoresis at alkaline pH using narrow range immobilized pH gradients." Proteomics **2**(2): 127-134.
- Hoving, S., H. Voshol and J. van Oostrum (2000). "Towards high performance two-dimensional gel electrophoresis using ultrazoom gels." Electrophoresis **21**(13): 2617-2621.
- Hsieh, S. and J. W. Jorgenson (1996). "Preparation and evaluation of slurry-packed liquid chromatography microcolumns with inner diameters from 12 to 33 microns." Anal Chem **68**(7): 1212-1217.
- Hu, Q., R. J. Noll, H. Li, A. Makarov, M. Hardman and R. Graham Cooks (2005). "The Orbitrap: a new mass spectrometer." J Mass Spectrom **40**(4): 430-443.
- Hu, Q. Z., R. J. Noll, H. Y. Li, A. Makarov, M. Hardman and R. G. Cooks (2005). "The Orbitrap: a new mass spectrometer." Journal of Mass Spectrometry **40**(4): 430-443.
- Huang, C. Y., P. M. Reuben, G. D'Ippolito, P. C. Schiller and H. S. Cheung (2004). "Chondrogenesis of human bone marrow-derived mesenchymal stem cells in agarose culture." Anat Rec A Discov Mol Cell Evol Biol **278**(1): 428-436.
- Huang da, W. (2007). "DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists." Nucleic Acids Res. **35**: W169-W175.
- Huang, S. and D. E. Ingber (1999). "The structural and mechanical complexity of cell-growth control." Nat Cell Biol **1**(5): E131-138.
- Hubner, N. C., S. Ren and M. Mann (2008). "Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis." Proteomics **8**: 4862-4872.
- Huey, D. J., J. C. Hu and K. A. Athanasiou (2012). "Unlike bone, cartilage regeneration remains elusive." Science **338**(6109): 917-921.

- Humphries, J. D., P. Wang, C. Streuli, B. Geiger, M. J. Humphries and C. Ballestrem (2007). "Vinculin controls focal adhesion formation by direct interactions with talin and actin." *J Cell Biol* **179**(5): 1043-1057.
- Hynes, R. (2002). "Integrins: bidirectional, allosteric signaling machines." *Cell* **110**: 673-687.
- Hynes, R. O. (2002). "Integrins: Bidirectional, allosteric signaling machines." *Cell* **110**(6): 673-687.
- Ingram, V. M. (1957). "Gene Mutations in Human Haemoglobin: the Chemical Difference Between Normal and Sickle Cell Haemoglobin." *Nature* **180**(4581): 326-328.
- Isacke, C. M. and H. Yarwood (2002). "The hyaluronan receptor, CD44." *The international journal of biochemistry & cell biology* **34**(7): 718-721.
- Janmey, P. A. (1998). "The cytoskeleton and cell signaling: component localization and mechanical coupling." *Physiol Rev* **78**(3): 763-781.
- Jiang, H. and A. M. English (2002). "Quantitative Analysis of the Yeast Proteome by Incorporation of Isotopically Labeled Leucine." *Journal of Proteome Research* **1**(4): 345-350.
- Jiménez, M. J. G., M. Balbín, J. M. López, J. Alvarez, T. Komori and C. López-Otín (1999). "Collagenase 3 Is a Target of Cbfa1, a Transcription Factor of the runt Gene Family Involved in Bone Formation." *Molecular and cellular biology* **19**(6): 4431-4442.
- Johnstone, B., T. M. Hering, A. I. Caplan, V. M. Goldberg and J. U. Yoo (1998). "In vitro chondrogenesis of bone marrow-derived mesenchymal progenitor cells." *Exp Cell Res* **238**(1): 265-272.
- Kawakami, Y., J. Rodriguez-León and J. C. I. Belmonte (2006). "The role of TGFβs and Sox9 during limb chondrogenesis." *Current Opinion in Cell Biology* **18**(6): 723-729.
- Keller, G. (2005). "Embryonic stem cell differentiation: emergence of a new era in biology and medicine." *Genes Dev* **19**(10): 1129-1155.
- Kersey, P. J., J. Duarte, A. Williams, Y. Karavidopoulou, E. Birney and R. Apweiler (2004). "The International Protein Index: an integrated database for proteomics experiments." *Proteomics* **4**(7): 1985-1988.
- Kim, I. S., F. Otto, B. Zabel and S. Mundlos (1999). "Regulation of chondrocyte differentiation by Cbfa1." *Mech Dev* **80**(2): 159-170.
- Kim, W., M. Kim and E. h. Jho (2013). "Wnt/β-catenin signalling: from plasma membrane to nucleus." *Biochemical Journal* **450**(1): 9-21.
- Köcher, T., R. Swart and K. Mechtler (2011). "Ultra-High-Pressure RPLC Hyphenated to an LTQ-Orbitrap Velos Reveals a Linear Relation between Peak Capacity and Number of Identified Peptides." *Analytical Chemistry* **83**(7): 2699-2704.
- Koelling, S., J. Kruegel, M. Irmer, J. R. Path, B. Sadowski, X. Miro and N. Miosge (2009). "Migratory chondrogenic progenitor cells from repair tissue during the later stages of human osteoarthritis." *Cell Stem Cell* **4**(4): 324-335.
- Konze, S. A., L. van Diepen, A. Schroder, R. Olmer, H. Moller, A. Pich, R. Weissmann, A. W. Kuss, R. Zweigerdt and F. F. Buettner (2014). "Cleavage of E-cadherin and beta-catenin by calpain affects Wnt signaling and spheroid formation in suspension cultures of human pluripotent stem cells." *Mol Cell Proteomics* **13**(4): 990-1007.
- Kosher, R. A., W. M. Kulyk and S. W. Gay (1986). "Collagen gene expression during limb cartilage differentiation." *J Cell Biol* **102**(4): 1151-1156.
- Kouzarides, T. (2007). "Chromatin modifications and their function." *Cell* **128**(4): 693-705.

