LUND UNIVERSITY

Global Proteome Survey -Transforming antibody-based affinity proteomics into a global discovery platform

Olsson, Niclas

2012

Department of Immunotechnology
Lund University

Global Proteome Survey

Transforming antibody-based affinity proteomics into a global discovery platform

Niclas Olsson

ACADEMIC THESIS which, by due permission of the Faculty of Engineering at Lund University,
will be publicly defended in Hörsal H 01, Health Sciences Centre,
Baravägen 3, Lund, Friday 4th of May 2012 at 09.15 a.m.

Faculty opponent is Dr. Thomas Joos, Natural and Medical Sciences Institute (NMI)
at the University of Tuebingen, Reutlingen, Germany.

LUND
UNIVERSITY

Organization
LUND UNIVERSITY
Department of Immunotechnology
BMC D13
SE-221 84 Lund
Sweden

Document name
DOCTORAL DISSERTATION

Date of issue
2012-05-04

Author(s)
Niclas Olsson

Sponsoring organization

Title and subtitle
Global Proteome Survey
Transforming antibody-based affinity proteomics into a global discovery platform

Abstract

Proteomics is expected to generate new insights into biological processes as well as identify novel biomarkers and therapeutic targets since most biological functions are transmitted through proteins. However, due to the complexity displayed by a proteome and inherent limitations associated with current methodologies, proteomic analyses often result in incomplete coverage and inconsistent measurements. Clearly, the development of novel high-performing proteomic platforms will be essential in order to successfully decipher the human proteome(s).

This thesis, based on four original papers denoted I to IV, describes the development and applicability of a novel proteomic technology platform entitled Global Proteome Survey (GPS) capable of transforming affinity proteomics into a global discovery engine. The GPS methodology combines the best features of affinity proteomics and mass spectrometry, and is based on using antibodies specific for short C-terminal amino acid peptide motifs shared by many proteins. This opens up the possibility to identify and quantify significant portions of a proteome, while still using a limited set of binders in a specie independent manner (Paper I-IV). Furthermore, structural models were generated for a large set of the experimentally verified captured peptides and the matching antibodies (Paper III). The data generated novel insights into antibody-peptide interactions and showed that a few key residues were essential for establishing specificity and acting as anchor residues.

The GPS assay reproducibility was extensively tested (Paper I, II, and IV) and displayed median coefficient of variations in the range of 10-20% in tissue profiling. In addition, the sensitivity was demonstrated by successfully targeting proteins present in a range of abundance values spanning over a million down to less than 50 copies per yeast cell (Paper I and II).

In Paper IV, the first clinical proteomic application of GPS was demonstrated by generating in-depth proteomic insights of 52 breast tumor tissues. While using only 9 antibodies, the GPS-platform enabled identification and quantification of over 1300 proteins, and most importantly established a link between a molecular signature and tumor progression. Highly relevant and promising cancer-associated protein signatures related to histologic grade, estrogen receptor, and HER2/*neu*-status were identified.

In conclusion, we have developed and demonstrated the quantitative capability, reproducibility, sensitivity, and coverage of the GPS methodology. It provides important methodological solutions to the complexity of proteome analysis and may act as a valuable tool for analyzing large numbers of clinical samples in an accurate, sensitive, and discovery-based manner. Hence, antibody-based affinity proteomics have been transformed into a global discovery platform and will pave the way for novel proteomic insights into complex molecular pathways in health and disease.

Key words
Proteomics, recombinant antibodies, affinity proteomics, mass spectrometry, proteome profiling, breast cancer, antibody specificity

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title

ISBN
978-91-7473-306-8

Recipient's notes

Number of pages
200

Price

Security classification

Distribution by (name and address)

Signature  Date 2012-03-26

Department of Immunotechnology
Lund University

Global Proteome Survey

Transforming antibody-based affinity proteomics into a global discovery platform

Niclas Olsson

LUND
UNIVERSITY

Lund 2012

Department of Immunotechnology
ISBN 978-91-7473-306-8
Lund University 2012

Contents

CONTENTS 3

MY CONTRIBUTION TO THE PAPERS IN THIS THESIS 5

ORIGINAL PAPERS 6

ABBREVIATIONS 7

1. INTRODUCTION 9

2. PROTEOMICS – MASS SPECTROMETRY 13

2.1 PROTEIN AND PEPTIDE BASED SEPARATION 14
2.1.1 Protein separation 14
2.1.2 Peptide based separation 15
2.2 MASS SPECTROMETRY AND ASSOCIATED MASS ANALYZERS 16
2.2.1 Time-of-flight (TOF) 16
2.2.2 Orbitrap 17
2.3 PEPTIDE IDENTIFICATION 18
2.4 CURRENT STATUS OF MASS SPECTROMETRY BASED PROTEOMICS 19

3. AFFINITY PROTEOMICS 21

3.1 AFFINITY PROBES 22
3.1.1 Polyclonal and monoclonal antibodies 22
3.1.2 Recombinant antibodies 23
3.2. ANTIBODY SPECIFICITY 24
3.3 CURRENT STATUS OF AFFINITY PROTEOMIC BASED ASSAYS 25

4. THE MARRIAGE OF AFFINITY PROTEOMICS AND MS 27

4.1 STABLE ISOTOPE STANDARD CAPTURE WITH ANTI-PEPTIDE ANTIBODIES 27
4.2 TERMINAL ANTI-PEPTIDE MOTIF SPECIFIC ASSAYS 28
4.2.1 Triple X Proteomics 29
4.2.2 Global proteome survey (GPS) 29
4.3 GPS CHARACTERISTICS 30

4.3.1 First generation of target motifs 30
4.3.2 CIMS-binder generation through phage display 30
4.3.3 CIMS-binder characteristics 32
4.3.4 GPS setup optimization 35
4.3.5 Label versus label-free 36
4.3.6 Assay reproducibility 37
4.3.7 Dynamic range and evaluation against SCX 38
4.3.8 First application of GPS - yeast 39
4.3.9 Next generation of motifs and binders 40

5. MOLECULAR PROFILING OF BREAST CANCER USING GPS 41

5.1 CURRENT STATUS OF BREAST CANCER ASSOCIATED MARKERS 42
5.1.1 BRCA1 and BRCA2 42
5.1.2 Transcriptomics derived classifiers 42
5.1.3 Protein based markers 43
5.1.4 Proteomics derived markers and current status 44
5.2 GPS AND BREAST CANCER 45
5.2.1 GPS identified cancer hallmarks 46
5.2.2 GPS as classifier of estrogen-status and HER2 48
5.2.3 Validation of GPS generated breast cancer signatures 48

6. CONCLUDING REMARKS 51

POPULÄRVETENSKAPLIG SAMMANFATTNING 54

ACKNOWLEDGEMENTS 57

REFERENCES 59

My contribution to the papers in this thesis

Paper I

I planned and performed all experiments, but the binder selection and protein array screening. I analyzed the data and participated in writing of the manuscript.

Paper II

I co-designed the study, planned and performed all experiments, analyzed the data and participated in the writing of the manuscript.

Paper III

I co-designed the study, planned and performed a majority of the experiments, and co-analyzed the data. I participated in the writing of the manuscript.

Paper IV

I planned, performed and analyzed all experiments together with PC. Further, I was co-main responsible for the writing of the manuscript.

Original papers

This thesis is based upon the following papers, which are referred to in the text by their Roman numerals (I-IV).

I. Niclas Olsson, Christer Wingren, Mikael Mattsson, Peter James, David O'Connell, Fredrik Nilsson, Dolores J. Cahill and Carl A.K. Borrebaeck. Proteomic analysis and discovery using affinity proteomics and mass spectrometry. **Mol Cell Proteomics** (2011) 10, M110.003962

II. Niclas Olsson, Peter James, Carl A.K. Borrebaeck and Christer Wingren. Quantitative proteomics targeting classes of motif-containing peptides using immunoaffinity-based mass spectrometry.
Submitted manuscript

III. Niclas Olsson, Stefan Wallin, Peter James, Carl A.K. Borrebaeck and Christer Wingren. Epitope-specificity of recombinant antibodies targeting tryptic proteomes reveals promiscuous peptide-binding properties.
Manuscript

IV. Niclas Olsson, Petter Carlsson, Peter James, Karin Hansson, Per Malmström, Mårten Fernö, Lisa Ryden, Christer Wingren and Carl A.K. Borrebaeck. Molecular portraits of histologic graded breast cancer tissues reflecting tumor progression.
Manuscript

Abbreviations

2DE	2-dimensional gel electrophoresis
CIMS	context independent motif specific
CDK1	cyclin dependent kinase 1
COFRADIC	combined fractional diagonal chromatography
CV	coefficient of variation
ER	estrogen receptor
ESI	electrospray ionization
FDR	false discovery rate
GPS	global proteome survey
HER2	human epidermal growth factor receptor 2
IEF	isoelectric focusing
IHC	immunohistochemistry
ITRAQ	isobaric tags for relative and absolute quantification
LC	liquid chromatography
LTQ	linear ion trap
mAb	monoclonal antibody
MALDI	matrix assisted laser-desorption ionization
MS	mass spectrometry
MS/MS	tandem mass spectrometry
pAb	polyclonal antibody
pI	isoelectric point
RP	reversed phase
scFv	single-chain fragment variable
SCX	strong cation exchange
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	stable isotope labeling with amino acids in cell culture
SISCAPA	stable isotope standards with capture by anti-peptide antibodies
SRM	selected reaction monitoring
TOF	time of flight
TXP	triple x proteomics
V_H	variable domain of the heavy chain
V_L	variable domain of the light chain

1. Introduction

Scientists are finally beginning to become equipped with highly sophisticated techniques and methodologies that may enable fundamental understanding of biological processes occurring in cells. Ever since the structure of the nucleic acid was reported back in 1953, crucial methodological and technological advancements have occurred and together evolved into the emerging "omics" technologies. The first eukaryotic genome i.e. *Saccharomyces cerevisiae* was published in 1996 (Goffeau et al., 1996) and more impressively just 5 years later, the first draft of the entire human genome was presented (Lander et al., 2001; Venter et al., 2001). This sequence generated information is now acting as blueprints daily used by many researchers. However, nucleotide sequence data alone will not be sufficient to elucidate how various biological processes are regulated and function in a cell. This has resulted in an effort to study expression levels of several active key components in a cell, i.e. mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics). Genome wide microarray expression analysis was developed in the mid-1990s (Schena et al., 1995) and has since then successfully generated numerous valuable insights into biological processes, classifications of tumors and disease states (Sorlie et al., 2001; Golub et al., 1999; van 't Veer et al., 2002; Ek et al., 2006). These mRNA based methodologies have advanced further and even enabled highly sensitive global transcriptome analysis of individual cells (Tang et al., 2009).

In parallel to transcript expression analysis, significant efforts have been made trying to decipher various proteomes. However, this task has turned out to be challenging, particularly for the human proteome(s). The term proteome was used for the first time in 1994 by Marc Wilkins to describe the entire complement of proteins expressed by a genome, cell, tissue or organism at a given time. Proteins are more diverse compared to DNA or mRNA, and can exist in various isoforms, splice variants, and have various post-translational modifications. While DNA is more or less static, each protein can in the same way as mRNA vary significantly in abundance. As a result, changes in the protein concentrations and their extent of post-translational modifications, such as, phosphorylation and glycosylation, greatly influence biological processes. Consequently, assessment of proteome changes

between healthy and diseased states is expected to be highly valuable for prognosis and diagnosis (Hanash, 2003; Borrebaeck and Wingren, 2009). In addition, identification of novel disease associated proteins may enable leads for new therapeutics, since proteins are very important therapeutic targets. However, in order to succeed and achieve complete proteome coverage, several highly sophisticated proteomic technologies and strategies will be required, due to the enormous complexity displayed by most proteomes.

So far, large proteome maps have successfully been generated for model organisms and human cell lines (de Godoy et al., 2008; Nagaraj et al., 2011b; Beck et al., 2011) thanks to the development of powerful mass spectrometry (MS)-based technology platforms. However, in order to achieve such extensive proteome coverage, laborious protein or peptide fractionation strategies together with extensive instrument analyzing times have frequently been required, generating key bottlenecks. Consequently these in-depth proteome discovery measurements have been limited to few proteomes, few samples, and few laboratories. Furthermore, these discovery efforts, to some extent, still suffer in terms of limited sensitivity and dynamic range. Alternative methodologies offering increased throughput and dynamic range have in parallel been developed by taking advantage of using antibodies unique binding properties. Affinity proteomics and, in particular, antibody microarrays (Wingren and Borrebaeck, 2004) have been established as a methodology offering highly sensitive proteomic analysis of some of the most challenging proteomes i.e. human serum (Wingren et al., 2007). However, the proteome coverage in antibody microarray-based approaches is limited by the availability of array optimized high-quality antibodies with desired specificities (Knezevic et al., 2001) and this is currently a bottleneck in order to enable discovery of biological responses on a global scale.

Nevertheless, cataloging of entire human proteome(s) might be within reach, giving mass spectrometry and antibody-based efforts another 5 to 10 years of development. But more importantly, how can we move forward and enable consistent, high-quality analyses of potentially all human proteins simultaneously in large cohorts of biologically clinically relevant samples? Large, statistically significant, sample sets (i.e. hundreds or thousands of samples per study) will be required in order for the proteomic generated biomarker discovery field to fully move forward. A biomarker is generally a molecule that is consistently modified or present at abnormal concentrations in specific illness or other health conditions.

With the aim of improving proteomic analysis I have in my thesis, based on four original papers denoted Paper I to IV, developed a proteomic workflow enabling fast, sensitive and reproducible protein profiling data, by combining affinity proteomics and MS. We termed the methodology global proteome survey (GPS) (Wingren et al., 2009) and it combines the use of recombinant antibodies and high resolution MS. Large groups of peptides were enriched by using antibodies specific for short terminal peptide motifs. The enriched peptides could then be quantified using MS and identified using tandem mass spectrometry (MS/MS), and subsequently be converted into proteomic maps revealing what proteins were present and their concentration. Within this thesis, I demonstrated that the developed GPS methodology fulfills several key criteria in order to become a powerful proteomic discovery approach.

In Paper I, a proof-of-concept study, the development of the GPS-methodology was described and we demonstrated its sensitivity and quantitative capability. In Paper II, we demonstrated the applicability of the GPS-methodology and focused on the quantitative capabilities and characteristics by profiling yeast proteomes. The sensitivity of the GPS-platform was demonstrated by successful quantification of proteins present in a range of abundance values spanning over a million down to less than 50 copies per cell. In addition, we demonstrated that quantitative data could be generated for the same set of peptides between biological replicates and that the GPS enabled complementary peptide coverage compared to classical methodologies. In Paper III, we characterized additional antibodies for the GPS platform and performed structural modeling of antibodies, detected peptides, and antibody-peptide interactions. Structural insights were revealed both for the captured peptides and used antibodies that could potentially explain the experimentally determined antibody specificities. In addition, new potential target motifs were evaluated *in silico* and the derived information will form the basis for the next round of selections and generation of GPS antibodies. Paper IV describes our first attempt to use the GPS-methodology for in-depth proteome analysis of histologic graded breast cancer tissues. By current proteomic standards, a relatively large cohort of samples (n=52) was profiled and over 1300 proteins were identified and quantified. The data showed that highly relevant differentially expressed cancer associated proteins could be identified and we established a link between a molecular signature and tumor cell progression. Several of the identified proteins may have a potential in future diagnostics and even as potential therapeutic targets.

Taken together, the methodological advancements and applications described in this thesis demonstrated that GPS has the potential to become a powerful technology in a

wide range of protein discovery profiling efforts. Furthermore, the GPS methodology also has a clear potential to be converted from a discovery platform into more targeted assay formats. This would then make GPS suitable for dedicated clinical test laboratories and thereby be a part of various future high throughput molecular signature screening, diagnostic as well as prognostic applications.