- Krijgsveld, J., S. Gauci, W. Dormeyer and A. J. R. Heck (2006). "In-Gel Isoelectric Focusing of Peptides as a Tool for Improved Protein Identification." Journal of Proteome Research **5**(7): 1721-1730.
- Kruegel, J., B. Sadowski and N. Miosge (2008). "Nidogen-1 and nidogen-2 in healthy human cartilage and in late-stage osteoarthritis cartilage." Arthritis and rheumatism **58**: 1422-1432.
- Kuettner, K. E. (1992). "Biochemistry of articular cartilage in health and disease." Clin Biochem **25**(3): 155-163.
- Kulyk, W. M., B. J. Rodgers, K. Greer and R. A. Kosher (1989). "Promotion of embryonic chick limb cartilage differentiation by transforming growth factor-beta." Dev Biol **135**(2): 424-430.
- Kumar, C. and M. Mann (2009). "Bioinformatics analysis of mass spectrometry-based proteomics data sets." FEBS Letters **583**: 1703-1712.
- Kumar, D. and A. B. Lassar (2009). "The transcriptional activity of Sox9 in chondrocytes is regulated by RhoA signaling and actin polymerization." Mol Cell Biol **29**(15): 4262-4273.
- Kumar, H. R., X. Zhong, D. J. Hoelz, F. J. Rescorla, R. J. Hickey, L. H. Malkas and J. A. Sandoval (2008). "Three-dimensional neuroblastoma cell culture: proteomic analysis between monolayer and multicellular tumor spheroids." Pediatr Surg Int **24**(11): 1229-1234.
- Kvist, A. J., A. Nyström, K. Hultenby, T. Sasaki, J. F. Talts and A. Aspberg (2008). "The major basement membrane components localize to the chondrocyte pericellular matrix--a cartilage basement membrane equivalent?" Matrix biology : journal of the International Society for Matrix Biology **27**: 22-33.
- Kyte, J. and R. F. Doolittle (1982). "A SIMPLE METHOD FOR DISPLAYING THE HYDROPATHIC CHARACTER OF A PROTEIN." Journal of Molecular Biology **157**(1): 105-132.
- Laemmli, U. K. (1970). "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4." Nature **227**(5259): 680-685.
- Lambrecht, S., G. Verbruggen, P. C. Verdonk, D. Elewaut and D. Deforce (2008). "Differential proteome analysis of normal and osteoarthritic chondrocytes reveals distortion of vimentin network in osteoarthritis." Osteoarthritis Cartilage **16**(2): 163-173.
- Lee, H., E. C. Yi, B. Wen, T. P. Reily, L. Pohl, S. Nelson, R. Aebbersold and D. R. Goodlett (2004). "Optimization of reversed-phase microcapillary liquid chromatography for quantitative proteomics." J Chromatogr B Analyt Technol Biomed Life Sci **803**(1): 101-110.
- Lee, H. J., B. H. Choi, B. H. Min and S. R. Park (2009). "Changes in surface markers of human mesenchymal stem cells during the chondrogenic differentiation and dedifferentiation processes in vitro." Arthritis Rheum **60**(8): 2325-2332.
- Lee, K. Y. and D. J. Mooney (2012). "Alginate: properties and biomedical applications." Prog Polym Sci **37**(1): 106-126.
- Lefebvre, V., W. Huang, V. R. Harley, P. N. Goodfellow and B. de Crombrughe (1997). "SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene." Mol Cell Biol **17**(4): 2336-2346.
- Lefebvre, V. and P. Smits (2005). "Transcriptional control of chondrocyte fate and differentiation." Birth Defects Res C Embryo Today **75**(3): 200-212.

- Leung, T., X. Q. Chen, E. Manser and L. Lim (1996). "The p160 RhoA-binding kinase ROK alpha is a member of a kinase family and is involved in the reorganization of the cytoskeleton." Molecular and cellular biology **16**(10): 5313-5327.
- Li, F., Y. Zhang and C. Wu (1999). "Integrin-linked kinase is localized to cell-matrix focal adhesions but not cell-cell adhesion sites and the focal adhesion localization of integrin-linked kinase is regulated by the PINCH-binding ANK repeats." Journal of Cell Science **112**(24): 4589-4599.
- Li, Y. F. and P. Radivojac (2012). "Computational approaches to protein inference in shotgun proteomics." BMC Bioinformatics **13 Suppl 16**: S4.
- Liu, H., R. G. Sadygov and J. R. Yates (2004). "A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics." Analytical Chemistry **76**(14): 4193-4201.
- Loo, D. T., S. B. Kanner and A. Aruffo (1998). "Filamin binds to the cytoplasmic domain of the beta(1)-integrin - Identification of amino acids responsible for this interaction." Journal of Biological Chemistry **273**(36): 23304-23312.
- Lubec, G. and L. Afjehi-Sadat (2007). "Limitations and pitfalls in protein identification by mass spectrometry." Chem Rev **107**(8): 3568-3584.
- Ly, T., Y. Ahmad, A. Shlien, D. Soroka, A. Mills, M. J. Emanuele, M. R. Stratton, A. I. Lamond and J. Pines (2014). "A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells." eLife **3**.
- Lyons, T. J., R. W. Stoddart, S. F. McClure and J. McClure (2005). "The tidemark of the chondro-osseous junction of the normal human knee joint." Journal of molecular histology **36**: 207-215.
- Ma, H.-L., S.-C. Hung, S.-Y. Lin, Y.-L. Chen and W.-H. Lo (2003). "Chondrogenesis of human mesenchymal stem cells encapsulated in alginate beads." Journal of biomedical materials research. Part A **64**: 273-281.
- Ma, H. L., S. C. Hung, S. Y. Lin, Y. L. Chen and W. H. Lo (2003). "Chondrogenesis of human mesenchymal stem cells encapsulated in alginate beads." J Biomed Mater Res A **64**(2): 273-281.
- Mackie, E. J., Y. A. Ahmed, L. Tatarczuch, K. S. Chen and M. Mirams (2008). "Endochondral ossification: how cartilage is converted into bone in the developing skeleton." Int J Biochem Cell Biol **40**(1): 46-62.
- Mackie, E. J., I. Thesleff and R. Chiquet-Ehrismann (1987). "Tenascin is associated with chondrogenic and osteogenic differentiation in vivo and promotes chondrogenesis in vitro." The Journal of cell biology **105**(6): 2569-2579.
- Makarov, A. (2000). "Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis." Analytical Chemistry **72**(6): 1156-1162.
- Makarov, A., E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat and S. Horning (2006). "Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer." Analytical chemistry **78**: 2113-2120.
- Makarov, A., E. Denisov, O. Lange and S. Horning (2006). "Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer." J Am Soc Mass Spectrom **17**(7): 977-982.
- Mallick, P. and B. Kuster (2010). "Proteomics: a pragmatic perspective." Nature biotechnology **28**: 695-709.
- Mammoto, T., A. Mammoto and D. E. Ingber (2013). "Mechanobiology and developmental control." Annu Rev Cell Dev Biol **29**: 27-61.
- Mann, M. (2006). "Functional and quantitative proteomics using SILAC." Nat Rev Mol Cell Biol **7**(12): 952-958.

- Mann, M. and O. N. Jensen (2003). "Proteomic analysis of post-translational modifications." Nature Biotechnology **21**(3): 255-261.
- March, R. E. (1997). "An Introduction to Quadrupole Ion Trap Mass Spectrometry." Journal of mass spectrometry **32**(4): 351-369.
- Marcotte, E. M. (2007). "How do shotgun proteomics algorithms identify proteins?" Nat Biotechnol **25**(7): 755-757.
- Mardis, E. R. (2011). "A decade/'s perspective on DNA sequencing technology." Nature **470**(7333): 198-203.
- Maroudas, A., P. Bullough, S. A. Swanson and M. A. Freeman (1968). "The permeability of articular cartilage." J Bone Joint Surg Br **50**(1): 166-177.
- Mather, J. P. and P. E. Roberts (1998). Introduction to cell and tissue culture: theory and technique, Springer Science & Business Media.
- Maul, R. S. and D. D. Chang (1999). "EPLIN, Epithelial protein lost in neoplasm." Oncogene **18**(54): 7838-7841.
- Mayya, V., K. Rezaul, Y. S. Cong and D. Han (2005). "Systematic comparison of a two-dimensional ion trap and a three-dimensional ion trap mass spectrometer in proteomics." Mol Cell Proteomics **4**(2): 214-223.
- McBeath, R., D. M. Pirone, C. M. Nelson, K. Bhadriraju and C. S. Chen (2004). "Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment." Developmental cell **6**: 483-495.
- McLuckey, S. (1992). "Principles of collisional activation in analytical mass spectrometry." Journal of the American Society for Mass Spectrometry **3**(6): 599-614.
- Mehlhorn, A. T., H. Schmal, S. Kaiser, G. Lepski, G. Finkenzeller, G. B. Stark and N. P. Sudkamp (2006). "Mesenchymal stem cells maintain TGF-beta-mediated chondrogenic phenotype in alginate bead culture." Tissue Eng **12**(6): 1393-1403.
- Meissner, F., R. A. Scheltema, H. J. Mollenkopf and M. Mann (2013). "Direct proteomic quantification of the secretome of activated immune cells." Science **340**(6131): 475-478.
- Meng, F., B. J. Cargile, S. M. Patrie, J. R. Johnson, S. M. McLoughlin and N. L. Kelleher (2002). "Processing Complex Mixtures of Intact Proteins for Direct Analysis by Mass Spectrometry." Analytical Chemistry **74**(13): 2923-2929.
- Metzker, M. L. (2010). "Sequencing technologies [mdash] the next generation." Nat Rev Genet **11**(1): 31-46.
- Michel, P. E., F. Reymond, I. L. Arnaud, J. Josserand, H. H. Girault and J. S. Rossier (2003). "Protein fractionation in a multicompartiment device using Off-Gel (TM) isoelectric focusing." Electrophoresis **24**(1-2): 3-11.
- Miyazono, K. (2000). "Positive and negative regulation of TGF-beta signaling." J Cell Sci **113 (Pt 7)**: 1101-1109.
- Morgner, N., F. Montenegro, N. P. Barrera and C. V. Robinson (2012). "Mass spectrometry--from peripheral proteins to membrane motors." J Mol Biol **423**(1): 1-13.
- Muhammad, H., B. Schminke and N. Miosge (2013). "Current concepts in stem cell therapy for articular cartilage repair." Expert Opin Biol Ther **13**(4): 541-548.
- Muir, H. (1995). "The chondrocyte, architect of cartilage. Biomechanics, structure, function and molecular biology of cartilage matrix macromolecules." BioEssays **17**(12): 1039-1048.
- Nagaraj, N., J. R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Paabo and M. Mann (2011). "Deep proteome and transcriptome mapping of a human cancer cell line." Mol Syst Biol **7**.

- Nesvizhskii, A. I. and R. Aebersold (2004). "Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS." Drug discovery today **9**(4): 173-181.
- Nesvizhskii, A. I. and R. Aebersold (2005). "Interpretation of shotgun proteomic data - The protein inference problem." Molecular & Cellular Proteomics **4**(10): 1419-1440.
- Neuhoff, V., N. Arold, D. Traube, W. Ehrhardt and D. Taube (1988). "Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250." Electrophoresis: 255-262.
- O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins." Journal of Biological Chemistry **250**(10): 4007-4021.
- Oberlender, S. A. and R. S. Tuan (1994). "Expression and functional involvement of N-cadherin in embryonic limb chondrogenesis." Development **120**(1): 177-187.
- Oh, C. D., S. N. Maity, J. F. Lu, J. Zhang, S. Liang, F. Coustry, B. de Crombrughe and H. Yasuda (2010). "Identification of SOX9 interaction sites in the genome of chondrocytes." PLOS one **5**(4): e10113.
- Olsen, J. V., L. M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning and M. Mann (2005). "Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap." Mol Cell Proteomics **4**(12): 2010-2021.
- Olsen, J. V., S.-E. Ong and M. Mann (2004). "Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues." Molecular & Cellular Proteomics **3**(6): 608-614.
- Olsen, J. V., J. C. Schwartz, J. Griep-Raming, M. L. Nielsen, E. Damoc, E. Denisov, O. Lange, P. Remes, D. Taylor, M. Splendore, E. R. Wouters, M. Senko, A. Makarov, M. Mann and S. Horning (2009). "A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed." Molecular & Cellular Proteomics **8**(12): 2759-2769.
- Ong, S. E., B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey and M. Mann (2002). "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics." Mol Cell Proteomics **1**(5): 376-386.
- Onyekwelu, I., M. B. Goldring and C. Hidaka (2009). "Chondrogenesis, joint formation, and articular cartilage regeneration." J Cell Biochem **107**(3): 383-392.
- Pampaloni, F., E. G. Reynaud and E. H. Stelzer (2007). "The third dimension bridges the gap between cell culture and live tissue." Nat Rev Mol Cell Biol **8**(10): 839-845.
- Paul, W. and H. Steinwedel (1953). "A new mass spectrometer without a magnetic field." Zeitschrift fuer Naturforschung (West Germany) Divided into Z. Naturforsch., A, and Z. Naturforsch., B: Anorg. Chem., Org. Chem., Biochem., Biophys. **8**.
- Pawson, T. (1993). "Signal transduction—a conserved pathway from the membrane to the nucleus." Developmental Genetics **14**(5): 333-338.
- Pearle, A. D., R. F. Warren and S. A. Rodeo (2005). "Basic science of articular cartilage and osteoarthritis." Clin Sports Med **24**(1): 1-12.
- Pittenger, M. F., A. M. Mackay, S. C. Beck, R. K. Jaiswal, R. Douglas, J. D. Mosca, M. A. Moorman, D. W. Simonetti, S. Craig and D. R. Marshak (1999). "Multilineage potential of adult human mesenchymal stem cells." Science **284**(5411): 143-147.
- Pollard, T. D. and E. D. Korn (1973). "Acanthamoeba Myosin : II. INTERACTION WITH ACTIN AND WITH A NEW COFACTOR PROTEIN REQUIRED FOR ACTIN ACTIVATION OF Mg²⁺ ADENOSINE TRIPHOSPHATASE ACTIVITY." Journal of Biological Chemistry **248**(13): 4691-4697.
- Poole, C. A. (1997). "Articular cartilage chondrons: form, function and failure." J Anat **191 (Pt 1)**: 1-13.

- Poole, C. A. (1997). "Review. Articular cartilage chondrons: form, function and failure." Journal of anatomy **191**(1): 1-13.
- Poole, C. A., S. Ayad and J. R. Schofield (1988). "Chondrons from articular cartilage: I. Immunolocalization of type VI collagen in the pericellular capsule of isolated canine tibial chondrons." Journal of Cell Science **90**(4): 635-643.
- Pouillet, P., A. Gautreau, G. Kadare, J. A. Girault, D. Louvard and M. Arpin (2001). "Ezrin interacts with focal adhesion kinase and induces its activation independently of cell-matrix adhesion." Journal of Biological Chemistry **276**(40): 37686-37691.
- R Development Core Team (2010). R: A language and environment for statistical computing, R Foundation for Statistical Computing.
- Rabilloud, T. (2009). "Membrane proteins and proteomics: Love is possible, but so difficult." Electrophoresis **30**(S1): S174-S180.
- Rabilloud, T., A. R. Vaezzadeh, N. Potier, C. Lelong, E. Leize-Wagner and M. Chevallet (2009). "Power and limitations of electrophoretic separations in proteomics strategies." Mass Spectrom Rev **28**(5): 816-843.
- Raikos, V., R. Hansen, L. Campbell and S. R. Euston (2006). "Separation and identification of hen egg protein isoforms using SDS-PAGE and 2D gel electrophoresis with MALDI-TOF mass spectrometry." Food Chemistry **99**(4): 702-710.
- Ranganath, S. H., O. Levy, M. S. Inamdar and J. M. Karp (2012). "Harnessing the mesenchymal stem cell secretome for the treatment of cardiovascular disease." Cell stem cell **10**(3): 244-258.
- Rappsilber, J., Y. Ishihama and M. Mann (2003). "Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics." Anal. Chem. **75**: 663-670.
- Reed, P. W., A. Densmore and R. J. Bloch (2012). "Optimization of large gel 2D electrophoresis for proteomic studies of skeletal muscle." Electrophoresis **33**(8): 1263-1270.
- Roepstorff, P. and J. Fohlman (1984). "PROPOSAL FOR A COMMON NOMENCLATURE FOR SEQUENCE IONS IN MASS-SPECTRA OF PEPTIDES." Biomedical Mass Spectrometry **11**(11): 601-601.
- Rogowska-Wrzesinska, A., M.-C. Le Bihan, M. Thaysen-Andersen and P. Roepstorff (2013). "2D gels still have a niche in proteomics." Journal of proteomics **88**(0): 4-13.
- Ryan, J. A. (2008). "Evolution of Cell Culture Surfaces." BioFiles: 38:21.
- Schaefer, L. and R. M. Schaefer (2010). "Proteoglycans: from structural compounds to signaling molecules." Cell Tissue Res **339**(1): 237-246.
- Scheel, C., E. N. Eaton, S. H. Li, C. L. Chaffer, F. Reinhardt, K. J. Kah, G. Bell, W. Guo, J. Rubin, A. L. Richardson and R. A. Weinberg (2011). "Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast." Cell **145**(6): 926-940.
- Schirle, M., M. A. Heurtier and B. Kuster (2003). "Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry." Molecular & Cellular Proteomics **2**(12): 1297-1305.
- Schmidt, C., C. Lenz, M. Grote, R. Lührmann and H. Urlaub (2010). "Determination of Protein Stoichiometry within Protein Complexes Using Absolute Quantification and Multiple Reaction Monitoring." Analytical Chemistry **82**(7): 2784-2796.
- Schmidt, C. and H. Urlaub (2009). "iTRAQ-labeling of in-gel digested proteins for relative quantification." Methods Mol Biol **564**: 207-226.

- Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen and M. Selbach (2011). "Global quantification of mammalian gene expression control." Nature **473**(7347): 337-342.
- Schwartz, J. C., M. W. Senko and J. E. P. Syka (2002). "A two-dimensional quadrupole ion trap mass spectrometer." Journal of the American Society for Mass Spectrometry **13**(6): 659-669.
- Scigelova, M. and A. Makarov (2006). "Orbitrap Mass Analyzer – Overview and Applications in Proteomics." Proteomics **6**(S2): 16-21.
- Sekiya, I., K. Tsuji, P. Koopman, H. Watanabe, Y. Yamada, K. Shinomiya, A. Nifuji and M. Noda (2000). "SOX9 Enhances Aggrecan Gene Promoter/Enhancer Activity and Is Up-regulated by Retinoic Acid in a Cartilage-derived Cell Line, TC6." Journal of Biological Chemistry **275**(15): 10738-10744.
- Senko, M. W., J. D. Canterbury, S. Guan and A. G. Marshall (1996). "A High-performance Modular Data System for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry." Rapid Communications in Mass Spectrometry **10**(14): 1839-1844.
- Shen, Y., R. Zhao, S. J. Berger, G. A. Anderson, N. Rodriguez and R. D. Smith (2002). "High-Efficiency Nanoscale Liquid Chromatography Coupled On-Line with Mass Spectrometry Using Nanoelectrospray Ionization for Proteomics." Analytical Chemistry **74**(16): 4235-4249.
- Shen, Y. F. and R. D. Smith (2002). "Proteomics based on high-efficiency capillary separations." Electrophoresis **23**(18): 3106-3124.
- Shevchenko, A., H. Tomas, J. Havlis, J. V. Olsen and M. Mann (2006). "In-gel digestion for mass spectrometric characterization of proteins and proteomes." Nature Protocols **1**(6): 2856-2860.
- Shi, Y. and J. Massagué "Mechanisms of TGF- β Signaling from Cell Membrane to the Nucleus." Cell **113**(6): 685-700.
- Shin, J. W., J. Swift, I. Ivanovska, K. R. Spinier, A. Buxboim and D. E. Discher (2013). "Mechanobiology of bone marrow stem cells: From myosin-II forces to compliance of matrix and nucleus in cell forms and fates." Differentiation **86**(3): 77-86.
- Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Mol Syst Biol **7**: 539.
- Sillero, A. and J. M. Ribeiro (1989). "Isoelectric Points of Proteins: Theoretical Determination." Anal Biochem **179**(2): 319-325.
- Siuti, N. and N. L. Kelleher (2007). "Decoding protein modifications using top-down mass spectrometry." Nat Methods **4**(10): 817-821.
- Slebos, R. J. C., J. W. C. Brock, N. F. Winters, S. R. Stuart, M. A. Martinez, M. Li, M. C. Chambers, L. J. Zimmerman, A. J. Ham, D. L. Tabb and D. C. Liebler (2008). "Evaluation of Strong Cation Exchange versus Isoelectric Focusing of Peptides for Multidimensional Liquid Chromatography-Tandem Mass Spectrometry." Journal of Proteome Research **7**(12): 5286-5294.
- Sleno, L. and D. A. Volmer (2004). "Ion activation methods for tandem mass spectrometry." Journal of mass spectrometry **39**(10): 1091-1112.
- Smetana, K., Jr. (1993). "Cell biology of hydrogels." Biomaterials **14**(14): 1046-1050.
- Smits, P., P. Li, J. Mandel, Z. Zhang, J. M. Deng, R. R. Behringer, B. de Crombrughe and V. Lefebvre (2001). "The transcription factors L-Sox5 and Sox6 are essential for cartilage formation." Dev Cell **1**(2): 277-290.

- Solursh, M., T. F. Linsenmayer and K. L. Jensen (1982). "Chondrogenesis from single limb mesenchyme cells." Dev Biol **94**(1): 259-264.
- Stafford Jr, G. C., P. E. Kelley, J. E. P. Syka, W. E. Reynolds and J. F. J. Todd (1984). "Recent improvements in and analytical applications of advanced ion trap technology." International Journal of Mass Spectrometry and Ion Processes **60**(1): 85-98.
- Stamm, S., S. Ben-Ari, I. Rafalska, Y. Tang, Z. Zhang, D. Toiber, T. A. Thanaraj and H. Soreq (2005). "Function of alternative splicing." Gene **344**: 1-20.
- Stastna, M. and J. E. Van Eyk (2012). "Analysis of protein isoforms: Can we do it better?" Proteomics **12**(19-20): 2937-2948.
- Stastna, M. and J. E. Van Eyk (2012). "Investigating the secretome: lessons about the cells that comprise the heart." Circ Cardiovasc Genet **5**(1): o8-o18.
- Steen, H. and M. Mann (2004). "The ABC's (and XYZ's) of peptide sequencing." Nature Reviews Molecular Cell Biology **5**(9): 699-711.
- Steward, A. J., D. R. Wagner and D. J. Kelly (2013). "The pericellular environment regulates cytoskeletal development and the differentiation of mesenchymal stem cells and determines their response to hydrostatic pressure." Eur Cell Mater **25**: 167-178.
- Stossel, T. P., J. Condeelis, L. Cooley, J. H. Hartwig, A. Noegel, M. Schleicher and S. S. Shapiro (2001). "Filamins as integrators of cell mechanics and signalling." Nat Rev Mol Cell Biol **2**(2): 138-145.
- Strahler, J. R. and S. M. Hanash (1991). "Immobilized pH gradients: Analytical and preparative use." Methods **3**(2): 109-114.
- Swiderski, R. E. and M. Solursh (1992). "Localization of type II collagen, long form alpha 1(IX) collagen, and short form alpha 1(IX) collagen transcripts in the developing chick notochord and axial skeleton." Dev Dyn **194**(2): 118-127.
- Tamura, M., J. Gu, K. Matsumoto, S. Aota, R. Parsons and K. M. Yamada (1998). "Inhibition of cell migration, spreading, and focal adhesions by tumor suppressor PTEN." Science **280**(5369): 1614-1617.
- Taylor, G. (1964). "Disintegration of Water Drops in an Electric Field." Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **280**(1382): 383-397.
- Tesche, F. and N. Miosge (2005). "New aspects of the pathogenesis of osteoarthritis: the role of fibroblast-like chondrocytes in late stages of the disease." Histol Histopathol **20**(1): 329-337.
- Thakur, S. S., T. Geiger, B. Chatterjee, P. Bandilla, F. Frohlich, J. Cox and M. Mann (2011). "Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation." Molecular & Cellular Proteomics **10**(8).
- The UniProt Consortium (2014). "Activities at the Universal Protein Resource (UniProt)." Nucleic acids research **42**(D1): D191-D198.
- Thiede, B., C. J. Koehler, M. Strozynski, A. Treumann, R. Stein, U. Zimny-Arndt, M. Schmid and P. R. Jungblut (2013). "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 Chromatography Coupled to an LTQ-OrbitrapMass Spectrometer." Molecular & Cellular Proteomics **12**(2): 529-538.
- Tosh, D. and J. M. W. Slack (2002). "How cells change their phenotype." Nat Rev Mol Cell Biol **3**(3): 187-194.
- Tran, J. C., L. Zamdborg, D. R. Ahlf, J. E. Lee, A. D. Catherman, K. R. Durbin, J. D. Tipton, A. Vellaichamy, J. F. Kellie, M. X. Li, C. Wu, S. M. M. Sweet, B. P. Early, N. Siuti, R. D. LeDuc, P. D. Compton, P. M. Thomas and N. L. Kelleher (2011). "Mapping intact protein