2. Proteomics – mass spectrometry

Proteomics is the large scale study of the proteome (the PROTEin complement expressed by a genOME (Wilkins et al., 1996)) and the proteins present in a given proteome will vary with time, stimuli, and the specific process a cell or organism undergoes. After some revisions in the original gene catalog from the Human-Genome Project, predictions were made that the human genome comprised approximately 20,500 protein coding genes (Clamp et al., 2007). At date, 20231 protein-encoding genes are listed (UniProtKB/SwissProt), although the precise number remains to be experimentally determined. Since proteins are closer to biological function than mRNA, it may be easier to prove hypotheses by monitoring proteins. Furthermore, due to alternative splicing and post-translational modifications, a single gene can be the parent of several protein products and this can only be addressed and resolved using proteomic methodologies (Pandey and Mann, 2000). Originally, it was believed that an increase or decrease in mRNA would result in the same change at the protein level. However, the correlation between mRNA and protein abundance levels remains an open question, since several studies have reported conflicting results (Gygi et al., 1999; Lu et al., 2007; Gry et al., 2009). In a recent study using yeast, cultivated under salt-stress conditions, Lee et al. demonstrated that transcript induction correlated with protein increase, while transcript reduction produced little to no change in some of the corresponding proteins in their model system (Lee et al., 2011). Hence, protein oriented studies are at least of equal or even higher importance compared to mRNA based studies in order to achieve increased understanding of gene functionality, biological processes and complex networks within a cell.

Within this chapter, I will briefly discuss some fundamental aspects within current MS-based proteomics, such as protein and peptide based separation, mass analyzers, and subsequent peptide identification and highlight current possibilities and limitations.

2.1 Protein and peptide based separation

Due to the inherent complexity displayed by a proteome, a wide array of methodologies has been developed to separate, identify, and quantify individual proteins, and study their associated characteristics. Some of the most commonly used protein and peptide based separation methods will be discussed below.

2.1.1 Protein separation

In a way, the birth of proteomics came already back in 1975 when the first large-scale protein technology, namely the two-dimensional (2DE) gel electrophoresis was developed (O'Farrell, 1975). Impressively, over 1100 different components from *Escherichia coli* were initially resolved by 2DE (O'Farrell, 1975). This methodology has, together with the implementation of the differential in-gel electrophoresis (DIGE) labeling system, enabled the separation and quantification of many thousands of proteins and been applied to a wide range of proteomes (Petrak et al., 2008; Bengtsson et al., 2007; Stella et al., 2011). 2DE-based separation of proteins is based on two orthogonal parameters, isoelectric point (pI) in the first dimension and molecular mass in the second dimension, by coupling isoelectric focusing (IEF) and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (O'Farrell, 1975). After separation, the proteins are visualized on the gels to evaluate the protein expression profiles that can be analyzed qualitatively and quantitatively. With parallel advancements of mass spectrometry for peptide identification, the 2D-gel electrophoresis combined with mass spectrometry quickly became the most widely used method of protein resolution and identification for several years (Hanash, 2000; Pandey and Mann, 2000). The methodology has over the years generated many potential biomarker candidates (Bengtsson et al., 2007; Petrak et al., 2008; Stella et al., 2011). One of the strengths with 2DE is its ability to resolve related proteins, such as differentially modified (i.e. phosphorylated) forms. However, the main inherent limitation of using 2DE-gels for protein profiling is that the method is limited to mainly high abundant proteins (Wilkins et al., 1998; Corthals et al., 2000). In addition, the methodology is relatively labor-intensive, has problems of co-migrating proteins, and limitations in detecting proteins of more hydrophobic character, like integral membrane proteins (Santoni et al., 2000) mainly due to low solubility in the aqueous-based buffers used.

2.1.2 Peptide based separation

In order to circumvent some of the shortcomings of protein based separation methodologies, parallel efforts were made on exploring strategies of peptide based separation using mainly non-gel based chromatography systems. In these so called shotgun proteomics experiment, the sample is first denatured to enable digestion and then digested using specific cleavage enzymes, i.e. trypsin, proven to cleave at unique sites (Olsen et al., 2004), in order to generate a mix of peptides of suitable length finally detected by mass spectrometry. The complexity of the sample is drastically increased since on average a single protein digested by trypsin generates 30-50 different peptides (Patterson and Aebersold, 2003) and thereby results in tens of thousands of different peptides, which often are too complex for a direct analysis. Consequently, the peptide mixture is frequently separated on a liquid chromatography (LC) column prior to identification by a tandem mass spectrometer (Aebersold and Mann, 2003). A major advancement toward such a comprehensive method came in 2001 when the multidimensional protein identification technology (MudPIT) was introduced (Washburn et al., 2001; Wolters et al., 2001). Over 1400 yeast proteins from all subcellular portions of the yeast were identified (Washburn et al., 2001). Noteworthy, even proteins with extremes in pI, MW, abundance, and hydrophobicity were all successfully identified.

Two of the most commonly used peptide separation methods during the last decade have been strong cation exchange (SCX) chromatography and reversed-phase (RP) chromatography (Washburn et al., 2001). Due to the direct compatibility between RP chromatography and mass spectrometry through electrospray ionization, the RP is frequently performed online using sophisticated LC-systems. We took advantage of using online RP-chromatography based peptide separation and sample cleanup when analyzing the majority of the eluted peptides from our GPS platform (Paper I-IV). Peptide separation systems using IEF, performed either in gel or off-gel, have also been developed (Hubner et al., 2008; Krijgsveld et al., 2006). In addition, the combined fractional diagonal chromatography (COFRADIC) system has elegantly demonstrated isolation of methionyl, cysteinyl and amino terminal peptides (Gevaert et al., 2002; Gevaert et al., 2005). The COFRADIC system has a very high potential to become very useful if combined with current powerful mass spectrometry instrumentations.

Taken together, the main advantage of working with any type of peptide based separation system is that peptides are more homogenous than proteins. The downside is that information, like protein mass and pI, which is retained in the protein based separation approaches, is completely lost in shotgun approaches.

2.2 Mass spectrometry and associated mass analyzers

Mass spectrometry is an analytical technique capable of forming, separating and detecting molecular ions based on their mass-to-charge ratio (m/z) and thereby the composition of a biomolecule can potentially be elucidated. In order to be successfully measured, the analyte of interest must first be ionized and transferred into the high vacuum system of the instrument, and this has to be done without destruction, which was initially proven to be difficult for proteins and peptides (Patterson and Aebersold, 2003). However, thanks to the development of soft ionization techniques called matrix-assisted laser desorption ionization (MALDI) (Karas and Hillenkamp, 1988; Tanaka et al., 1988) and electrospray ionization (ESI) (Fenn et al., 1989), proteins and peptides could be analyzed without completely breaking the molecules. Both Koichi Tanaka and John B Fenn were awarded the Noble prize in chemistry in 2002 as inventors of MALDI and ESI, respectively. Since then, a plethora of commercial MS-instruments have been developed with robust ESI or MALDI ionization sources and become widely available to the protein chemistry community (Patterson and Aebersold, 2003). With implementation of tandem mass spectrometry, in which further fragmentation is induced, structural information could be obtained from polypeptides. These advancements essentially replaced Edman degradation (Edman, 1949) for peptide sequencing, due to the higher offered sensitivity, sample throughput, and capacity to handle more complex samples. Currently, a wide variety of tandem mass spectrometers exist and their performance in terms of sensitivity, resolving power and mass accuracy depends on many parameters. Several of the instruments used for generating the data presented in Paper I-IV were so called “hybrid” tandem mass spectrometers combining different types of mass analyzers to enhance the capability or performance of the instrument.

2.2.1 Time-of-flight (TOF)

MALDI is commonly coupled with a time-of-flight (TOF) analyzer or tandem TOF/TOF analyzers (Medzihradszky et al., 2000) and offers high-throughput capabilities, robustness, and sensitivity. The principle is based on that larger ions take longer time to travel through the vacuum tube than the smaller ions (i.e. the time of flight) and in the case of MS/MS, the generated fragments are analyzed in the second TOF analyzer. In Paper I, we used MALDI-TOF for characterization of various synthetic peptide mixtures. Due to effects of ion suppression MALDI-MS has often been used to analyze simpler peptide mixtures, while integrated LC-ESI-

MS systems often are preferred for highly complex peptide mixtures. Consequently, for some of the more complex samples (i.e. colon) analyzed in Paper I, we in addition used an LC-ESI-QTOF setup. The QTOF is a hybrid mass analyzer with a quadrupole connected to a TOF analyzer offering high mass accuracy. However, we found that even the QTOF setup was outperformed by the newer, faster and more sensitive LTQ-Orbitrap setup when analyzing the GPS eluates.

2.2.2 Orbitrap

The proof of principle of the Orbitrap analyzer was first described by Makarov (Makarov, 2000) and the LTQ-Orbitrap was commercially introduced in 2005. This particular configuration has since then made significant contributions to the proteomics field (de Godoy et al., 2008; Nagaraj et al., 2011b) and become the clear choice for many proteomics applications (Scigelova and Makarov, 2006). The LTQ-Orbitrap is a hybrid instrument enabling two mass analyzers to work in parallel, e.g. a high resolution / mass accuracy spectrum of the precursor is acquired in the Orbitrap, while the linear ion trap (LTQ) carries out fragmentation and detection of MS/MS (or higher order MSn) spectra of selected peptides (Scigelova and Makarov, 2006). By using the Orbitrap as a detector for the precursor mass measurement, accuracies of less than 2 ppm can easily be achieved for analyses of complex peptide mixtures (Yates et al., 2006). Such a high precision of the fragmentation spectra of peptides might not be necessarily required for database searches and thereby fragmentation is recorded with the sensitive and faster LTQ-detector that delivers 3–5 MS/MS spectra per second. Since the introduction of the LTQ-Orbitrap, additional hybrid configurations have emerged with possibilities of additional fragmentation techniques i.e. electron-transfer dissociation (ETD) and higher energy collision dissociation (HCD). In Paper I-IV, we used a LC-ESI-LTQ-Orbitrap setup for analyzes of yeast, colon, and breast tumor samples. When compared to the QTOF-setup the LTQ-Orbitrap-setup delivered significantly more identified peptides from several of the GPS eluates. The high precursor mass accuracy from the Orbitrap clearly improved our results in terms of significantly identified peptides. However, an equally important reason for the improved performance was the gained sensitivity and scan speed offered by the LTQ. The possibility to perform more MS/MS per second enabled far more reproducible identifications compared to the QTOF setup when analyzing replicate captures.

2.3 Peptide identification

During 1993, five different groups independently presented the idea of using mass spectral data combined with database searches of theoretical masses generated by *in silico* cleavage of proteins in order to identify and map peptides to a protein (James et al., 1993; Henzel et al., 1993; Mann et al., 1993; Pappin et al., 1993; Yates et al., 1993). This methodology is now referred to as peptide mass fingerprinting (PMF). However, mapping only intact peptide masses might not be enough. To this end, sophisticated algorithms were in parallel developed that matched fragmented peptide (MS/MS) spectra to its corresponding protein in sequence databases (Mann and Wilm, 1994; Eng et al., 1994; Perkins et al., 1999). However, certain caution needs to be taken in order to ensure data quality. First of all, peptides can often be mapped to several proteins, in particular if only the parent mass is determined and no MS/MS is performed. Furthermore, a single identified peptide sequence per protein, compared to two or more peptides, has, at least in the past, by many been considered as lower quality and publication strategies have even been implemented demanding "two-peptide" rule unless individual annotated spectra are provided (Carr et al., 2004; Wilkins et al., 2006). However, critics have recently been raised and suggestions made that the "two-peptide" rule should be abandoned (Gupta and Pevzner, 2009). Several approaches have been adopted in order to improve data quality for peptide identifications and parallel searches are currently often performed in shuffled or reversed databases to estimate False Discovery Rate (FDR) (Elias and Gygi, 2007). In addition, using multiple search engines may increase identifications and a combined FDR cut off can then be implemented (Hakkinen et al., 2009).

The importance of data analysis became apparent in a recent large inter-laboratory study (Bell et al., 2009), where a standardized mixture, made up of 20 highly purified recombinant proteins, where each protein contained at least one tryptic peptide of 1250 Da, was distributed among 27 laboratories for analysis. Initially, only 7 labs reported all 20 proteins correctly, and only 1 lab reported all expected tryptic peptides of 1250 Da. However, centralized data analysis later revealed that all 20 proteins and most of the 1250 Da peptides had, in fact, been detected in all 27 labs. Clearly, data handling and analysis is far from optimized and alternative search algorithms are frequently released (Cox, 2011). In order to handle generated data (Paper I, III and IV), we used the Proteios Software Environment (Hakkinen et al., 2009) (http://www.proteios.org) and this is a very powerful platform for analysis and management of proteomics data. Decoy searches and FDR cutoffs were implemented for all peptide identification data generated in this thesis (Paper I-IV).

In addition, combined database searches using both Mascot (Perkins et al., 1999) and X!Tandem (Craig and Beavis, 2004) were used in paper I, III and partially for paper IV. The workflow of using combined database searches enabled more peptide identifications and this was particularly seen in Paper IV.

2.4 Current status of mass spectrometry based proteomics

Shotgun-based proteomic approaches combined with high-resolution MS has improved at a rapid pace during the last decade and represents a very powerful hypothesis generating tool for in-depth characterization of biological systems (Aebersold and Mann, 2003; Mallick and Kuster, 2010). The shotgun approach has typically been used for comprehensive proteomic discovery surveys, limited to few samples and comparison of two different biological states. In terms of proteome coverage, the increase has closely followed technological improvements (hence, similar trends as the Moore's law, roughly meaning a doubling every 2 years). Assuming improvements of the instrumentation continue to follow such a trend, given the methodology another 5 to 10 years, might potentially enable close to complete coverage of certain proteomes. Recently, two independent studies succeeded in identifying over 10,000 human proteins in two human cell line proteomes (Nagaraj et al., 2011b; Beck et al., 2011). However, these single proteome studies required significant amount of work (i.e. sample fractionation) and furthermore in the case of the Nagaraj et al., a massive 288 hours of instrument measuring time. Hence, 12 days were needed to profile a single sample. Such workloads and measuring times are not feasible when large and more complex sample cohorts (e.g. hundreds of patient tissues or plasma samples) compared to a cell line have to be analyzed.

Analyzes of tissues and in particular, plasma samples, have turned out to be much more challenging and normally result in very incomplete proteome coverage and limit the analysis to the most abundant proteins (Whiteaker et al., 2011a). The dynamic range of protein concentrations in plasma exceeds at least 10 orders of magnitude and this is substantially more challenging than the five orders of magnitude in yeast (Ghaemmaghami et al., 2003) or the seven orders of magnitude measured in the above discussed cell line proteome by Beck et al. Hence, the methodology is not yet capable of characterizing challenging sample formats, such as entire plasma proteomes and complex tissues, where it continues to be highly dependent on laborious setups using fractionation/enrichment steps (Carr and Anderson, 2008).

The usage of combinatorial peptide ligand libraries (hexa-peptide libraries) have recently been demonstrated to increase proteome coverage (Righetti et al., 2006) and to some extent overcome some of the complexity in challenging proteomes, like human erythrocytes (Roux-Dalvai et al., 2008). However, it remains to be seen whether this approach has full quantitative capabilities.

Selected reaction monitoring mass spectrometry (SRM-MS), also called multiple reaction monitoring (MRM-MS), performed on triple quadrupole mass spectrometers has emerged as a highly sensitive, robust, and quantitatively precise methodology when targeting prototypic peptides (Lange et al., 2008b; Lange et al., 2008a). These types of targeted analyses are capable of delivering measurements of a defined set of proteins in every sample and thereby circumvent the dilemma of incomplete measurements experienced within classical shotgun proteomics. Impressive system biological insights have been generated with the SRM approach (Picotti et al., 2009; Malmstrom et al., 2009; Costenoble et al., 2011). In addition, SRM assays have been developed for measuring some of the major plasma proteins (Kuzyk et al., 2009) and furthermore demonstrated high reproducibility between different laboratories (Addona et al., 2009). However, even the SRM-methodology is not able to detect the majority of the lowest abundant analytes present in challenging samples e.g. plasma samples without any sample fractionation or enrichment steps (Whiteaker et al., 2011c).

Clearly, when it comes to analyses of highly complex protein samples using mass spectrometry, certain areas are still in need of further improvements. The limitations I particularly would like to see being addressed are summarized below:

- further improved throughput on current LC-ESI-MS/MS based systems is needed, e.g. number of samples or proteomes possible to analyze within a certain time
- move towards minimalistic sample preparation workflows, e.g. few or no fractionation steps
- enable successful measurements of large sets of low abundant analytes in highly complex samples

Mass spectrometry based proteomics is and will continue to be an extremely powerful methodology. However in order to achieve its full power and capabilities when analyzing highly complex proteome samples, above issues have to be efficiently solved.