- isoforms in discovery mode using top-down proteomics." *Nature* **480**(7376): 254-U141.
- Tudor, C., J. te Riet, C. Eich, R. Harkes, N. Smisdom, J. Bouhuijzen-Wenger, M. Ameloot, M. Holt, J. S. Kanger, C. G. Figdor, A. Cambi and V. Subramaniam (2014). "Syntenin-1 and ezrin link Activated Leukocyte Cell Adhesion Molecule to the actin cytoskeleton." *Journal of Biological Chemistry*.
- Twomey, J. D., P. I. Thakore, D. A. Hartman, E. G. Myers and A. H. Hsieh (2014). "Roles of type VI collagen and decorin in human mesenchymal stem cell biophysics during chondrogenic differentiation." *Eur Cell Mater* **27**: 237-250; discussion 249-250.
- van Hoof, D., J. Krijgsveld and C. Mummery (2012). "Proteomic analysis of stem cell differentiation and early development." *Cold Spring Harb Perspect Biol* **4**(3).
- Vandermarliere, E., M. Mueller and L. Martens (2013). "GETTING INTIMATE WITH TRYPsin, THE LEADING PROTEASE IN PROTEOMICS." *Mass Spectrometry Reviews* **32**(6): 453-465.
- Wasinger, V. C., S. J. Cordwell, A. Cerpa-Poljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams and I. Humphery-Smith (1995). "Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*." *Electrophoresis* **16**(7): 1090-1094.
- Watt, F. M., P. W. Jordan and C. H. O'Neill (1988). "Cell shape controls terminal differentiation of human epidermal keratinocytes." *Proc Natl Acad Sci U S A* **85**(15): 5576-5580.
- Westman-Brinkmalm, A. and G. Brinkmalm (2008). Tandem Mass Spectrometry. *Mass Spectrometry*, John Wiley & Sons, Inc.: 89-103.
- Wezeman, F. H. (1998). "Morphological foundations of precartilag development in mesenchyme." *Microscopy research and technique* **43**(2): 91-101.
- Wilm, M., G. Neubauer and M. Mann (1996). "Parent Ion Scans of Unseparated Peptide Mixtures." *Analytical Chemistry* **68**(3): 527-533.
- Wong, P. S. and R. Graham Cooks (1997). "Ion trap mass spectrometry." *Current separations* **16**: 85-92.
- Woods, A., G. Wang and F. Beier (2005). "RhoA/ROCK signaling regulates Sox9 expression and actin organization during chondrogenesis." *J Biol Chem* **280**(12): 11626-11634.
- Woods, A., G. Wang and F. Beier (2007). "Regulation of chondrocyte differentiation by the actin cytoskeleton and adhesive interactions." *Journal of cellular physiology* **213**(1): 1-8.
- Wozniak, M. A., K. Modzelewska, L. Kwong and P. J. Keely (2004). "Focal adhesion regulation of cell behavior." *Biochim Biophys Acta* **1692**(2-3): 103-119.
- Wu, L. and D. K. Han (2006). "Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics." *Expert Rev Proteomics* **3**(6): 611-619.
- Xu, J., W. Wang, M. Ludeman, K. Cheng, T. Hayami, J. C. Lotz and S. Kapila (2008). "Chondrogenic differentiation of human mesenchymal stem cells in three-dimensional alginate gels." *Tissue Eng Part A* **14**(5): 667-680.
- Xu, P., D. M. Duong and J. Peng (2009). "Systematical Optimization of Reverse-Phase Chromatography for Shotgun Proteomics." *Journal of Proteome Research* **8**(8): 3944-3950.
- Yamashita, M. and J. B. Fenn (1984). "ELECTROSPRAY ION-SOURCE - ANOTHER VARIATION ON THE FREE-JET THEME." *Journal of Physical Chemistry* **88**(20): 4451-4459.

- Yates, J. R. (2004). "MASS SPECTRAL ANALYSIS IN PROTEOMICS." Annual review of biophysics and biomolecular structure **33**(1): 297-316.
- Yates, J. R., 3rd (2013). "The revolution and evolution of shotgun proteomics for large-scale proteome analysis." J Am Chem Soc **135**(5): 1629-1640.
- Yates, J. R., 3rd, S. Speicher, P. R. Griffin and T. Hunkapiller (1993). "Peptide mass maps: a highly informative approach to protein identification." Anal Biochem **214**(2): 397-408.
- Zaidel-Bar, R., S. Itzkovitz, A. Ma'ayan, R. Iyengar and B. Geiger (2007). "Functional atlas of the integrin adhesome." Nature Cell Biology **9**(8): 858-868.
- Zhang, Z., W. Jin, J. Beckett, T. Otto and B. Moed (2011). "A proteomic approach for identification and localization of the pericellular components of chondrocytes." Histochemistry and cell biology **136**(2): 153-162.
- Zhu, H., M. Bilgin and M. Snyder (2003). "Proteomics." Annu Rev Biochem **72**: 783-812.
- Zhu, K., J. Zhao, D. M. Lubman, F. R. Miller and T. J. Barder (2005). "Protein pI Shifts due to Posttranslational Modifications in the Separation and Characterization of Proteins." Analytical Chemistry **77**(9): 2745-2755.
- Zougman, A., B. Pilch, A. Podtelejnikov, M. Kiehnopf, C. Schnabel, C. Kumar and M. Mann (2008). "Integrated analysis of the cerebrospinal fluid peptidome and proteome." J Proteome Res **7**(1): 386-399.

6 Appendixes

Appendix 1. Comparison of protein isoforms identified in any of the 1DE analyses of both biological replicates of CPC-2D / CPC-3D or using the 2DE approach.