3. Affinity proteomics

Antibodies, also known as immunoglobulins, have a central role in the immune system where they specifically recognize and bind a variety of foreign substances. Owing to the unique binding properties of antibodies, they have been utilized in a vast array of research applications, clinical diagnostics and therapeutics (Borrebaeck, 2000; Carter, 2001). The primary applications where antibodies are used as affinity reagents for detection of proteins of interest in complex biological samples are assays, such as enzyme-linked immunosorbent assay (ELISA), protein microarrays, immunohistochemistry (IHC), western blot and flow cytometry. The accepted 'gold standard' for sensitive single-protein measurements is clearly the immunoassay with attributes such as flexibility, ease of use, low-cost and throughput. These attributes have been difficult to match for other methodologies such as MS-based protein assays. Consequently, there has been a massive interest in developing highly multiplexed immunoassay measurements during the last decade (Kingsmore, 2006; Borrebaeck and Wingren, 2011). These efforts have developed into a strongly expanding research field termed affinity proteomics. In particular antibody-based microarrays (Haab, 2005; Wingren and Borrebaeck, 2006; Hartmann et al., 2009) have emerged as a fundamental affinity-based platform for multiplexed protein profiling. The underlying principle of miniaturized, solid-phase, immunoassays in an array format were developed in the mid-1980s (hence even before the DNA microarrays) (Ekins and Chu, 1991) and have since then become a well-established proteomic research tool, providing sensitive, rapid and multiplexed analysis of protein abundance (Haab et al., 2001; Wingren and Borrebaeck, 2004; Borrebaeck and Wingren, 2007; Yu et al., 2010). Antibody microarrays are produced by depositing small drops (pL scale) of individual antibodies onto a solid support in a defined pattern, an array. The immobilized antibodies can then capture its specific target when exposed to a biological sample. Following antigen binding and stringent washing steps, the detection (and quantification) of the captured protein is then performed. By having labeled the biological sample chemically with e.g. biotin, prior to the incubation step on the array, the final detection step can be performed by adding fluorescently labeled streptavidin, which after binding to biotin can be measured with a confocal scanner.

Within this chapter, I will briefly discuss some fundamental aspects within affinity based proteomic assays, such as probe format, specificity and assay-setups.

3.1 Affinity probes

Affinity-based proteomics approaches to proteome analysis need systematic generation of binders and in principle, against all genome-encoded proteins and their variant forms in order to enable analysis of global protein expression patterns. However, despite multiple efforts in the academic and industrial sectors, the availability of well-characterized, assay optimized, validated specific antibodies is a limiting factor for affinity proteomic strategies often resulting in limited proteome coverage (Colwill et al., 2011; Dubel et al., 2010; Ramachandran et al., 2008). Significant large scale efforts, such as the Human Protein Atlas (www.proteinatlas.org) (Uhlen et al., 2005) and the HUPO human antibody initiative (www.hupo.org/research/hai/) have therefore been initiated toward systematic generation and validation of a resource of antibodies to the entire human proteome. Furthermore, initiatives such as the Affinomics (www.affinomics.org/) focus on generating comprehensive binder sets for certain dedicated protein groups and make a strong effort in implementing recombinant binder formats. In parallel to these highly coordinated efforts many more focused efforts within single academic labs have generated plenty of recombinant binder reagents during the last decade (Soderlind et al., 2000; Persson et al., 2009; Gustavsson et al., 2011).

3.1.1 Polyclonal and monoclonal antibodies

From a structural point of view, an antibody consists of two identical heavy- and light chains (approximately 50 kDa and 25 kDa) respectively, resulting in a Y-shaped molecule with two identical antigen binding sites. The light chain consists of one variable (V_L) and one constant (CL) domain, while the heavy chain, contains one variable domain (V_H) and three or four constant domains (CH1, CH2, CH3, CH4) depending on isotype. In general, the two most commonly used types of antibodies as probes are (i) polyclonal antibodies (pAb) made by immunization of animals and (ii) monoclonal antibodies (mAb) made by immunization and screening of antibody producing mouse or rabbit cells (Kohler and Milstein, 1975; Milstein, 2000). The aim of the Swedish Human Protein Atlas project is to generate affinity purified polyclonal rabbit antibodies (entitled mono-specific antibodies) to all human non-redundant proteins (Uhlen and Ponten, 2005; Nilsson et al., 2005).

These binders have so far been generated by using protein epitope signature tags (PrESTs), comprising 100-150 consecutive amino acid residues of a protein, as both immunogens and affinity ligands (Agaton et al., 2003). The current release version (9.0) includes more than 15,500 antibodies, targeting proteins from more than 12,200 human genes (~61% of the human protein-coding genes). Binder reagents together with associated information regarding individual protein localization and expression pattern continuously become available to the proteomic research community (Ponten et al., 2008; Ponten et al., 2011). Such information and binder resources are going to be highly valuable. However, since these antibodies are still polyclonal and thereby non-renewable, this might be a critical issue, since each new produced batch, generated by additional immunization, might result in batch-to-batch variations (Stoevesandt and Taussig, 2007; Saerens et al., 2008). A monoclonal antibody expressed by a hybridoma cell-line basically ensures an essentially unlimited supply of a specific antibody and thereby circumvent the batch to batch issue by pAbs. However, generating the large number of monoclonal (mAb) or polyclonal (pAb) binders required for global proteome analysis is a costly process and will continue to be labor intensive and logistically challenging (Phelan and Nock, 2003). In addition, certain antigens might generate insufficient immune response. Furthermore, it has been observed that only a fraction of commercially available pAbs and mAbs may be possible to use in conventional antibody array setups since large proteins like pAbs and mAbs unfortunately tend to denature when deposited in small drops and allowed to dehydrate on hydrophobic surfaces, (Haab et al., 2001; MacBeath, 2002). Finally, there are also the ethical aspects regarding the massive use of animals dedicated to antibody generation to consider.

3.1.2 Recombinant antibodies

In order to overcome several of the above discussed issues associated with pAbs and mAbs, a third major type of antibodies, i.e. recombinant antibodies, made through *in vitro* selection approaches has been developed and increased in popularity. After over 35 years of optimization, the hybridoma technology has approached its limits, while the *in vitro* display systems currently provide tremendous potential for miniaturization and acceleration of binder generation (Dubel et al., 2010). Recombinant binders have inherent advantages over their animal-derived counterparts and the use of *in vitro* technologies offers several properties not available in regular animal-based antibody generation methods. First and foremost, it has been shown that specific high-affinity renewable antibodies can be produced quickly and efficiently and the technology allows for generation of antibodies

without sacrificing animals (Soderlind et al., 2000; Colwill et al., 2011). Hence, recombinant antibodies can already be generated at a lower cost to monoclonal hybridoma antibodies (Dubel et al., 2010) and the approach permits control over all selection and screening conditions. Secondly, recombinant antibodies are a fully renewable probe source, thereby eliminating the risk of potential batch-to-batch variation which is crucial if binders are used in diagnostic and therapeutic applications. In contrast to both pAbs and mAbs, subsequent genetic modifications to improve properties (e.g. affinity and stability) or assay compatibility (e.g. fusion to alternative tags) can also easily be made (Borrebaeck and Wingren, 2011). In particular the ability to improve affinity and specificity has made the *in vitro* antibody technologies powerful (Bradbury et al., 2011) and this was clearly exemplified by successful isolation of an anti-fluorescein scFv with an impressive affinity in the femtomolar range (Boder et al., 2000). One of the most common recombinant binder formats is the single-chain fragment variable (scFv) that is composed of the antigen binding domains, V_L and V_H, fused together by a 15-25 amino acid long linker (Bird et al., 1988; Soderlind et al., 2000). Recombinant scFv antibodies, based upon a single fixed framework, have proven to be an excellent, stable and reproducible probe resource for antibody microarray set-ups (Steinhauer et al., 2002; Wingren and Borrebaeck, 2006). In addition to scFvs, several alternative recombinant binding reagents e.g. affibodies (Nord et al., 1997) and aptamers (Ellington and Szostak, 1990) have been developed and used in array formats (Renberg et al., 2007; Gold et al., 2010), although not as commonly used as scFvs or regular antibodies.

3.2. Antibody specificity

Antibodies are clearly very powerful capture agents and the issue of specificity is central when working with antibodies as probes. The functionality of each individual antibody can vary depending on assay format and sample targeted. Cross reactivity has been observed (Michaud et al., 2003; Kijanka et al., 2009) and recently a high proportion of commercial antibodies demonstrated poor specificity in western blots (Hamdani and van der Velden, 2009; Jensen et al., 2009). Proteins might interact with other proteins with low affinity causing weak and unintended binding (Zichi et al., 2008), and such unspecific binding could potentially impair the use of certain antibodies in research and clinical diagnostics. Consequently, it is important to always use optimal, functional reagents and highly stringent assays, and considerable care must always be taken in protein marker validation by antibody based methods (Ackermann and Berna, 2007). Screening and validation with spike-

in / dilution experiments or if possible either proteome array screening or mass spectrometry associated analyses of the captured antigens is recommended prior to using a binder in any type of affinity based application. Recombinant antibodies, in contrast to many commercial antibodies, generally go through extensive screening and validation prior to their use in affinity-applications. In order to ensure as high standard data quality as possible from current affinity based platforms, it is also important that the MIAPAR, the minimum information about a protein affinity reagent (Bourbeillon et al., 2010) guidelines become implemented. This would then, in the same manner as MIAPE (minimal information about a proteomic experiment) (Taylor et al., 2007), fulfill certain of the requirements frequently enforced on MS-based proteomic profiling experiments.

3.3 Current status of affinity proteomic based assays

The array-technology has been applied in clinical applications, demonstrating its potential for e.g. biomarker discovery and classification (Carlsson et al., 2008). The methodological capacity was demonstrated in recent clinical profiling efforts generating distinct disease associated or prognostic molecular portraits of the highly complex serum proteome (Carlsson et al., 2011b; Carlsson et al., 2011a; Sanchez-Carbayo et al., 2006). In addition, the platform has proven to be successful in targeting the urine proteome (Kristensson et al., 2012) and several membrane proteins (Dexlin et al., 2008; Dexlin-Mellby et al., 2011). Unfortunately, the majority of successful antibody microarray analysis has so far generally been limited to target only a couple of hundred analytes. One of the current bottlenecks is that the availability of high performing, specific, array-optimized high affinity antibodies is still limited. Consequently, this has so far limited the analysis to focused analysis and excluding large proteome profiling approaches.

An alternative to the conventional protein microarray format is array in solution coupled to flow cytometry sensing (Templin et al., 2004; Schwenk et al., 2007; Hartmann et al., 2009). In such set-ups, using antibody functionalized color-coded beads, large numbers of different bioassays can be performed and analyzed simultaneously. Currently, a commercially available platform allows multiplexing of up to 500 unique bioassays within a single sample (www.luminexcorp.com). In this context, antibodies within the Human Protein Atlas project have successfully been used together with the color-coded bead system for various proteomic profiling studies (Schwenk et al., 2008).

In addition to the above discussed assay setups, the usage of the nucleic acid aptamer platform has been demonstrated to successfully being used in highly parallelized assays with high specificity and sensitivity (Gold et al., 2010; Kraemer et al., 2011). When it comes to using limited amounts of sample or performing single cell *in situ* analysis, the elegant proximity ligation assay also offers an impressive sensitivity (Fredriksson et al., 2002; Fredriksson et al., 2007). The methodology might become an attractive alternative in biomarker research, particular due to its recent progress in successfully profiling 74 putative biomarkers with four 24-plex assays (Lundberg et al., 2011).

Clearly, any type of global affinity based approaches will need additional development before coming close to generating validated binder reagents against all human genome-encoded proteins and their variant forms and states. Issues such as:

- successful generation of all purified antigens needed for antibody generation
- potential off-target binding
- failed immunizations (in the case of pAbs and mAbs)

all need to be addressed and resolved. In addition, once all binder reagents are available potential further assay development /optimization will likely be needed in order to minimize issues such as cross-reactivity in order to successfully perform proteomic analysis. Furthermore, several of the current affinity based setups frequently require labeling of the samples and this might introduce additional assay variation or might even hamper certain epitopes.

4. The marriage of affinity proteomics and MS

The impact of proteomic research will be highly dependent on the availability of high quality methodologies that can successfully decipher any proteome of interest. Obviously, it will take some time before defining when a proteome of a cell type, body fluid or tissue is completely studied. Since, in the human body alone there exists over 200 different cell types, each presumably expressing only a subset of the over 20,000 predicted proteins at a given time (Nilsson et al., 2010). So far, as described in chapter 2 and 3, MS-based and affinity-based methodologies have been the two major workhorses individually used for proteomic studies. Both approaches are moving forward analyzing proteomes with impressive coverage and results (Beck et al., 2011; Ponten et al., 2011). However, a clear need for additional, complementing, improved proteomic assays and workflows still exists. Particularly, if measurements of the entire vast concentration range of proteins present in complex proteomes and large patient cohorts are going to be handled successfully. In order to bypass earlier discussed technological hurdles and advance further, some of the most attractive features of affinity proteomics and MS could be combined. The biological sample would then be digested and exposed to peptide-specific antibodies, where after any specifically enriched peptides would be detected, quantified and identified using MS and MS/MS. This would also enable the possibility of adopting label-free proteomic workflows. Another clear advantage of using such a combined approach is the usage of the mass spectrometer as the detector, since it will function as a "second antibody" and in most setups result in an actual identification of the targeted analyte. Hence, the approach would not only validate a specific binding interaction, but could potentially also compensate for off-target binding to certain antibodies.

4.1 Stable isotope standard capture with anti-peptide antibodies

Peptide immune affinity enrichment coupled with mass spectrometry based readout was demonstrated already in 2004 in the stable isotope standard capture with anti-peptide antibodies (SISCAPA) set-up (Anderson et al., 2004). The usage of such

antibodies can increase the sensitivity of the assay by > 10^4, which will be necessary for making successful measurements in clinical specimens (Anderson et al., 2009b). Anderson and colleagues recently proposed the initiation of the human Proteome Detection and Quantitation project (hPDQ) based on developing a complete suite of MS-based assays, e.g. two peptides from the protein product of each of the over 20,000 human genes. In order to achieve such an impressive goal the importance of using SISCAPA was emphasized (Anderson et al., 2009a). SISCAPA has now emerged as a technology with great potential for robust quantitative proteomic assays and has been further developed in terms of multiplexing and automation capabilities (Whiteaker et al., 2010). Additional binder reagents have been generated and validated (Whiteaker et al., 2011b) and the most recent significant advancements demonstrate and outline the utility of this targeted peptide immune-affinity SRM-assay for, in particular, large biomarker verification experiments (Whiteaker et al., 2011c). Both mAbs and pAbs have been used with the SISCAPA-platform. Polyclonal antibodies was the selected probe source in the recent SISCAPA large scale binder generation effort (Whiteaker et al., 2011b). This might cause potential batch-to-batch binder variation and impair performance in diagnostic applications. Notably, like conventional antibody array setups (Borrebaeck and Wingren, 2011), the SISCAPA platform also relies on the usage of one binder per unique peptide/protein, creating logistical issues when scaling up for global profiling efforts. In fact, the actual number of ready available peptide specific antibodies is still relatively scarce and it remains to be seen whether successful immunizations and binders can be generated for all intended targets. Hence, the SISCAPA is currently not viewed as a fully viable option for global discovery profiling efforts. In addition, these peptide specific antibodies have so far mainly been generated against human derived peptides resulting in that binders for additional model organisms will have to be generated in parallel efforts creating additional logistical issues.

4.2 Terminal anti-peptide motif specific assays

In order to circumvent the need of having to generate numerous antibodies, we (Wingren et al., 2009) and others (Poetz et al., 2009), recently presented concepts of using antibodies directed against short terminal peptide-motifs (epitopes) shared among up to hundreds of different peptides/proteins. When combined with mass spectrometry based detection, this approach would provide an inherent capability of probing any proteome in a discovery mode, even in a species independent manner, while still using a relatively limited number of antibodies. Furthermore, the enrichment generated from such capture step would result in less complex peptide

samples. The poor reproducibility in terms of overlap of MS/MS identified peptides between two technical LC-MS/MS runs of complex peptide mixtures is well known due to the stochastic nature of data dependent sampling (Tabb et al., 2010). Hence, peptide motif specific enrichments would result in less complex mixes and thereby enable far more reproducible measurements of the same peptides in several biological samples and in addition result in increased dynamic range compared to classical shotgun based methodologies.

4.2.1 Triple X Proteomics

The group of Joos et al termed their peptide motif-immunoaffinity methodology Triple X Proteomics (TXP) (Poetz et al., 2009) and focused on enriching groups of peptides that share a common epitope of three to four amino acids at the N- or C-terminal end. The first step in the development of a proteomics application based on the use of such peptide class-specific antibodies was demonstrated by Hoeppe et al and colleagues in 2011. They used polyclonal derived TXP antibodies as binder reagent and by specifically targeting β-catenin-derived motifs, the expected β-catenin peptides and a set of 38 additional epitope-containing peptides were successfully identified from trypsin-digested cell lysates. In addition, several of the identified peptides were previously not reported in the PeptideAtlas database, demonstrating the potential strength of the methodology (Hoeppe et al., 2011). In a recent *in silico* study, they also demonstrated (theoretically) that almost 50% of the non-redundant human proteome could be covered by only 100 different motifs (binders) (Planatscher et al., 2010). In order to achieve complete coverage, the number of required binder reagents was estimated to be about 2,000 solutions (motifs), whereby multi-peptide coverage was in addition achieved for 13,800 proteins (Planatscher et al., 2010). However, the choice of using pAbs might cause potential future batch-to-batch variation and to some extent impair performance in diagnostic applications. Another potential limitation of selecting pAbs as the probe source is that there is little chance of improving a generated binder using molecular based design and affinity maturation steps.

4.2.2 Global proteome survey (GPS)

We decided to focus on targeting only C-terminal tryptic peptides using selection motifs based on 4 or 6 amino acids. In contrast to the TXP-platform, we decided to develop the GPS-platform around fully renewable human recombinant scFv

antibodies, based upon a single fixed framework, generated through phage display technology, as our probe source (Paper I). This format has proven to be an excellent and stable probe resource for conventional antibody microarray set-ups (Steinhauer et al., 2002; Wingren and Borrebaeck, 2006) and we therefore expected the same stability and functionality in the intended GPS-setup. Furthermore, we expected this type of peptide binders to work well in most types of proteomes and potentially even in a specie independent manner and therefore termed them context independent motif specific (CIMS) scFv antibodies. A schematic outline of the GPS-assay workflow can be seen in Figure 1.

4.3 GPS characteristics

Below, I will discuss in more detail different aspects of our GPS-platform that have been addressed during the development and evaluation of the GPS methodology.

4.3.1 First generation of target motifs

Initially, 27 selection peptide motifs were designed, consisting of either 4 or 6 amino acids, and all motifs had a C-terminal lysine or arginine to mimic tryptic peptides, in order to be suitable for the intended proteomic workflow (Paper I). In order to evaluate terminal peptide motif frequencies, we took advantage of using the first manually annotated representation of all known human protein-coding genes (release 56.1 of UniProtKB/Swiss-Prot, composed of 20,325 nonredundant proteins). By performing an *in silico* tryptic digest of the database a total of 1 193 062 tryptic peptides were generated, whereof 58.5 % were represented within a mass fraction between 500-3500 Da. The frequency of different motifs was then addressed particularly within this mass range that fit in the detection range of the mass spectrometer.

4.3.2 CIMS-binder generation through phage display

The n-CoDeR® phage display library (BioInvent International AB, Lund, Sweden) has successfully been used for generating binders against proteins, haptens, carbohydrates, and peptides (Soderlind et al., 2000; Borrebaeck and Ohlin, 2002).

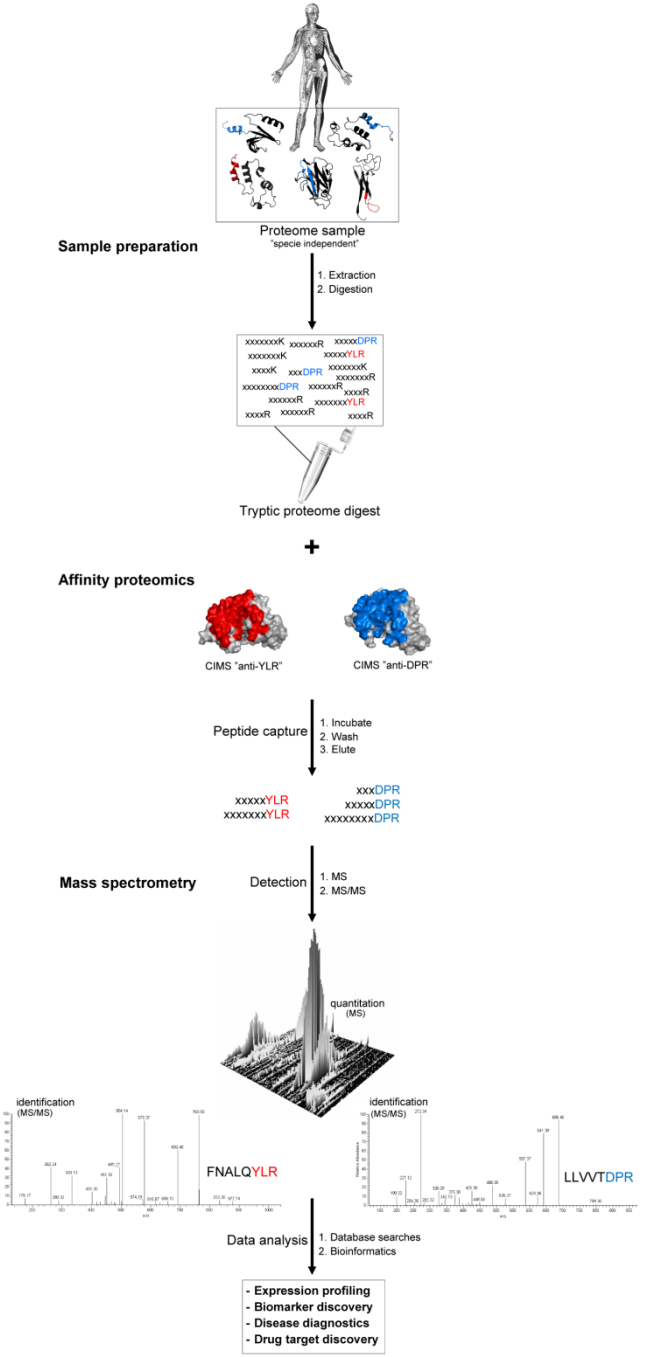

Figure 1. Concept of the GPS-assay.
A proteome sample (e.g. cell, tissue, plasma) is extracted followed by digestion with trypsin and then incubated with a set of CIMS-antibodies. These recombinant antibody fragments (scFv) specifically target short C-terminal amino-acid motifs and capture a wide range of tryptic peptides containing the target motif. Hence, the peptides captured can originate from up to several hundred different proteins. After removing unbound non-motif containing peptides by washing, the captured peptides are eluted and then directly detected and quantified using mass spectrometry. Through analyses of MS/MS-spectra and sophisticated database searches each peptide sequence can be determined and mapped back to its corresponding protein. By generating proteomic maps from both healthy and diseased samples, differentially expressed proteins can be identified. This may enable disease diagnostics, biomarker discovery and valuable proteomic insights of biological processes occurring in cells.

Several hundred antibodies derived from the n-CoDeR® library have also been successfully used as probes on conventional micro- and nanoarrays (Borrebaeck and Wingren, 2011). Hence, the n-CoDeR® library was therefore a natural choice to start with when generating peptide specific scFv reagents for the GPS platform. The library has been estimated to consist of approximately 2 x 10^{10} antibody clones and is referred to as a semi-synthetic library, combining a natural and synthetic approach through PCR-shuffling (Soderlind et al., 2000; Borrebaeck and Ohlin, 2002). The phage display technology was demonstrated already in 1985 by George Smith performing display of short peptides on individual phages (Smith, 1985), and five years later with display of antibody fragments (McCafferty et al., 1990) by introducing antibody DNA into the phage genomes via vectors. Phages have since been the most commonly used display system for antibody fragments (Bradbury and Marks, 2004). Having a direct "link" (genotype to phenotype) between the antibody, the gene encoding the antibody and the antigen recognized by the antibody, makes phage display an ideal platform for deriving antibodies in a high throughput manner (Hallborn and Carlsson, 2002; Bradbury et al., 2011). Unlike immunization, *in vitro* display methodologies provide direct availability of the sequence encoding the antibody and in addition well defined selection conditions can be adjusted to reflect the final use of the affinity reagent. The screening is often performed by incubating the phage display library with the immobilized antigen, in our case peptides (Paper I), on magnetic beads, followed by stringent wash steps to remove unbound clones (phage particles) followed by elution of bound phage particles. These eluted phages can then infect new fresh *E.coli* culture and the selection process is then repeated a number of times on beads before individual clones can be isolated. In Paper I, a total of 91 different recombinant binder molecules (CIMS) were generated. From this cohort of scFvs, a smaller dedicated set of clones that demonstrated promising characteristics in early screenings were further tested in detail.

4.3.3 CIMS-binder characteristics

The generated CIMS-binders were found to display affinities in the μM range against their corresponding selection peptide (Paper I). Recombinant antibodies against proteins often display affinities in the nM–pM range (Soderlind et al., 2000; Colwill et al., 2011) and such high affinities might not be expected by anti-peptide antibodies directly selected from scFv phage-display libraries unless subjected to additional affinity maturation. One reason to this is simply due to the size of the actual antigen and potential available epitopes on a protein compared to smaller

peptides. Affinities for anti-peptide monoclonal rabbit derived antibodies have recently been demonstrated to display affinities in the low nM range (Pope et al., 2009; Razavi et al., 2011). Despite the µM-range affinity displayed by the CIMS-binders, they still enabled very consistent and specific capture of tryptic peptides containing a narrow set of motif-like sequences in a clone dependent manner when exposed to crude tryptic digests (mouse, yeast, and human) (Paper I-IV).

In the majority of our immunoaffinity-peptide capture experiments, a clear trend was observed among the captured peptides. Besides the C-terminal arginine or lysine, it was apparent that relatively few key amino acid positions were essential. Often, only two additional amino acid positions, denoted anchor residues, appeared to be important for generating selectivity. For binders capturing 50 % arginine and 50 % lysine ending peptides, no permanent conclusion regarding selectivity in the terminal position could be drawn since lysine or arginine is present as the last terminal amino acid in any tryptic digest. Hence, a few binders displayed a binding pattern of peptides associated with certainty for only one very dominant amino acid position next to the C-terminal position and these characteristics were in particular illustrated with some of the "sister clones" (Paper I–III). We defined sister clones as binders selected against the same selection peptide, but displaying unique assay characteristics. In order to further evaluate the experimental observed specificities, we generated theoretical structural models for all selection peptides and for over 100 different captured peptides (Paper III). Based on the structural alignments some key structural properties were observed. Frequently, two amino acid positions next to each other were found to display their side chains in opposite directions (see Figure 2). Hence, this indicated a plausible explanation to the observed differences between some of the "sister clones" generated against the same selection peptide motif. Clearly, this might cause reduced selectivity in one of the two positions. Likely, there will often be a chance for certain flexibility in one of the last four terminal positions when compared to the original target motif unless other binder selection strategies, e.g. strict counter selections, are implemented during the binder generation.

Our findings were further confirmed when we generated WebLogo (Crooks et al., 2004) sequences of peptides captured from polyclonal derived TXP-antibodies (Hoeppe et al., 2011) where key positions also could be determined and the identity of certain neighboring residues was less crucial.

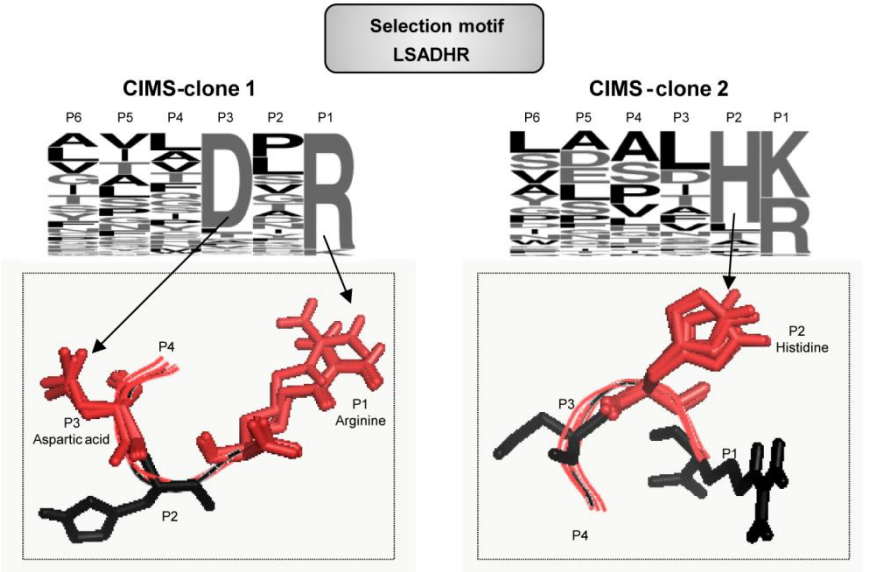

Figure 2. Binding patterns for two sister clones. C-terminal amino acid frequency analysis (Crooks et al., 2004) of captured peptides for two sister clones (Paper III). The same selection peptide was used for generating both antibodies. According to the output from the WebLogo peptide frequency sequences, it appeared as the arginine and aspartic acid were most important for clone 1, while the histidine appeared to be the most important for clone 2. Furthermore, peptide structure analysis for five captured peptides (red) and the selection peptide (black) is displayed. The peptide structure models were generated using PEP-FOLD (Maupetit et al., 2009) and consequently displayed and aligned on the last four terminal residues using PyMOL (The PyMOL Molecular Graphics System). The ribbon for the last four residues are displayed and aligned against the selection peptide. The most important residues are illustrated as sticks in red for the two cases. It appears as the histidine is displayed in a slightly different direction compared to the aspartic acid and the arginine. Hence, this might explain the selectivity observed between the two sister clones. Since clone 2 captured about 50 % arginine ending and 50 % lysine ending peptides, it might be reasonable to speculate whether the histidine reduced the selectivity in the last position.

Unless using binders with high affinity, stringent binding pattern, and a proteomic assay workflow adapting highly stringent washes, a certain wobbling would likely be observed. The wobbling should in general be seen as an attractive feature, since relatively few binders can generate relatively large proteome coverage (Paper II and Paper IV). This would thereby reduce the number of binders needed in order to achieve complete proteome coverage. However, for cases when targeting extremely low abundant analytes in highly challenging proteomes, i.e. plasma, the wobbling might hinder fully efficient captures and detection due to potential interference from high abundant analytes. Taken together, the results indicated that 3 amino acid

motifs spread over 4 positions could easily be achieved (Paper I, (Hoeppe et al., 2011), while a fixed motif of 4 consecutive amino acids might be harder to achieve. Furthermore, it is important to emphasize that the CIMS-binders generated in Paper I were not selected in terms of generating maximum affinities. As demonstrated in Paper I-IV, such 2-3 amino acid motif terminal binders still enabled highly successful and reproducible capture of many low abundant proteins within both the yeast and human proteome.

4.3.4 GPS setup optimization

Certain key methodological parameters have been addressed during the development of our GPS assay set-up. Initially, an offline affinity-packed micro-column for capturing the peptides was used (Paper I) followed by elution and manual transfer to the MS. Affinity-packed columns directly interfaced (in-line) with a mass spectrometry instrument might be highly valuable due to the possibility for direct sample injections and minimal sample loss. However, due to the amount of manual labor needed for manufacturing a column for each binder and issues (e.g. back pressure and reproducibility) associated with running the capture assay offline, we chose to adopt the platform into a much more flexible magnetic bead-based setup. This significantly reduced the manual lab work and furthermore opened up the possibility for improved throughputs, automation capabilities, and significantly improved reproducibility.

Anti-peptide antibodies with high affinity and low off-rates is of critical importance in order to retain peptides through the washing steps prior peptide elution and MS analysis, particularly if plasma and low abundant analytes are targeted. In order to minimize sample losses and compensate for the µM range affinities displayed by the CIMS, we chose to limit the number of washing cycles and downstream sample cleanup. When combined with a powerful LC-step, offering additional peptide separation and cleanup, prior to the powerful ESI-MS/MS analysis there was generally no need for performing extensive washings when analyzing tissue proteomes. The problems associated with ion-suppression commonly seen on reasonable complex peptide mixtures if using MALDI as detector was thereby to some extent also reduced with the online-LC-step. Hence, our developed setup and minimalistic workflow enabled highly successful and reproducible detection. The use of a MALDI based system would likely need an additional two cycles of washing or at minimum a c18-cleanup step in order to reduce potential ion suppression effects.