1DE-r1 indicates 1DE analysis of the first biological replicate, 1DE-r2 indicates the 1DE analysis of the second biological replicate. Sequence identity was calculated from the sequence alignment of the isoforms using the Clustal Omega sequence alignment software, which does not consider gaps (no gap penalty) and its Uniprot implementation which considers gaps for the calculation of the identity score (gap penalty). Protein isoforms were ranked based on the number of razor and unique peptides mapping to each isoform (isoform with highest number of peptides having rank 1 and a decreasing rank corresponding to a decreasing number of peptides). When more than two isoforms for gene product were identified, sequence alignment was performed between each isoform and isoform with rank 1 only. Protein PEP is the product of the PEP of the razor and unique peptides that map to the protein.

Gene names	Protein ID	Sequence identity [%]		Razor and unique peptides			Protein PEP			Rank
		gap penalty	no gap penalty	1DE-r1	1DE-r2	2DE	1DE-r1	1DE-r2	2DE	
SOD2	F5GYZ5	64.49	95.81	0	1	2	NA	1.07E-03	2.56E-06	2
	P04179			17	21	24	1.58E-304	1.58E-304	1.58E-304	1
PLEC	Q15149			349	343	356	1.58E-304	1.58E-304	1.58E-304	1
	Q15149-3	96.39	98.86	1	1	0	3.84E-14	3.84E-14	NA	3
	Q15149-4	96.45	99.38	2	2	0	1.63E-34	1.63E-34	NA	2
	Q15149-8	96.32	99.78	0	1	0	NA	1.90E-03	NA	4
TNC	P24821			31	76	30	1.58E-304	1.58E-304	1.58E-304	1
	P24821-4	95.86	100	0	1	0	NA	6.68E-26	NA	2
	P24821-5	75.19	100	0	1	0	NA	4.91E-23	NA	3
POSTN	Q15063	96.53	99.88	0	1	0	NA	4.45E-02	NA	2
	B1ALD9			10	28	3	1.58E-304	1.58E-304	6.06E-74	1
	F5H628	87.62	94.53	0	1	0	NA	8.38E-06	NA	3
NDUFV3	P56181	13.29	58.88	3	12	10	1.58E-304	4.65E-04	1.58E-304	1
	P56181-2			0	1	0	NA	1.33E-03	NA	2
NDUFS7	F5GXJ1	76.01	92.12	0	0	1	NA	NA	7.34E-01	2
	O75251			5	5	3	1.20E-81	8.11E-66	1.77E-62	1
MACF1	H3BQK9			0	44	4	NA	1.58E-304	3.88E-05	1
	Q9UPN3	97.19	99.96	0	0	1	NA	NA	3.20E-01	2
	Q9UPN3-3	70.24	99.33	0	0	1	NA	NA	7.62E-02	3
LMNA	P02545			73	76	94	1.58E-304	1.58E-304	1.58E-304	1
	P02545-2	85.39	99.13	1	1	0	8.16E-54	8.16E-54	NA	2
	Q5TCI8	68.37	92.46	0	1	0	NA	6.61E-02	NA	3
IMMT	C9J406	88.22	100	0	1	0	NA	1.09E-03	NA	3

Appendixes

	Q16891	98.54	100	0	1	0	NA	6.17E-12	NA	2
	Q16891-2			34	46	55	1.58E-304	1.58E-304	1.58E-304	1
FN1	P02751			85	109	80	1.58E-304	1.58E-304	1.58E-304	1
	P02751-10	91.15	99.95	0	1	0	NA	4.01E-23	NA	3
	P02751-15	96.28	99.96	0	2	0	NA	2.76E-18	NA	2
DNAJB11	F5G XK8	4.16	20.55	0	1	0	NA	1.89E-04	NA	3
	H7C2Y5	40.87	94.34	0	1	0	NA	1.40E-02	NA	2
	Q9UBS4			0	9	17	NA	1.58E-304	1.58E-304	1
UQCRB	B7Z2R2	55.9	81.08	0	0	1	NA	NA	4.65E-02	2
	P14927			7	8	1	1.67E-240	9.28E-248	2.67E-02	1
TRPT1	A8MU17	34.36	81.65	0	1	0	NA	9.98E-05	NA	1
	H0YGY9			0	1	1	NA	1.23E-04	1.23E-04	2
TRAPPC4	Q9Y296	82.64	100	0	3	10	NA	1.91E-07	2.46E-74	1
	J3KP27			0	0	1	NA	NA	1.38E-01	2
TOR1AIP1	J3KN66	97.16	99.83	0	1	1	NA	1.43E-21	1.43E-21	2
	Q5JTV8			12	15	19	1.58E-304	1.58E-304	1.58E-304	1
TMPO	P42166	34.11	58.25	0	2	1	NA	1.62E-17	1.99E-01	2
	P42167			4	13	4	1.04E-10	2.35E-131	1.82E-17	1
THOC3	Q6NZ53	85.18	91.44	1	0	0	1.28E-01	NA	NA	2
	Q96J01			1	1	12	1.39E-11	1.39E-11	1.25E-45	1
SYNCRIP	O60506	88.12	97.69	15	30	18	1.58E-304	1.58E-304	1.58E-304	1
	O60506-3			0	1	0	NA	4.20E-02	NA	2
STXBP5	Q5T5C0	19.93	99.58	0	1	4	NA	2.66E-02	1.95E-09	1
	Q5JRH0			0	0	1	NA	NA	1.92E-03	2
SPTAN1	Q13813	99.15	99.96	14	137	42	1.58E-304	1.58E-304	1.58E-304	1
	Q13813-3			0	1	0	NA	2.25E-03	NA	2
SNAP23	H3BV99	67.29	100	0	0	1	NA	NA	2.71E-01	2
	O00161			2	10	7	2.78E-19	1.58E-190	4.11E-59	1
SMC1A	H0Y7K8	20.11	94.66	0	1	0	NA	8.84E-02	NA	2
	Q14683			1	33	2	3.82E-03	1.10E-179	6.70E-05	1
SDCBP	O00560	93.1	100	9	10	16	1.58E-304	1.58E-304	1.58E-304	1
	G5EA09			1	1	1	2.18E-163	2.18E-163	2.18E-163	2
RTN4	Q9NQC3-2	95.15	100	10	14	7	1.58E-304	1.58E-304	1.58E-304	1
	Q9NQC3-5			1	1	0	2.93E-07	2.93E-07	NA	2
RRBP1	Q9P2E9	99.78	100	27	63	31	1.58E-304	1.58E-304	1.58E-304	1
	Q9P2E9-2			0	1	0	NA	6.35E-04	NA	2
RPS3	E9PJN9	14.81	100	0	0	1	NA	NA	8.55E-15	2
	P23396			20	25	26	1.58E-304	1.58E-304	1.58E-304	1
RPL3	H7C422	64.41	99.64	0	0	1	NA	NA	6.67E-02	2