Furthermore, the multiplexing capability was initially demonstrated by mixing two different CIMS-antibodies resulting in identifications of over 90 human proteins (Paper I). However, in paper IV, a different approach was taken regarding the actual combination of binders used. When possible "sister-clones" were mixed and used together to potentially generate a synergistic effect. With this approach one of the mixes could quantify over 600 proteins using only two binders (Paper IV). Since peptides are expected to consistently bind and dissociate from the antibodies, the improved multiplexing results might potentially be explained by the fact that the binders together enabled to retain more low affinity peptides. In addition, the increased coverage could partly also be attributed to a further improved chromatography of the used LC-system. Even three different CIMS-binders (generated by different selection peptides) were successfully multiplexed (paper IV) and we have so far successfully mixed up to 4 CIMS-binders (data not shown). Potentially even more binders (depending on binder characteristics, desired assay performance and target abundances) might be combined. But in order to maintain the same total identification coverage as for individual captures, features like inclusion lists may then have to be implemented.

4.3.5 Label versus label-free

During the last decade, a wide variety of quantitative mass spectrometry based proteomics methods have been developed, adopting either label-free workflows or labeling workflows, such as metabolic labeling and isobaric chemical labeling using iTRAQ™ (isobaric tags for relative and absolute quantification (Ross et al., 2004)). Label-free quantification is widely used because it is applicable to samples from any source and there is no need for laborious and costly processes associated with labeling the sample. In contrast to label-free approaches, multiplexed isotopically labeled samples can be simultaneously analyzed resulting in an increased throughput. One of the most popular and widely adopted metabolic labeling approaches is the stable isotope labeling by amino acids in cell culture (SILAC) methodology (Ong et al., 2002) which has successfully been implemented in model organisms such as yeast, drosophila, and mouse (Gruhler et al., 2005; Sury et al., 2010; Kruger et al., 2008). The main advantage of SILAC is the possibility to combine several samples already prior to protein extraction and thereby reduce downstream experimental variation. The accuracy of SILAC-based experiments might in some cases become impaired by the metabolic conversion of arginine to proline, resulting in additional heavy labeled proline peptide satellites. This may

however be circumvented by either using lower amount of arginine or addition of label free proline (Lossner et al., 2011).

Within this thesis, I have tested both a label free strategy (Paper I, IV) and a labeling approach (Paper II). The common feature between these two methodologies is that both are based on quantifying the MS1 intensity. In Paper I and IV, we demonstrated that the label-free GPS-methodology delivered highly reproducible quantitative data. For label-free experiments, a sophisticated feature alignment software is needed and we took advantage of using the Progenesis LC-MS software which was demonstrated to perform well in a recent evaluation (Sandin et al., 2011). In Paper II, we took advantage of using the SILAC methodology and associated MaxQuant software (Cox and Mann, 2008). The SILAC approach also generated highly accurate quantitative data. Taken together, both quantification strategies proved to be successful when combined with our GPS platform and this leaves the door open for implementing any of the strategies in future GPS-based discovery applications.

4.3.6 Assay reproducibility

Substantial efforts have been undertaken within this thesis to address key issues, such as reproducibility and dynamic range of the GPS-platform (Paper I, II, and IV). The reproducibility in terms of quantification was identified to be highly satisfactory (Paper I, II and IV). This was in particular demonstrated in Paper II, were two proteomes were mixed at different ratios by taking advantage of the SILAC approach. The GPS enabled highly accurate measurements of the two mixed proteomes over a wide range of protein abundances. High correlation between replicate captures were achieved and thereby clearly indicated that the capture step did not introduce any drastic increase in variance when simultaneously capturing different analytes spanning 3-4 orders of magnitude (Paper II).

The reproducibility of the assay was also demonstrated for the multiplexed label-free setup (Paper IV) where a median CV of 10.8 % was achieved for three independent captures and quantification of over 600 proteins. When data from all mixes (in total 9 binders split on 4 mixes) were combined, this resulted in a median CV of 10.8% when 1264 proteins were included (Paper IV). This is comparable to other existing methods, such as the gold-standard SRM, where CV values around 10-15% have routinely been achieved (Addona et al., 2009). However, the above discussed CV-value range for SRM based assays has frequently been for optimized

assays, high performing peptides, and at most a couple of hundred proteins while our data was based on 1264 proteins and no optimization. Clearly, it might be possible to improve the CVs even further by excluding some of the poor performing peptides and combining the GPS with SRM-based readout. When we distributed triplicate capture runs over a week (one antibody mix on the pooled tumor sample, Paper IV) or a month of instrument time (all combined data for four antibody mixes on pool sample, Paper IV) the total median CV for the same 1264 proteins increased to 22.8 %. The slight dip in performance could be attributed to the actual LC-system with drifts in retention times and buildup of contaminants. In addition, the LC-column was replaced twice during these runs. Taken together, this clearly demonstrated the reproducibility of the entire GPS assay and in particular the capture step, since there will always be some variation associated with the actual LC-MS/MS analysis step.

4.3.7 Dynamic range and evaluation against SCX

The dynamic range of the GPS methodology and individual CIMS-capture experiments was addressed both in paper I and II by comparing identified proteins from our yeast capture experiments to absolute protein abundances generated by orthogonal methods (Ghaemmaghami et al., 2003). The results demonstrated that proteins could be detected in a wide range of abundance values spanning from over a million copies / cell down to less than 50 copies / cell. Hence, we demonstrated that the entire known concentration range within a yeast cell could be targeted and that several of the binders were not limited to only the most abundant analytes. This is promising, since a need for developing SRM-targeted assays in order to enable successful measurements of low abundant analytes has often been expressed (Picotti et al., 2009).

In addition, we evaluated the GPS generated data against the conventional SCX methodology (Paper II). Not surprisingly, the SCX-based fractionation setup generated a significantly larger total coverage, since the GPS only used a small set of binders and less instrument time. However, the aim was not to benchmark in terms of total coverage, but rather to compare the overlapping quantified peptides and proteins. Good agreement was generally displayed both at peptide and protein level between the two methodologies and we could conclude that the capture step did not introduce any biased or increased quantitative assay variation compared to conventional methodologies. Even more importantly, the entire GPS assay was capable of generating reproducible quantitative data based on the same set of

peptides between biological replicates. This is a crucial methodological feature required for generating quantitatively reliable information over many biological samples.

Noteworthy, the number of peptides previously not reported was found to be substantial when comparing against the PeptideAtlas database (Deutsch et al., 2008). Furthermore, the median length of a majority of all GPS detected peptides was found to be substantially shorter (Paper I-IV) when compared to the peptides identified with either the SCX methodology (Paper II) or classical shotgun methodologies reported in the PeptideAtlas database. The possibility for detecting shorter peptides was particularly seen in Paper II and IV. This characteristic indicated that GPS enabled analysis of an entirely new set of peptides. The reason behind that the peptides had not previously been observed might be that shorter peptides are masked by longer peptides in more complex mixtures using conventional proteomic MS-analysis. Furthermore, due to potentially higher affinities these shorter peptides may also be favorably enriched by antibodies compared to longer peptides. Consequently, when analyzed by the mass analyzer, the shorter peptides naturally will have a higher chance of being detected due to their increased abundance.

4.3.8 First application of GPS - yeast

Yeast has been used extensively for evaluating methodologies within proteomics (Washburn et al., 2001; Ghaemmaghami et al., 2003; de Godoy et al., 2008; Picotti et al., 2009; Nagaraj et al., 2011a) particularly due to the facts that the yeast proteome is less complex and the ease with which yeast can be experimentally manipulated. Since the analysis of mammalian cell proteomes is much more complex and technically challenging, yeast can be viewed as a window into the future of proteomics-based biological research. The applicability of the GPS-platform was demonstrated by applying it to a fundamental biological question within yeast biology, i.e. what proteins and pathways are differentially expressed when yeast uses glucose or ethanol as its main carbon source (Paper II). We demonstrated that a rather large qualitative and quantitative reproducible map of the yeast proteome could be generated in a rapid manner with just six CIMS-binders. The GPS technology successfully detected several differentially expressed proteins and as expected many were related to the tricarboxylic acid cycle and glyoxylate cycle, displaying a marked induction upon growth in the presence of ethanol as its carbon source. Similar expression trends have been observed using either 2D-gels

(Kolkman et al., 2005) or sophisticated SRM-assays (Costenoble et al., 2011). In addition, a set of uncharacterized proteins were identified to be significantly up-or down-regulated and several of these proteins and their functions may be of significant interest to further explore for development of potential ethanol tolerant yeast strains.

4.3.9 Next generation of motifs and binders

Based on experimental data (Paper I-III), we continued to theoretically generate and evaluate new terminal amino acid motifs in Paper III. Since mainly peptides of shorter length were observed in the initial studies (Paper I-III), we limited the new targeted tryptic cohort of peptides to be between 800-2000 Da. In addition, a software script was developed in order to evaluate large cohorts of motifs and whether a specific motif was present in some of the most high abundant plasma proteins or not (Paper III). Plasma and serum are the overwhelmingly predominant clinical specimens available for routine molecular analysis and all molecules present in plasma (or serum) have clearly the widest diagnostic potential (Anderson, 2010). However, the abundance of proteins in plasma proteomes exceeds 10 orders of magnitude and this has, so far, prevented detection of many low-abundant proteins in proteome analyses (Anderson and Anderson, 2002). One reason behind this is that the top 10–15 highest abundant proteins such as albumin (or human serum albumin, HSA), immunoglobulins (Ig), transferrin, etc. represent only < 0.1% of the diversity of proteins, yet constitute more than 95% of the mass of total plasma proteins (Anderson and Anderson, 2002). These high abundant proteins, and in particular albumin, thereby mask or interfere with the detection of other low abundant proteins. Immunoaffinity depletion methodologies have been used during the last decade in order to remove some of the most abundant analytes (Liu et al., 2006; Tu et al., 2010), however untargeted proteomic analysis using LC-MS/MS platforms even with immunodepletion cannot efficiently discover low-abundant disease-specific biomarkers in plasma (Tu et al., 2010). Hence, if CIMS-binders against motifs present in medium and low-abundant plasma analytes could successfully be generated our GPS platform could potentially also be highly suitable for future plasma proteome analysis. Efforts in generating CIMS-binder reagents against the newly designed motifs will be continuously ongoing.

5. Molecular profiling of breast cancer using GPS

Recently, it was estimated that one third of all people in Sweden will be diagnosed with cancer during their lifetime (Bergman et al., 2009). The outcome for a specific cancer patient is primarily related to how early and correct the diagnosis can be set (Etzioni et al., 2003). However, in reality the diagnosis might be ambiguous and the prognosis of the cancer disease and treatment response may in many cases be difficult to determine. The existing traditional clinic pathological parameters are far from perfect and the identification of highly specific disease associated markers together with prognostic markers (at genomic or proteomic level) will be crucial in order to advance. A biomarker is generally a molecule that is consistently modified or present at abnormal concentrations in specific illness or other health conditions. Such information could directly influence treatment selections, minimize patient suffering, maximize the effect of each treatment and fully enable the concept of personalized medicine/individualized therapy to be implemented. The quest to decipher protein alterations in cancer has spanned well over half a century and massive amounts of complex and heterogeneous proteomic data are currently being generated (Hanash and Taguchi, 2010). Clearly, future cancer treatment will be based on personalized approaches that begin with identifying critical molecules necessary for tumor growth and survival that then consequently decide optimal molecularly directed therapies. One step towards such personalized cancer treatment was recently demonstrated in a study using whole tumor genome sequencing on patients followed by treatment determined by convening a sequencing tumor board of experts (Roychowdhury et al., 2011). The scientists were able to find key mutations within 24 days of biopsy to a cost of ~$3600 per patient and potentially matched to clinical trials, hence a short enough time and reasonable amount of money spent to be clinically useful. Given the rapid advances of next-generation sequencing and the drastically reduced prices, the day may soon come when full genome and tumor genome sequencing could become routinely adopted in clinics. In parallel, the entire field of cancer proteomics also clearly has a chance to offer cutting-edge capabilities to accelerate the translation of basic discoveries into daily clinical practice (Cho, 2007).

Within this chapter, I will briefly cover some of the current breast cancer associated markers, past breast cancer proteomic efforts and then continue to discuss some of the results generated from our proteomic profiling efforts on breast cancer tissue (Paper IV) using the developed GPS platform.

5.1 Current status of breast cancer associated markers

Breast cancer is by far the most commonly diagnosed cancer (roughly 30 %) worldwide among women (Jemal et al., 2011). Treatment usually involves surgery to remove the tumor frequently followed by radiation therapy, endocrine therapy and/or chemotherapy. Traditional clinic pathological parameters, such as histological grading, tumor size, age, and lymph node involvement are used to decide treatment and prognosis. Furthermore, a panel of molecular markers has become available that may aid in tumor classification, treatment choices, and predict prognosis.

5.1.1 BRCA1 and BRCA2

The first evidence for a gene associated with hereditary breast cancer came with the identification of truncating mutations in the coding sequence of BRCA1 (Miki et al., 1994), and shortly thereafter BRCA2 (Wooster et al., 1994), in families with multiple cases of breast cancer. Both these tumor suppressor genes are involved in the maintenance of genomic integrity and DNA repair (Narod and Foulkes, 2004) and the discovery has increased the understanding of both hereditary and sporadic forms of breast cancer. BRCA mutations occur at a frequency of 1 in 250 women (Narod and Foulkes, 2004) and genetic testing for breast cancer susceptibility has to some extent become implemented into the practice of clinical oncology and enabled mapping of certain populations and risk groups (Marroni et al., 2004). After positively testing for BRCA1 or BRCA2, certain preventive and risk reducing measures like increased frequency of mammography, chemoprevention or prophylactic surgery can be taken (Tuttle et al., 2010).

5.1.2 Transcriptomics derived classifiers

Breast cancer is a heterogeneous disease and key evidence, at the transcriptomic level, established that estrogen-positive and estrogen-negative tumors are

fundamentally different (Perou et al., 2000). Shortly thereafter, a classification based on five different molecular subtypes was reported: luminal A, luminal B, normal-breast-like, human epidermal growth factor receptor 2 (HER2), and basal-like (Sorlie et al., 2001). In addition, the first successful prognostic gene signature was identified by van't Veer and colleagues separating a group of patients with good prognosis and minimal risk of development of distant metastasis (van 't Veer et al., 2002). Currently, two major breast cancer gene expression assays are available: 1) the MammaPrint® (Agendia, Amsterdam, The Netherlands) using the 70-gene expression profile developed by van´t Veer (van 't Veer et al., 2002) and 2) the Oncotype DX™ (Genomic Health, Redwood City, CA, USA) quantifying 21 genes using RT-PCR. The Oncotype DX™ test generates a score that correlates with the likelihood of a woman's chances of a recurrence, and the likelihood that she will benefit from adding chemotherapy to her hormonal therapy. MammaPrint® was the first assay based on microarrays that succeeded in going all the way from the research lab to FDA approval for clinical application. The full potential and results from clinical usage of these assays are currently evaluated in the MINDACT trial (enrollment of over 6000 patients for evaluating MammaPrint®) and in the TAILORx study (enrollment of over 10000 patients for evaluating Oncotype DX™) (Colombo et al., 2011).

5.1.3 Protein based markers

The main biological markers recommended for use in breast cancer decision making of treatment choices, are estrogen receptor (ER), progesterone receptor (PR) and HER2 (Hondermarck et al., 2008). Estrogen positive patients generally respond to treatment with anti-estrogen, such as tamoxifen (Ciocca and Elledge, 2000). In addition, about 20 % of breast tumors overexpress the surface marker ErbB2 (commonly referred to as HER2 or HER2/*neu*) (Hondermarck et al., 2008) and these tumors have proven to be rather successfully treated with Herceptin, a blocking-antibody, directed against ErbB2 (Slamon et al., 2001). Most clinical laboratories are using IHC for determining the HER2-status, however, a recent study by Philips et al 2009 revealed that one in five HER2 tests might generate incorrect results. In addition, the scandal in Newfoundland, Canada, where nearly 400 of 1000 patients received incorrect test results of the ER-status of their tumors (Hede, 2008), attests to the problems associated with consistency and reproducibility in the area of quantitative IHC (Leong and Zhuang, 2011). Additional tests for HER2 using fluorescence *in situ* hybridization (FISH) detection systems (Allison, 2010) or a proximity-based assay (HERmark) measuring the total amount of HER2 (Huang et

al., 2010) have been developed, but researchers have yet to come to an agreement regarding if there is a single optimal HER2 test.

5.1.4 Proteomics derived markers and current status

The output, in terms of FDA approved biomarkers derived from proteomics, has so far been very disappointing (Whiteaker et al., 2011a). This can partly be assigned to some of the short-comings (see chapter 2) that classical mass spectrometry based proteomics have suffered from. The poor quality seen in early proteomic studies have highlighted the need for relevant study designs, precise sample handling, stringent data analysis, and robust technology platforms offering high sensitivity and specificity (Mischak et al., 2007; Borrebaeck and Wingren, 2009). Further fueling the debate concerning proteomic limitations, came when simple meta-analysis was performed on 186 published 2-DE experiments on human, mouse and rat tissues (Petrak et al., 2008). The results from the meta-analysis demonstrated that certain proteins and protein families were strikingly overrepresented as differentially expressed in most of the studies regardless of the experiment, tissue or species. The most recurrent protein, enolase 1, was found to be identified in every third experiment on both human and rodent tissues (Petrak et al., 2008). Other common proteins observed were heat-shock proteins and keratins. Valid concerns were raised that some findings might be attributed to technical artifact and limitation of the 2DE-method or alternatively the top candidate proteins represent universal cellular sensors (Petrak et al., 2008). Since few putative biomarkers undergo rigorous validation it also results in that the literature is flooded with lengthy lists of candidate markers without follow up (Polanski and Anderson, 2006). Hence, caution and extensive validations should always be taken prior to defining a protein as a marker or uniquely disease associated and this is going to be crucial if any proteomic derived markers are going to be successfully implemented into the clinics.

The challenges of handling breast cancer tissues in comprehensive proteomics approaches have been well described (Hondermarck, 2003). The tumor biopsies generally consists of epithelial cells, nerve fibers, myoepithelial cells, circulating cells, adipocytes, fibroblasts, and endothelial cells and thereby form very heterogeneous proteomic samples. Protein extracts from such biopsies have frequently been analyzed with 2DE and facilitated analysis of hundreds of proteins (Deng et al., 2006; Kim et al., 2009). However, the analysis has often been limited to high abundant proteins, due to problems with dynamic range and sensitivity.

Many candidate markers has been reported, however, none has yet been fully validated and approved. In order to minimize the heterogeneity between individual patients and between different cell types, laser capture microdissection (LCM) has been tested, but with generally lower proteome coverage (Hill et al., 2011). Furthermore, profiling experiments were recently performed on needle-biopsied breast cancer tissues and this resulted in identification of 2,331 proteins (Yang et al., 2012). However, the study needed massive instrument time and still just reported profiles for 11 samples and the quantitative data was determined by spectral counts. Labeling approaches, like iTRAQ, combined with LC-MS/MS adopted workflows, have also been tested (Bouchal et al., 2009) resulting in identification of 605 proteins when comparing low-grade breast primary tumor tissues with and without metastases. Analyzes of breast cancer cell line proteomes together with SDS-PAGE and a LTQ-Orbitrap setup have enabled larger coverage with identification of 3,481 proteins (Strande et al., 2009). Recently, a different approach was taken by mixing SILAC-labeled cell lines with breast tumor samples (Geiger et al., 2010). This enabled quantification of 4,336 protein groups by implementing peptide fractionation, long elution gradients (190 min) and the usage of state of art mass spectrometry instrumentation. The impressive results from the super-SILAC approach indicated that the methodology could enable accurate quantification of proteome differences between tumor types, but in order to identify true biomarker candidates, more than single tumor tissues need to be analyzed and thereby require massive instrument time. Finally, when it comes to serum profiling of breast cancer patients, one study by Carlsson and colleagues is worth mentioning since they elegantly demonstrated a serum biomarker signature able to classify patients with primary breast cancer according to their risk of developing distant recurrence (Carlsson et al., 2011a).

5.2 GPS and breast cancer

Histological grading is currently one of the most frequent classical clinical parameters used to evaluate a breast tumor. It is a combined score, based on microscopic evaluation of morphologic and cytologic features of tumor cells, illustrating the aggressive behavior of a tumor (Elston and Ellis, 1991). The sum of these scores stratifies breast tumors into: grade 1 (slow growing and well differentiated), grade 2 (moderately differentiated), and grade 3 (highly proliferative and poorly differentiated). However, the scores are hard to determine and could thereby potentially become operator dependent (Robbins et al., 1995; Frierson et al., 1995). In paper IV we used our developed GPS platform in order to see whether we

could generate molecular protein profiles that could be used to distinguish the different tumors based on histological grade, ER-status and HER2-status. In total, 52 samples were profiled. Over 2,000 protein groups were identified using only 9 CIMS-antibodies and the label-free workflow enabled quantitative data for over 1,300 proteins. A protein signature consisting of 49 analytes, displaying differential expression among the three histological graded groups of samples, was defined. The majority of these 49 analytes displayed either a continuously increasing or decreasing trend over the three histologic graded tumor types. Hence, we could potentially establish a link between a molecular signature and tumor cell progression. The biological processes for some of the proteins in the generated signature together with some of the findings regarding the ER-status and HER2 delineated signatures are briefly discussed below.

5.2.1 GPS identified cancer hallmarks

Tumors were originally viewed as masses of proliferating cancer cells that had acquired a set of capabilities and essential alterations in their physiology that collectively dictate malignant growth (Hanahan and Weinberg, 2000). However, it has become apparent that the biology of tumors can no longer be understood simply by focusing on the traits of the cancer cells and should instead encompass the entire tumor microenvironment (Hanahan and Weinberg, 2011). Tumors are clearly highly complex tissues composed of multiple distinct cell types involved in heterotypic interactions with one another (Hanahan and Weinberg, 2011). The six classical suggested cancer hallmarks (Hanahan and Weinberg, 2000) complemented with two emerging hallmarks and two enabling ones (Hanahan and Weinberg, 2011) can all be seen in Figure 3. Noteworthy, 10 of the 49 proteins identified as differentially expressed between the three histological grades (paper IV) were mapped to the hallmark groups in Figure 3.

Interestingly, a majority of the genes in the MammaPrint® signature could also be mapped to most of the cancer hallmarks. Genes in a gene signature are often interchangeable, demonstrated in a reanalysis of the van't Veer et al. (2002) data set (Ein-Dor et al., 2005), and it is often instead the functional processes captured by a gene signature and not the individual genes that are more of importance for successful signatures (Wirapati et al., 2008). We found that cell proliferation associated proteins, like cyclin dependent kinase 1 (CDK1) and minichromosome maintenance complex component 3 (MCM3) displayed increased expression trends in tumors of higher histological grade. Proliferation has been recognized as one of

the key prognostic factors in breast cancer, and has been found to be one of the major components of several prognostic gene expression signatures (Desmedt et al., 2008; Wirapati et al., 2008).

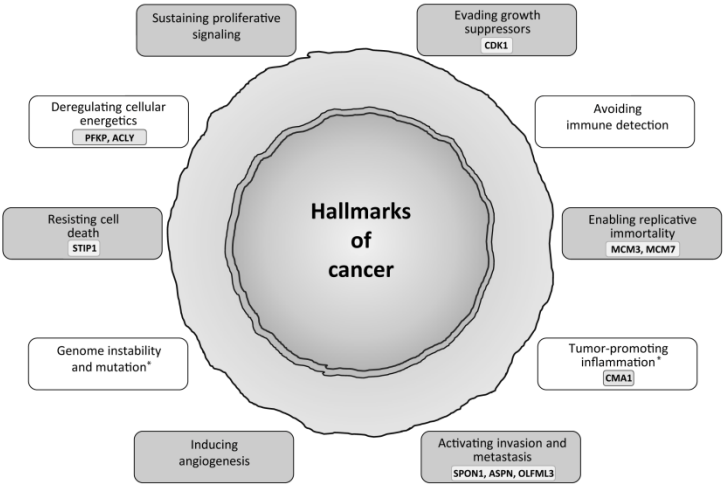

Figure 3. Hallmarks of cancer. Depicted are 10 of the 49 significantly differentiated proteins identified between low and high graded tumors (Paper IV), mapped and associated to a certain cancer hallmark. The original six cancer hallmarks (capabilities acquired during the multistep development of human tumors (Hanahan and Weinberg, 2000)) were indicated with grey boxes. Recent research have suggested that two additional emerging hallmarks of cancer (i.e. deregulating cellular energetics, and avoiding immune detection) are involved in the pathogenesis of cancers and were therefore also illustrated (Hanahan and Weinberg, 2011). Additionally, two consequential characteristics* are also illustrated (i.e. tumor-promoting-inflammation, and genome instability and mutation) of neoplasia facilitate acquisition of both core and emerging hallmarks (Hanahan and Weinberg, 2011).

Another key observation made within our study was that several extracellular matrix related proteins, such as asporin (ASPN), spondin (SPON1) and olfactomedin-like protein 3 (OLFML3) displayed a decreasing expression trend with drastically decreased expression levels in tumors of highest histological grade. Stroma cells clearly contribute to the development and expression of certain hallmark capabilities, and we demonstrated further support for this at the protein level since the extracellular matrix appeared to go through drastic changes between the lowest and highest graded tumors (paper IV). Furthermore, proteins important for adopting metabolic changes like ATP-citrate synthase (ACLY) were observed to display increased expression in higher graded tumors. Cancer cells have a higher need for energy than normal differentiated adult cells and in order to maintain their rapid

growth and proliferation the increased demand is, in part, met by an altered metabolic program.

5.2.2 GPS as classifier of estrogen-status and HER2

In addition to the above discussed findings, we also identified 39 proteins as differentially expressed between the ER-positive and ER-negative samples. One particularly interesting differentially expressed protein was GREB1, since it displayed a similar profile as reported in the literature with highest measured expression in the ER-positive tumors. GREB1 is an estrogen-regulated gene that mediates estrogen-stimulated cell proliferation and has been suggested as a surrogate marker for ER, a marker for response to endocrine therapy as well as a potential therapeutic target (Hnatyszyn et al., 2010; Rae et al., 2005). Furthermore, the difference between HER2-positive and HER2-negative samples was evaluated and resulted in 5 significantly differentially expressed proteins. Noteworthy, one of the proteins in this signature was HER2 itself and the GPS indicated high HER2-expression level for three of five positive samples.

Taken together, this indicated that all three generated signatures, i.e. for the histologic grade, ER-status, and HER2, have not only generated new insights at the protein level into breast cancer biology, but could also have a potential to complement or even replace some of the current pathological IHC-based evaluation systems. Having an approach capable of generating fully quantitative protein values is anticipated to enable far more precise classifications of each individual tumor compared to classical IHC-based methodologies.

5.2.3 Validation of GPS generated breast cancer signatures

Proteomics generated data should always be validated. Often, western blots are used, however such analysis may be less quantitative, less specific, and therefore less informative (Mann, 2008). Performing GPS with SRM quantification on a new set of independent samples offers a far more attractive alternative, but was beyond the scope of this first GPS clinical application study. As an initial validation of our findings, we took advantage of using large publicly available gene expression data sets (Ringner et al., 2011). When analyzing all three signatures, i.e. for the histologic grade, ER-status, and HER2, high agreement between mRNA and protein was seen for a majority of the analytes. The large cohorts of independent breast tumor

samples present in the transcript profiling database (total 1881 samples, whereby 1411 samples defined with histologic grade) very clearly supported the data and trends observed in Paper IV. It is worth mentioning that some analytes did not correlate, like serum amyloid P-component with observed down-regulation at the protein level in the most advanced tumor grade but no differential mRNA expression in the database. Nevertheless, the general agreement between protein data (for all three signatures) and the mRNA for most of the analytes was striking.

In conclusion, we have successfully used our GPS platform for generating valuable information associated to breast cancer biology and a molecular signature reflecting tumor cell progression was established. Clearly, the entire tumor micro-environment provides highly valuable information. Several of the reported proteins have a potential for improving prognosis and even therapeutic planning. Some of the proteins, like CDK1, could also be highly relevant as novel potential therapeutic targets for breast cancer patients (Johnson et al., 2011). The greatest challenge in developing clinically valid signatures to be used for breast cancer diagnosis, prognosis, and prediction of treatment responses will be related to the heterogeneity of breast cancer. Since, what is actually a fully true tumor signature and what cells / genomes / proteomes should it be based on? In a recent proteomics study by Kang et al focused on the interface zone, the region between an invading tumor front and normal tissue, reported a unique molecular profile potentially related to progression and metastasis (Kang et al., 2010). The optimal future diagnostic sample format would clearly be plasma, but when it comes to discovery profiling efforts tissues will continue to provide many novel insights into tumor progression.

6. Concluding remarks

During the last decade, the possibilities of performing global analyses of entire genomes and associated mRNA transcripts have become a reality and many of these tools are today fully mature and offer highly consistent and sensitive analysis of large cohorts of samples. Unfortunately, the same statement cannot be made for the field of proteomics, where there still exists a need for further methodological, technological, bioinformatics, and standardization advancements. Hundreds of millions of dollars have already been spent on proteomics research in hope of generating potential new biomarkers and drug targets, but so far the outcome, i.e. fully validated disease specific markers, has been very limited (Anderson, 2010; Whiteaker et al., 2011a). This large gap between candidate biomarker proteomics and clinical biomarker output has partly been attributed to technical barriers. Hence, what has been explored so far is to some extent just the tip of the iceberg and the human proteome(s) are clearly much more complex.

The aim of this thesis was to develop a novel and sensitive proteomic methodology. By combining some of the most attractive features of affinity proteomics and mass spectrometry, I have in this thesis demonstrated proof-of-concept, developments, and biological applications of our proteomic discovery methodology entitled GPS. The data presented herein demonstrate a semi-high-throughput bead-based immunoaffinity assay offering high reproducibility, sensitivity and quantitative capability transforming affinity proteomics into a global discovery engine. The sensitivity was demonstrated when peptides, originating from proteins present in a range of abundance values spanning over a million down to less than 50 copies per cell, could be identified and quantified in the yeast proteome. In addition, valuable information regarding binder characteristics and peptide properties was discovered. This information will be valuable in future design of selection peptides and binder generation efforts.

The applicability of our peptide immunoaffinity-based MS workflow was demonstrated in two proteome profiling studies and we report a number of highly interesting proteins particularly in the breast cancer study. To the best of our knowledge, our breast cancer study is one of few large scale crude tissue proteome

profiling efforts succeeding in measuring a number of expected low abundant proteins without using laborious sample preparation, fractionation strategies, and long instrument times. Even more importantly, several highly relevant cancer associated proteins were identified discriminating between the histologic grades. We could thereby potentially establish a link between a molecular signature and tumor cell progression. These findings are going to be of great importance not only for a deeper understanding of tumor biology, but also in directing the development of therapeutic targets and future diagnostics. Although it is too early to know if any of the identified proteins will be of clinical value, or as in the case of the yeast, industrial value, we believe these initial findings are promising and the data will be validated in independent samples.

Taken together, the GPS-platform displays several attractive features, for both focused and global discovery proteome analyses, and can be used in a specie independent manner enabling rapid and sensitive analyses of a relatively large set of proteins and samples. This was particularly demonstrated in Paper II and IV, where a very limited set of binders and minimal instrumentation time successfully enabled relatively deep proteomic insights in many samples. The methodology can either be used alone or work in concert with other existing methodologies in the current proteomic research and diagnostic field (see Figure 4). In addition, the platform will in the near future be expanded with a number of additional binder reagents, thus raising the potential and profiling capabilities of the GPS assay even further. Furthermore, it can be envisioned that GPS has the potential for being introduced and fine-tuned in clinical settings by using a SRM-based readout for targeted sets of proteins.

The complex battle against cancer and, in particular, breast cancer will continue for many years to come, but hopefully some of our identified differentially expressed proteins may shed some new light. Future research will be needed in order to fully determine the functional roles, potential diagnostic value and potential therapeutic targeted value several of these cancer associated molecules may have. I generally believe there still is a chance for many proteomic generated biomarker success stories but it remains to be seen if any of the reported candidates will fulfill requirements and make it into the clinical arena.