Appendixes

	P39023			6	14	1	8.64E-38	2.27E-172	1.50E-02	1
RCN1	B7Z1M1	76.13	90	0	1	0	NA	2.36E-02	NA	2
	Q15293			8	19	13	3.92E-196	1.58E-304	1.58E-304	1
RAB6A	P20340	98.55	98.56	0	1	0	NA	9.63E-50	NA	2
	P20340-2			11	12	9	1.58E-304	1.58E-304	1.41E-275	1
RAB1B	E9PLD0	84.08	100	0	1	0	NA	2.61E-02	NA	2
	Q9H0U4			7	8	4	2.13E-205	5.65E-208	2.27E-172	1
PYCRL	H0Y6C3	89.68	99.21	0	1	0	NA	9.02E-03	NA	2
	Q53H96			0	7	4	NA	1.05E-40	9.19E-22	1
PSMG4	C9J2F8	50.4	75.61	0	1	0	NA	2.93E-13	NA	1
	E2QRC7			0	1	0	NA	1.14E-03	NA	2
POR	H0Y4R2	79.73	95.76	0	0	1	NA	NA	5.95E-01	2
	P16435			9	29	3	1.86E-157	1.58E-304	1.35E-07	1
PLS3	B4DGB4	94.83	99.67	0	1	0	NA	4.63E-02	NA	2
	P13797			9	32	31	1.02E-204	1.58E-304	1.21E-266	1
PLIN3	O60664	57.83	100	3	19	19	1.60E-141	1.58E-304	1.58E-304	1
	O60664-2			0	0	1	NA	NA	3.12E-04	2
PKM2	P14618	95.85	95.86	37	57	48	1.58E-304	1.58E-304	1.58E-304	1
	P14618-2			1	4	3	1.04E-97	4.24E-105	9.15E-103	2
PITRM1	C9JSL2	20.73	92.27	22	45	34	3.27E-302	1.58E-304	1.58E-304	1
	Q5JRX3			0	1	0	NA	1.07E-01	NA	2
PICALM	Q13492	45.3	99.67	2	15	4	2.21E-149	1.58E-304	7.60E-160	1
	H0YEH1			0	1	0	NA	9.29E-49	NA	2
PFN2	P35080-2	93.57	93.57	4	5	9	4.43E-51	4.77E-107	8.86E-118	1
	P35080			0	0	1	NA	NA	5.03E-03	2
PEX26	E9PRC5	42.9	72.28	1	0	0	2.32E-01	NA	NA	2
	Q7Z412			0	3	0	NA	4.17E-28	NA	1
PDIA3	P30101	20	82.11	40	47	50	1.58E-304	1.58E-304	1.58E-304	1
	H7BZJ3			1	1	1	1.53E-235	1.53E-235	1.53E-235	2
P4HA1	P13674	98.12	98.13	1	2	1	1.36E-06	9.16E-17	1.36E-06	2
	P13674-2			13	34	35	1.58E-304	1.58E-304	1.58E-304	1
OTUB1	J3KR44	53.61	61.17	1	11	2	1.40E-05	5.90E-160	9.57E-08	1
	Q96FW1-2			0	0	1	NA	NA	3.82E-02	2
OAT	P04181	68.56	100	18	25	27	1.58E-304	1.58E-304	1.58E-304	1
	P04181-2			0	0	1	NA	NA	7.10E-04	2
NIT2	H7C579	44.86	96.51	0	1	0	NA	6.55E-02	NA	2
	Q9NQR4			3	13	13	1.49E-101	1.58E-304	1.58E-304	1
NDUFS4	H0Y9M8	64.57	97.41	0	0	1	NA	NA	2.52E-01	2
	O43181			2	5	1	3.88E-12	2.40E-21	3.36E-08	1

Appendixes

NDUFAB1	I3L505	44.87	100	0	0	1	NA	NA	6.94E-04	2
	O14561			3	4	3	1.83E-12	6.78E-21	1.03E-17	1
MYO1C	I3L501	15.9	98.22	0	1	0	NA	1.29E-02	NA	2
	O00159-3			16	57	15	1.58E-304	1.58E-304	1.58E-304	1
MYEF2	Q9P2K5	31.33	100	0	0	3	NA	NA	3.91E-07	1
	Q9P2K5-3			0	0	1	NA	NA	2.06E-01	2
MTHFD1L	H0Y327	33.6	99.4	0	0	1	NA	NA	2.21E-02	2
	Q6UB35			12	25	23	1.65E-123	1.85E-234	9.40E-215	1
MTCH2	E9PIE4	86.79	100	0	1	1	NA	6.42E-06	6.42E-06	2
	Q9Y6C9			11	13	15	4.35E-202	1.18E-225	2.62E-229	1
MRPL47	Q9HD33	56	100	1	9	0	1.19E-04	4.13E-39	NA	1
	Q9HD33-3			0	1	0	NA	7.20E-02	NA	2
MFGE8	Q08431	80.62	100	25	31	26	1.58E-304	1.58E-304	1.58E-304	2
	Q08431-2			0	0	1	NA	NA	1.59E-111	2
METTL9	Q9H1A3	99.68	100	0	1	0	NA	6.94E-05	NA	2
	Q9H1A3-2			0	4	3	NA	2.18E-15	2.23E-08	1
METAP2	F8VQZ7	99.79	100	0	1	0	NA	6.05E-03	NA	2
	P50579			0	12	9	NA	1.88E-162	1.00E-93	1
MAT2A	B4DN45	75.69	100	1	0	0	1.24E-01	NA	NA	2
	P31153			0	11	7	NA	4.13E-123	1.75E-67	1
LSS	C9J315	34.92	98.08	0	0	1	NA	NA	1.15E-02	2
	P48449			22	33	15	2.41E-297	1.58E-304	2.93E-216	1
LMO7	Q8WWI1	80.15	100	1	0	0	1.21E-01	NA	NA	2
	Q8WWI1-3			3	25	15	2.95E-59	1.58E-304	1.81E-176	1
KRAS	P01116	89.94	90.43	0	1	0	NA	3.49E-02	NA	2
	P01116-2			1	1	1	1.34E-09	1.34E-09	1.34E-09	1
JUP	F5GWP8	44.71	61.96	0	13	6	NA	1.18E-126	8.22E-33	1
	P14923			0	9	7	NA	3.03E-187	1.44E-143	2
ITGA2	E7ESP4	79.67	99.89	0	1	0	NA	2.83E-66	NA	2
	P17301			5	38	0	8.10E-46	1.58E-304	NA	1
ILF3	G5E9M5	76.5	97.86	22	37	18	5.80E-206	1.58E-304	1.04E-155	1
	Q12906			0	1	0	NA	1.18E-02	NA	2
IKBIP	Q70UQ0	35	42.81	10	23	20	2.02E-140	1.66E-258	1.78E-168	1
	Q70UQ0-4			7	15	5	1.66E-209	3.18E-283	9.62E-96	2
HNRNPR	O43390	99.52	100	12	19	13	1.58E-304	1.58E-304	4.49E-252	1
	O43390-2			0	1	0	NA	6.64E-09	NA	2
HMGA1	P17096	88.78	98.96	4	9	2	3.55E-40	1.76E-128	3.05E-35	1
	P17096-2			1	1	0	2.01E-06	2.01E-06	NA	2
GUK1	Q16774-2	90.36	100	1	8	11	1.52E-20	8.04E-100	3.82E-103	1

Appendixes

	Q16774			0	0	1	NA	NA	4.82E-04	2
GSN	P06396	58.82	94.85	27	40	35	1.58E-304	1.58E-304	1.58E-304	1
	Q5T0H9			0	0	1	NA	NA	7.13E-02	2
GLS	O94925	82.96	92.81	18	22	21	1.58E-304	1.58E-304	1.58E-304	1
	O94925-3			1	2	3	2.01E-78	1.68E-179	4.64E-181	2
GANAB	Q14697	97.72	100	1	1	1	3.84E-86	3.84E-86	3.84E-86	2
	Q14697-2			44	56	44	1.58E-304	1.58E-304	1.58E-304	1
GALNT7	H0YAH3	66.51	96.26	0	1	0	NA	4.48E-05	NA	2
	Q86SF2			0	10	0	NA	1.62E-135	NA	1
FBN2	D6REJ2	11.77	96.35	0	0	1	NA	NA	4.20E-01	2
	P35556			1	32	0	3.45E-06	1.97E-242	NA	1
FBN1	H0YND0	27.2	91.56	0	1	0	NA	2.64E-03	NA	2
	P35555			70	116	20	1.58E-304	1.58E-304	1.58E-304	1
FBLN1	B1AHL2	80.29	87.12	0	2	0	NA	2.50E-16	NA	2
	P23142			1	14	2	5.09E-03	1.23E-258	8.44E-09	1
FAM82B	B4DZW6	90.44	100	0	0	1	NA	NA	5.38E-02	2
	Q96DB5			6	16	17	2.84E-108	1.58E-304	3.27E-208	1
EHD1	C9JC03	65.69	100	0	0	1	NA	NA	4.49E-02	2
	Q9H4M9			3	22	19	6.16E-48	2.35E-204	3.75E-128	1
EEF1D	E9PRY8	36.87	100	5	8	10	6.93E-121	6.39E-238	3.28E-237	1
	P29692-3			0	1	0	NA	5.82E-05	NA	2
ECI2	C9J000	57.61	100	0	0	2	NA	NA	5.06E-03	2
	O75521			5	19	26	2.70E-105	1.58E-304	1.58E-304	1
CUX1	H0YL76	29.26	71.7	0	3	0	NA	9.54E-51	NA	2
	P39880-3			0	4	5	NA	1.96E-41	1.54E-50	1
COQ5	Q5HYK3	43.33	58.85	1	6	17	4.02E-20	6.59E-186	1.35E-183	1
	Q5HYK3-2			0	0	1	NA	NA	1.06E-01	2
COL6A3	E9PCV6	93.7	100	1	2	2	3.52E-15	3.80E-18	3.80E-18	2
	P12111			173	242	198	1.58E-304	1.58E-304	1.58E-304	1
COL6A2	P12110	81.72	91.47	43	56	56	1.58E-304	1.58E-304	1.58E-304	1
	P12110-2			2	2	5	1.10E-33	1.10E-33	1.16E-43	2
CCT3	B4DUR8	91.74	100	0	0	1	NA	NA	8.34E-02	2
	P49368			8	25	27	2.60E-138	3.93E-279	1.25E-217	1
CCDC90B	E9PKQ5	62.99	76.92	0	0	1	NA	NA	6.29E-02	2
	Q9GZT6			4	10	12	7.94E-35	1.23E-73	1.07E-58	1
CBFB	Q13951	89.3	91.76	0	3	5	NA	1.32E-70	2.86E-75	1
	Q13951-2			0	1	0	NA	2.39E-05	NA	2
CAP1	Q01518	99.78	100	9	19	26	1.34E-261	1.58E-304	1.58E-304	1
	Q01518-2			0	1	0	NA	5.68E-20	NA	2

Appendixes

CALD1	F5H1Z9	94.17	98.29	0	1	0	NA	7.60E-02	NA	2
	Q05682-5			7	19	16	3.69E-121	1.58E-304	1.58E-304	1
C8orf82	Q6P1X6	70.58	76.85	4	8	6	7.10E-86	8.05E-109	5.56E-91	1
	Q6P1X6-2			0	0	1	NA	NA	3.38E-01	2
C12orf10	F8VQQ3	80.05	100	1	0	0	2.36E-01	NA	NA	2
	Q9HB07			0	5	6	NA	7.99E-32	1.12E-43	1
ATPAF1	B7Z7I6	10.82	84.44	1	0	0	1.67E-01	NA	NA	2
	I3L448			0	5	10	NA	2.64E-232	4.50E-249	1
ATP5C1	P36542	99.66	100	14	19	16	1.72E-261	6.20E-278	1.10E-212	1
	P36542-2			1	1	1	1.28E-15	1.28E-15	1.28E-15	2
ARG1	P05089-2	71.51	100	0	5	0	NA	2.60E-24	NA	1
	P05089-3			1	0	0	4.14E-04	NA	NA	2
AP2A2	E9PQP4	9.58	100	0	0	1	NA	NA	2.38E-02	2
	O94973			0	27	9	NA	1.58E-304	2.86E-204	1
ANXA6	A6NN80	99.1	100	1	2	1	9.61E-08	1.57E-42	9.61E-08	2
	P08133			56	55	67	1.58E-304	1.58E-304	1.58E-304	1
ANXA4	B4DDF9	74.29	100	0	0	1	NA	NA	4.93E-05	2
	P09525			29	25	46	1.58E-304	1.58E-304	1.58E-304	1
ACTN1	P12814	97.59	100	29	42	12	1.58E-304	1.58E-304	1.97E-201	1
	P12814-3			0	1	0	NA	8.89E-02	NA	2

Acknowledgement

I would like to express my gratitude to the following people:

Prof. Henning Urlaub for giving me the opportunity to work in his laboratory. By being part of his group, I was able to learn a lot about mass spectrometry and proteomics for which I am very grateful. Furthermore, I acknowledge his support for sending me to conferences and summer schools. I not only met interesting people and gained knowledge on these occasions but also had a lot of fun.

The members of my supervising committee, Prof. Henning Urlaub, Prof. Michael Kessel and Prof. Nicolai Miosge, for their time and advice, also besides the regular reports.

Prof. Henning Urlaub, Prof. Michael Kessel, Prof. Ahmed Mansouri, Prof. Peter Rehling, Prof. Kai Tittmann and Prof. Jürgen Wienands for being members of my examination committee.

Prof. Nicolai Miosge and Christa Bode for their help on the CPC project.

Prof. Bernhard Küster and Dr. Amin Moghaddas Gholami from the TU München for their cooperation on the 2DE project.

Prof. Dieter Heineke for his help with the submission procedure.

My colleagues, Miro, Katha and Romi for the time that we spend together. I had a lot of fun with you guys especially at the summer school in Brixen.

Sunit and Kundan, my „Indian connection“, for your help with my thesis and for all the good moments together.

Saadia for her wise thoughts on life, they always brought something new to my perspective.

Samir for his sharp jokes and his numerous scientific questions.

Monika, Uwe and Johanna for their help in the lab.

All of you, the bioanalytical mass spec group, for the nice and interesting time I spend with you.

The „Bulgarian lunch group“ for bringing a bit of Bulgaria to the MPI.

Rebecca for supporting me unconditionally, you are the best.

My parents and my brother for supporting and encouraging me throughout my life.

Curriculum Vitae

Personal data

Name Ilian Ivanov Atanassov
Date of birth June 08, 1986
Place of birth Sofia / Bulgaria

Education

04/2010 – 05/2014 **PhD project**, Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Göttingen;

10/2008 - 04/2010 **M.Sc** Molecular Biology, within the Molecular Biology IMPRS program at the Georg-August University Göttingen

10/2005 – 06/2008 **B.Sc** in Biochemistry and Cell Biology, Jacobs University Bremen

09/2000 – 06/2005 National High School for Mathematics and Natural Sciences, Sofia / Bulgaria

Publications

Atanassov, I. and Urlaub, H. (2013). Increased proteome coverage by combining PAGE and peptide isoelectric focusing: Comparative study of gel-based separation approaches. *PROTEOMICS* 13(20): 2947–2955.

Atanassov, I., Atanassov, I., et al. (2009). A simple, flexible and efficient PCR-fusion/Gateway cloning procedure for gene fusion, site-directed mutagenesis, short sequence insertion and domain deletions and swaps. *Plant Methods* 5:11.

Declaration

This thesis has been written independently and with no other sources and aids than quoted.

Ilian Atanassov

Göttingen, May 27th, 2014