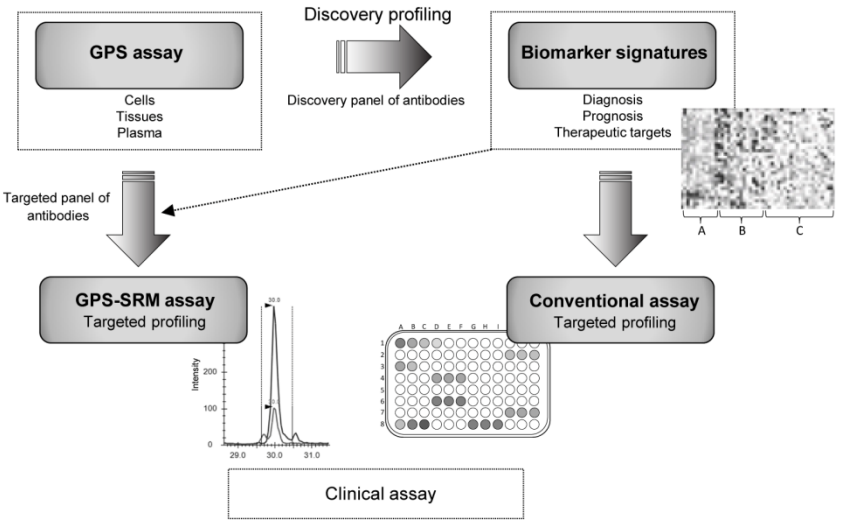

Figure 4. Possible GPS workflows and clinical applications. The GPS methodology is suitable both for discovery efforts and more targeted analysis. One attractive workflow would be to first perform a discovery study and when a set of key targets has been identified, transfer profiling efforts into a GPS-SRM assay for validation and as a clinical assay. The outcome from a discovery effort might naturally also be implemented into the clinic using conventional assay setups. However, one of the advantages of the GPS-SRM setup would be the minimal need for extensive new assay development, since the same CIMS-binding reagents and sample format could be used. It could also be envisioned that biomarker signatures adopted for conventional assays or GPS-SRM assays will complement or in the long term run outcompete current IHC analysis.

In order to fully understand a certain phenotype a genomic, proteomic or metabolomic view alone is often not sufficient. However, now when more and more high-performing omics-approaches become available, these methods might finally together enable increased understanding of complex biological systems, like breast cancer. Let's hope that current and future proteomic molecular profiling efforts can finally pave the way for unique disease associated classification signatures, therapeutic targets and personalized treatment of patients. This would then fulfill the hopes and goals that the proteomics community, to some extent so far, have failed to deliver back to the patients.

Populärvetenskaplig sammanfattning

En av de stora vetenskapliga bedrifterna hittills under 2000-talet är onekligen att man efter många års arbete för första gången lyckades kartlägga (sekvensera) samtliga mänskliga gener. Det mänskliga DNA:t är lokaliserat till cellkärnan och i princip så kan man likna en cellkärna med ett kontrollrum fyllt av manualer (DNA) som talar om cellens sammansättning och hur dess maskineri ska byggas ihop. Trots att det gått mer än 10 år sen det första humana genomet kartlades är majoriteten av genernas funktioner fortfarande okända. För att få en ökad förståelse måste parallella studier på genprodukterna dvs. proteinerna göras. Beroende på typen av inkommande signaler till cellkärnan kommer nämligen olika manualer (gener) börja läsas av vilket leder till att s.k. budbärarmolekyler (mRNA) tillverkas och dessa molekyler transporteras i sin tur ut från cellkärnan till specifika ställen i cellen där de kan uttrycka sitt budskap, dvs. budskapet kodas av, och ett specifikt protein kan bildas. Proteiner är de molekyler som aktivt utför cellens önskade order/funktioner och uppskattningsvis så finns det över 20 000 olika humana proteiner. I en mänsklig cell antar proteinerna en mängd olika former och denna enorma komplexitet ställer därmed mycket stora krav på de analysverktyg som behöver användas för att studera ett helt proteom dvs. alla protein som finns i en viss celltyp vid en viss tidpunkt. Eftersom de flesta läkemedel är riktade mot proteiner samt det faktum att många av nuvarande diagnostiska tester för en människas olika sjukdomstillstånd är baserade på protein-analyser gör sammantaget att behovet av nya metoder för heltäckande proteomanalyser är mycket stort. Nyttan av att snabbt och pålitligt kunna studera så många olika proteiner som möjligt i t.ex. ett tumörvävnadsprov eller blodprov kommer bli mycket central inom den framtida medicinska diagnostiken. Sådana proteinanalyser kommer kunna ge en betydligt mer komplett bild av den specifike individens tillstånd vilket kommer möjligöra en mer individanpassad skräddarsydd behandling dvs. val av korrekt behandling för att uppnå maximal effekt och minimalt lidande.

I min avhandling som bygger på fyra vetenskapliga arbeten beskrivs grunden, utvecklandet och de första stegen för användandet av en ny metodik för proteomanalyser. Målet var att utveckla ett analysverktyg där stora delar av

proteomet skulle kunna analyseras med hjälp av s.k. affinitets-proteomik och masspektrometri. Detta i hopp om att möjliggöra enklare, snabbare, känsligare och förhoppningsvis i slutändan heltäckande analyser för att t.ex. kunna identifiera nya sjukdomsassocierade proteiner. Den utvecklade metodiken kallar vi Global Proteome Survey (GPS) och i de presenterade manuskripten visas hur metoden fungerar i sin helhet för analys av biologiska prover. Många av de tidigare discovery-analysverktygen har ofta varit begränsade till att studera proteiner som återfinns i högre koncentrationer i en cell vilket medför att man missar möjligheten att mäta betydligt ovanligare men minst lika viktiga proteiner (nålen i höstack fenomenet). För att kunna mäta låguttryckta proteiner har man ofta fått använda sig av antikroppar (bindar-molekyler) som kan fånga de aktuella proteinerna. Men detta har i sin tur gjort att man varit tvungen att bestämma sig (val av antikroppar) redan innan experimenten gjorts exakt för vilka protein man ska analysera vilket givetvis begränsat möjligheterna för att upptäcka helt nya okända sjukdomsassocierade proteiner. GPS metodiken löser detta och öppnar upp för globala analyser genom att använda sig av en unik variant av antikroppar som kan fånga korta terminala aminosyrasekvenser (motif) hos olika peptider. Dessa unika bindarmolekyler har vi döpt till CIMS-bindare (eng. Context Independent Motif Specific). Proteiner kan nämligen klippas upp i mindre fragment s.k. peptider genom att exponeras för specifika enzymer (t.ex. trypsin) som klipper alla protein i mindre beståndsdelar (peptider) vid specifika positioner. Dessa peptider kan vi sedan fånga med våra CIMS-bindare och därefter detektera och kvantifiera med hjälp av en masspektrometer dvs. ett instrument som kan mäta molekylers massa med extremt hög noggrannhet. Man börjar mäta massan för hela peptiden (intakt) och därefter även olika fragment som uppkommer från peptiden i instrumentet. Genom att tolka alla de uppkomna fragmentens massor kan man bestämma hela peptidens aminosyrasammansättning (sekvensordning) och med lite tur och i kombination med avancerade databassökningar spåra från vilket protein som peptiden ursprungligen kom ifrån. Databaserna man använder sig av för att söka i har kunnat byggas upp tack vare att det mänskliga genomet blev känt vilket visar på den stora betydelse sekvenseringen av genomet har och kommer fortsätta ha många år framöver.

För att påvisa GPS styrka i form av enkelhet, robusthet och känslighet studerades initialt protein extrakt från flera olika organismer (muslever, tarmvävnad från människa, samt jästceller) och vi kan nu med säkerhet säga att metodiken kommer att kunna fungera på i princip vilken typ av organism som helst (levande på vår planet). I början lades fokus på att studera ett enkelt system dvs. jäst som odlades i närvaro av glukos eller etanol. Tydliga skillnader för många nyckel-proteiner detekterades. Dessutom påvisades metodens analyskänslighet genom framgångsrik

detektion av proteiner uppskattade till att finnas i mindre än 50 kopior per jästcell. I hopp om att kunna förfina och förbättra bindarmolekylerna och utveckla GPS metodiken ytterligare gjordes sedan studier på flera bindarmolekyler samt strukturella datamodellerings-studier av både bindarmolekyler och de fångade peptiderna. Resultatet av dessa modelleringsstudier gjorde att vi till viss del kan förklara varför vissa peptider fångades av vissa CIMS-molekyler och denna information kommer nu kunna användas för att generera nästa generation av nya CIMS-bindarmolekyler.

Slutligen så applicerades den utvecklade GPS metodiken för analys på bröstcancer tumörer. Bröstcancer är den absolut vanligaste cancertypen hos kvinnor och det är idag erkänt att det är en mycket heterogen och komplicerad cancer form. Desto tidigare diagnos ställs och val av korrekt terapi görs ökar drastiskt på chanserna för överlevnad. Tyvärr så är en del av de nuvarande traditionella metoderna otillräckliga och läkarna behöver få tillgång till nya kraftigare analysverktyg som kan vägleda dem bättre i val av diagnos och behandling. Genom att använda GPS teknologin på över 50 olika patient-tumör prover (fördelade i tre olika grupper (grader) av tumörer) kunde vi jämföra mer än 1300 olika proteiner mellan samtliga prover. Unika protein signaturer (differentiellt uttryckta proteiner) mellan den lindrigaste (grad I) samt svåraste (grad III) av tumör-typerna kunde tydligt påvisas. Proteiner med både kända och delvis okända funktioner identifierades och mycket intressanta mönster observerades och sammantaget gav detta en tydlig indikation på ett antal proteiner viktiga för tumörernas progression. Det finns en stor potential för att flera av dessa proteiner kan ev. kunna komma användas som biomarkörer och med fördjupade studier kan vissa t.o.m. kanske fungera som attraktiva målmolekyler för framtida läkemedel.

Sammanfattningsvis kan man säga att grunden och de initiala stegen för GPS metodiken och dess breda användningsområden och styrka har påvisats i denna avhandling. Utvecklandet av tekniken kommer fortsätta genom genererande av fler bindarmolekyler samt förfinade masspektrometri instrument och data analyser. Förhoppningsvis kommer denna metodik utgöra en av flera hörnstenar för både breda och fokuserade proteinanalyser under många år framåt.

Acknowledgements

First of all, a big hurrah, to all of you who made the great achievement of actually reading all previous 56 pages - *before* turning to this page. Congratulations! I would now like to THANK all fantastic people that, in different ways, contributed to this thesis. In particular, I would like to express my gratitude to the following persons:

Christer, min handledare, TACK för ditt engagemang, kloka råd samt positiva anda under denna resa. Din dörr har varit öppen för att diskutera stort som smått, även tack för att du trodde på jästexperimenten samt för all hjälp under slutspurten. Det har varit en spännande tid där vi båda onekligen försökt pressa oss till det yttersta enligt bästa Tour de France maner.

Carl, min biträdande handledare, TACK för alla kloka råd samt att du gett projektet stora möjligheter och fria ramar. Ett extra tack för att du vid kritiska tillfällen även satt tryck på en viss engelsman så att vi fick tillgång till de lite mer lyxiga instrumenten. Dessutom har du skapat en otroligt trevlig avdelning med tillgång till både ”cutting edge” kollegor och vetenskaplig utrustning.

Peter James, all your input and ability to take your time and discuss various problems have been invaluable. I have learnt incredible much. THANKS! I have also fully enjoyed your sense of humor and hilarious stories. Oh yes, also big thanks for letting me use the Orbi.

Mats Ohlin, tack för värdefull input, din oändliga kunskap samt alla spännande frågor under alla år.

Petter, TACK för all hjälp med bröst cancer projektet, för ditt alltid lika glada sinne samt delandet av frustrationen då Orbin inte riktigt ville som vi.

Mats Mågård och Karin Hansson, TACK för att ni tagit hand om de kära massspek-instrumenten så otroligt bra. Er kunskap, assistans och dedikering har varit helt enastående. Liselotte, tack för hjälpen med Maldi instrumenten.

Fredrik L, TACK för alla givande diskussioner och all hjälp med Proteios SE samt för ditt alltid lika glada och öppna sinne.

Ett stort tack till ALLA medlemmar i protein chip gruppen, gamla som nya, för all input och diskussioner under dessa åren. Ett extra tack till mina två "roomies" Frida och Jacob för trevliga stunder, diskussioner och input. Anders, extra tack för allt arbete med script och Peter E för värdefulla diskussioner under första åren. Även extra tack till Anna och Helena för att ni "hittat rätt" och anslutit till "GPS-spåret".

Tack alla externa sammarbetspartners. Speciellt, Fredrik och Mikael på BioInvent för att ni initierade och startade upp detta projekt. Stefan, tack för ditt arbete med modelleringarna. Alla på onkologen, tack för att ni lät GPS testas på era kliniska prover. Simon E och Thomas L, tack för intressant sammarbete.

Stefanie Gaertner, my phenomenal master thesis student, thanks for all your fantastic work. Rene Gniouth, thank you for exciting mass spec experiments. It was a real pleasure to supervise both of you.

Jag vill även tacka ALLA medlemmar på protein-teknologi sidan, gamla som nya. Extra tack till Sofia, Johan M, Linn, Paulo o Kristoffer W för ert stora intresse och trevliga diskussioner under åren. Marianne, tack för att du lär oss andra statistik. An extra thanks to Roberto, Giorgio and Cinzia for all late nights in the lab.

Tack ALLA på avdelningen, nya som före detta, för alla trevliga stunder. Extra TACK till Ann-Charlott, Marianne o Tommie för att ni har allt under kontroll. Ni är SUPER! Ett särkilt tack till alla doktorand kollegor under dessa år för alla roliga events både på och utanför jobbet.

Till ALLA vänner utanför avdelningen som alltid visat intresse.

Ett speciellt tack till Ola, Olof och Philippe för alla groovy-jam-sessioner under åren. Även hela salsa gänget och speciellt blåssektionen för alla trevliga rep. Familjerna Sjöstrands och Nilssons för alla fantastiska middagar. Eva o Lars-Erik, Clary o Sune, Eva I och Nils-Olof, för att ni alltid finns där och för ert stora intresse. Familjerna Sandin o Svantesson för alla trevliga stunder under dessa år.

Maria, världens bästa syster! Tack till dig, Fredrik och Alfonz för alla trevliga stunder och för att ni alltid finns till och stöttar.

Mina fantastiska föräldrar. TACK för att ni alltid finns till hands och ställer upp. Jag älskar er jättemycket.

References

Ackermann, B.L., and Berna, M.J. (2007). Coupling immunoaffinity techniques with MS for quantitative analysis of low-abundance protein biomarkers. Expert Rev Proteomics *4*, 175-186.

Addona, T.A., Abbatiello, S.E., Schilling, B., Skates, S.J., Mani, D.R., Bunk, D.M., Spiegelman, C.H., Zimmerman, L.J., Ham, A.J., Keshishian, H., *et al.* (2009). Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat Biotechnol *27*, 633-641.

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198-207.

Agaton, C., Galli, J., Hoiden Guthenberg, I., Janzon, L., Hansson, M., Asplund, A., Brundell, E., Lindberg, S., Ruthberg, I., Wester, K., *et al.* (2003). Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. Mol Cell Proteomics *2*, 405-414.

Allison, M. (2010). The HER2 testing conundrum. Nat Biotechnol *28*, 117-119.

Anderson, N.L. (2010). The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. Clin Chem *56*, 177-185.

Anderson, N.L., and Anderson, N.G. (2002). The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics *1*, 845-867.

Anderson, N.L., Anderson, N.G., Haines, L.R., Hardie, D.B., Olafson, R.W., and Pearson, T.W. (2004). Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). J Proteome Res *3*, 235-244.

Anderson, N.L., Anderson, N.G., Pearson, T.W., Borchers, C.H., Paulovich, A.G., Patterson, S.D., Gillette, M., Aebersold, R., and Carr, S.A. (2009a). A human proteome detection and quantitation project. Mol Cell Proteomics *8*, 883-886.

Anderson, N.L., Jackson, A., Smith, D., Hardie, D., Borchers, C., and Pearson, T.W. (2009b). SISCAPA peptide enrichment on magnetic beads using an in-line bead trap device. Mol Cell Proteomics *8*, 995-1005.

Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011). The quantitative proteome of a human cell line. Mol Syst Biol *7*, 549.

Bell, A.W., Deutsch, E.W., Au, C.E., Kearney, R.E., Beavis, R., Sechi, S., Nilsson, T., and Bergeron, J.J. (2009). A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. Nat Methods *6*, 423-430.

Bengtsson, S., Krogh, M., Szigyarto, C.A., Uhlen, M., Schedvins, K., Silfversward, C., Linder, S., Auer, G., Alaiya, A., and James, P. (2007). Large-scale proteomics analysis of human ovarian cancer for biomarkers. J Proteome Res *6*, 1440-1450.

Bergman, O., Hont, G., and Johansson, E. (2009). Cancer i siffror. Cancerfonden, Sa Stockholm, Socialstyrelsen.

Bird, R.E., Hardman, K.D., Jacobson, J.W., Johnson, S., Kaufman, B.M., Lee, S.M., Lee, T., Pope, S.H., Riordan, G.S., and Whitlow, M. (1988). Single-chain antigen-binding proteins. Science *242*, 423-426.

Boder, E.T., Midelfort, K.S., and Wittrup, K.D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proc Natl Acad Sci U S A *97*, 10701-10705.

Borrebaeck, C.A.K (2000). Antibodies in diagnostics - from immunoassays to protein chips. Immunology today *21*, 379-382.

Borrebaeck, C.A.K., and Ohlin, M. (2002). Antibody evolution beyond Nature. Nat Biotechnol *20*, 1189-1190.

Borrebaeck, C.A.K., and Wingren, C. (2007). High-throughput proteomics using antibody microarrays: an update. Expert review of molecular diagnostics *7*, 673-686.

Borrebaeck, C.A.K., and Wingren, C. (2009). Transferring proteomic discoveries into clinical practice. Expert Rev Proteomics *6*, 11-13.

Borrebaeck, C.A.K., and Wingren, C. (2011). Recombinant antibodies for the generation of antibody arrays. Methods Mol Biol *785*, 247-262.

Bouchal, P., Roumeliotis, T., Hrstka, R., Nenutil, R., Vojtesek, B., and Garbis, S.D. (2009). Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. Journal of proteome research *8*, 362-373.

Bourbeillon, J., Orchard, S., Benhar, I., Borrebaeck, C., de Daruvar, A., Dubel, S., Frank, R., Gibson, F., Gloriam, D., Haslam, N., *et al.* (2010). Minimum information about a protein affinity reagent (MIAPAR). Nat Biotechnol *28*, 650-653.

Bradbury, A.R., and Marks, J.D. (2004). Antibodies from phage antibody libraries. J Immunol Methods *290*, 29-49.

Bradbury, A.R., Sidhu, S., Dubel, S., and McCafferty, J. (2011). Beyond natural antibodies: the power of in vitro display technologies. Nat Biotechnol *29*, 245-254.

Carlsson, A., Wingren, C., Ingvarsson, J., Ellmark, P., Baldertorp, B., Ferno, M., Olsson, H., and Borrebaeck, C.A.K. (2008). Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays. Eur J Cancer *44*, 472-480.

Carlsson, A., Wingren, C., Kristensson, M., Rose, C., Ferno, M., Olsson, H., Jernstrom, H., Ek, S., Gustavsson, E., Ingvar, C., *et al.* (2011a). Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. Proc Natl Acad Sci U S A *108*, 14252-14257.

Carlsson, A., Wuttge, D.M., Ingvarsson, J., Bengtsson, A.A., Sturfelt, G., Borrebaeck, C.A.K, and Wingren, C. (2011b). Serum protein profiling of systemic lupus erythematosus and systemic sclerosis using recombinant antibody microarrays. Mol Cell Proteomics *10*, M110 005033.

Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K., and Nesvizhskii, A. (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*, 531-533.

Carr, S.A., and Anderson, L. (2008). Protein quantitation through targeted mass spectrometry: the way out of biomarker purgatory? Clin Chem *54*, 1749-1752.

Carter, P. (2001). Improving the efficacy of antibody-based cancer therapies. Nat Rev Cancer *1*, 118-129.

Cho, W.C. (2007). Contribution of oncoproteomics to cancer biomarker discovery. Molecular cancer *6*, 25.

Ciocca, D.R., and Elledge, R. (2000). Molecular markers for predicting response to tamoxifen in breast cancer patients. Endocrine *13*, 1-10.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M.F., Kellis, M., Lindblad-Toh, K., and Lander, E.S. (2007). Distinguishing protein-coding and noncoding genes in the human genome. Proc Natl Acad Sci U S A *104*, 19428-19433.

Colombo, P.E., Milanezi, F., Weigelt, B., and Reis-Filho, J.S. (2011). Microarrays in the 2010s: the contribution of microarray-based gene expression profiling to breast cancer classification, prognostication and prediction. Breast cancer research : BCR *13*, 212.

Colwill, K., Persson, H., Jarvik, N.E., Wyrzucki, A., Wojcik, J., Koide, A., Kossiakoff, A.A., Koide, S., Sidhu, S., Dyson, M.R., *et al.* (2011). A roadmap to generate renewable protein binders to the human proteome. Nat Methods *8*, 551-558.

Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F., and Sanchez, J.C. (2000). The dynamic range of protein expression: a challenge for proteomic research. Electrophoresis *21*, 1104-1115.

Costenoble, R., Picotti, P., Reiter, L., Stallmach, R., Heinemann, M., Sauer, U., and Aebersold, R. (2011). Comprehensive quantitative analysis of central carbon and amino-acid metabolism in Saccharomyces cerevisiae under multiple conditions by targeted proteomics. Mol Syst Biol *7*, 464.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Craig, R., and Beavis, R.C. (2004). TANDEM: matching proteins with tandem mass spectra. Bioinformatics *20*, 1466-1467.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome research *14*, 1188-1190.

de Godoy, L.M., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Frohlich, F., Walther, T.C., and Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature *455*, 1251-1254.

Deng, S.S., Xing, T.Y., Zhou, H.Y., Xiong, R.H., Lu, Y.G., Wen, B., Liu, S.Q., and Yang, H.J. (2006). Comparative proteome analysis of breast cancer and adjacent normal breast tissues in human. Genomics, proteomics & bioinformatics / Beijing Genomics Institute *4*, 165-172.

Desmedt, C., Haibe-Kains, B., Wirapati, P., Buyse, M., Larsimont, D., Bontempi, G., Delorenzi, M., Piccart, M., and Sotiriou, C. (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clinical cancer research : an official journal of the American Association for Cancer Research *14*, 5158-5165.

Deutsch, E.W., Lam, H., and Aebersold, R. (2008). PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. EMBO reports *9*, 429-434.

Dexlin-Mellby, L., Sandstrom, A., Antberg, L., Gunnarsson, J., Hansson, S.R., Borrebaeck, C.A.K, and Wingren, C. (2011). Design of recombinant antibody microarrays for membrane protein profiling of cell lysates and tissue extracts. Proteomics *11*, 1550-1554.

Dexlin, L., Ingvarsson, J., Frendeus, B., Borrebaeck, C.A.K, and Wingren, C. (2008). Design of recombinant antibody microarrays for cell surface membrane proteomics. J Proteome Res *7*, 319-327.

Dubel, S., Stoevesandt, O., Taussig, M.J., and Hust, M. (2010). Generating recombinant antibodies to the complete human proteome. Trends Biotechnol *28*, 333-339.

Edman, P. (1949). A method for the determination of amino acid sequence in peptides. Archives of biochemistry *22*, 475.

Ein-Dor, L., Kela, I., Getz, G., Givol, D., and Domany, E. (2005). Outcome signature genes in breast cancer: is there a unique set? Bioinformatics *21*, 171-178.

Ek, S., Andreasson, U., Hober, S., Kampf, C., Ponten, F., Uhlen, M., Merz, H., and Borrebaeck, C.A.K. (2006). From gene expression analysis to tissue microarrays: a rational approach to identify therapeutic and diagnostic targets in lymphoid malignancies. Mol Cell Proteomics *5*, 1072-1081.

Ekins, R.P., and Chu, F.W. (1991). Multianalyte microspot immunoassay--microanalytical "compact disk" of the future. Clin Chem *37*, 1955-1967.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods *4*, 207-214.

Ellington, A.D., and Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature *346*, 818-822.

Elston, C.W., and Ellis, I.O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology *19*, 403-410.

Eng, J.K., McCormack, A.L., and Yates, J.R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrum, 976-989.

Etzioni, R., Urban, N., Ramsey, S., McIntosh, M., Schwartz, S., Reid, B., Radich, J., Anderson, G., and Hartwell, L. (2003). The case for early detection. Nat Rev Cancer *3*, 243-252.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science *246*, 64-71.

Fredriksson, S., Dixon, W., Ji, H., Koong, A.C., Mindrinos, M., and Davis, R.W. (2007). Multiplexed protein detection by proximity ligation for cancer biomarker validation. Nat Methods *4*, 327-329.

Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S.M., Ostman, A., and Landegren, U. (2002). Protein detection using proximity-dependent DNA ligation assays. Nat Biotechnol *20*, 473-477.

Frierson, H.F., Jr., Wolber, R.A., Berean, K.W., Franquemont, D.W., Gaffey, M.J., Boyd, J.C., and Wilbur, D.C. (1995). Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. American journal of clinical pathology *103*, 195-198.

Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R., and Mann, M. (2010). Super-SILAC mix for quantitative proteomics of human tumor tissue. Nature methods *7*, 383-385.

Gevaert, K., Van Damme, J., Goethals, M., Thomas, G.R., Hoorelbeke, B., Demol, H., Martens, L., Puype, M., Staes, A., and Vandekerckhove, J. (2002). Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 Escherichia coli proteins. Mol Cell Proteomics *1*, 896-903.

Gevaert, K., Van Damme, P., Martens, L., and Vandekerckhove, J. (2005). Diagonal reverse-phase chromatography applications in peptide-centric proteomics: ahead of catalogue-omics? Anal Biochem *345*, 18-29.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., *et al.* (1996). Life with 6000 genes. Science *274*, 546, 563-547.

Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E.N., Carter, J., Dalby, A.B., Eaton, B.E., Fitzwater, T., *et al.* (2010). Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One *5*, e15004.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., *et al.* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531-537.

Gruhler, A., Olsen, J.V., Mohammed, S., Mortensen, P., Faergeman, N.J., Mann, M., and Jensen, O.N. (2005). Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol Cell Proteomics *4*, 310-327.

Gry, M., Rimini, R., Stromberg, S., Asplund, A., Ponten, F., Uhlen, M., and Nilsson, P. (2009). Correlations between RNA and protein expression profiles in 23 human cell lines. BMC genomics *10*, 365.

Gupta, N., and Pevzner, P.A. (2009). False discovery rates of protein identifications: a strike against the two-peptide rule. J Proteome Res *8*, 4173-4181.

Gustavsson, E., Ek, S., Steen, J., Kristensson, M., Algenas, C., Uhlen, M., Wingren, C., Ottosson, J., Hober, S., and Borrebaeck, C.A.K. (2011). Surrogate antigens as targets for proteome-wide binder selection. N Biotechnol *28*, 302-311.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. Molecular and cellular biology *19*, 1720-1730.

Haab, B.B. (2005). Antibody arrays in cancer research. Mol Cell Proteomics *4*, 377-383.

Haab, B.B., Dunham, M.J., and Brown, P.O. (2001). Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol *2*, RESEARCH0004.

Hakkinen, J., Vincic, G., Mansson, O., Warell, K., and Levander, F. (2009). The proteios software environment: an extensible multiuser platform for management and analysis of proteomics data. J Proteome Res *8*, 3037-3043.

Hallborn, J., and Carlsson, R. (2002). Automated screening procedure for high-throughput generation of antibody fragments. BioTechniques *Suppl*, 30-37.

Hamdani, N., and van der Velden, J. (2009). Lack of specificity of antibodies directed against human beta-adrenergic receptors. Naunyn Schmiedebergs Arch Pharmacol *379*, 403-407.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Hanash, S. (2003). Disease proteomics. Nature *422*, 226-232.

Hanash, S., and Taguchi, A. (2010). The grand challenge to decipher the cancer proteome. Nat Rev Cancer *10*, 652-660.

Hanash, S.M. (2000). Biomedical applications of two-dimensional electrophoresis using immobilized pH gradients: current status. Electrophoresis *21*, 1202-1209.

Hartmann, M., Roeraade, J., Stoll, D., Templin, M.F., and Joos, T.O. (2009). Protein microarrays for diagnostic assays. Anal Bioanal Chem *393*, 1407-1416.

Hede, K. (2008). Breast cancer testing scandal shines spotlight on black box of clinical laboratory testing. J Natl Cancer Inst *100*, 836-837, 844.

Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C., and Watanabe, C. (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*, 5011-5015.

Hill, J.J., Tremblay, T.L., Pen, A., Li, J., Robotham, A.C., Lenferink, A.E., Wang, E., O'Connor-McCourt, M., and Kelly, J.F. (2011). Identification of vascular breast tumor markers by laser capture microdissection and label-free LC-MS. Journal of proteome research *10*, 2479-2493.

Hnatyszyn, H.J., Liu, M., Hilger, A., Herbert, L., Gomez-Fernandez, C.R., Jorda, M., Thomas, D., Rae, J.M., El-Ashry, D., and Lippman, M.E. (2010). Correlation of GREB1 mRNA with protein expression in breast cancer: validation of a novel GREB1 monoclonal antibody. Breast cancer research and treatment *122*, 371-380.

Hoeppe, S., Schreiber, T.D., Planatscher, H., Zell, A., Templin, M.F., Stoll, D., Joos, T.O., and Poetz, O. (2011). Targeting peptide termini, a novel immunoaffinity approach to reduce complexity in mass spectrometric protein identification. Mol Cell Proteomics *10*, M110 002857.

Hondermarck, H. (2003). Breast cancer: when proteomics challenges biological complexity. Molecular & cellular proteomics : MCP *2*, 281-291.

Hondermarck, H., Tastet, C., El Yazidi-Belkoura, I., Toillon, R.A., and Le Bourhis, X. (2008). Proteomics of breast cancer: the quest for markers and therapeutic targets. Journal of proteome research *7*, 1403-1411.

Huang, W., Reinholz, M., Weidler, J., Yolanda, L., Paquet, A., Whitcomb, J., Lingle, W., Jenkins, R.B., Chen, B., Larson, J.S., *et al.* (2010). Comparison of central HER2 testing with quantitative total HER2 expression and HER2 homodimer measurements using a novel proximity-based assay. American journal of clinical pathology *134*, 303-311.

Hubner, N.C., Ren, S., and Mann, M. (2008). Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. Proteomics *8*, 4862-4872.

James, P., Quadroni, M., Carafoli, E., and Gonnet, G. (1993). Protein identification by mass profile fingerprinting. Biochem Biophys Res Commun *195*, 58-64.

Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. CA: a cancer journal for clinicians *61*, 69-90.

Jensen, B.C., Swigart, P.M., and Simpson, P.C. (2009). Ten commercial antibodies for alpha-1-adrenergic receptor subtypes are nonspecific. Naunyn Schmiedebergs Arch Pharmacol *379*, 409-412.

Johnson, N., Li, Y.C., Walton, Z.E., Cheng, K.A., Li, D., Rodig, S.J., Moreau, L.A., Unitt, C., Bronson, R.T., Thomas, H.D., *et al.* (2011). Compromised CDK1 activity sensitizes BRCA-proficient cancers to PARP inhibition. Nature medicine *17*, 875-882.

Kang, S., Kim, M.J., An, H., Kim, B.G., Choi, Y.P., Kang, K.S., Gao, M.Q., Park, H., Na, H.J., Kim, H.K., *et al.* (2010). Proteomic molecular portrait of interface zone in breast cancer. Journal of proteome research *9*, 5638-5645.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem *60*, 2299-2301.

Kijanka, G., Ipcho, S., Baars, S., Chen, H., Hadley, K., Beveridge, A., Gould, E., and Murphy, D. (2009). Rapid characterization of binding specificity and cross-reactivity of antibodies using recombinant human protein arrays. J Immunol Methods *340*, 132-137.

Kim, D.H., Bae, J., Lee, J.W., Kim, S.Y., Kim, Y.H., Bae, J.Y., Yi, J.K., Yu, M.H., Noh, D.Y., and Lee, C. (2009). Proteomic analysis of breast cancer tissue reveals upregulation of actin-remodeling proteins and its relevance to cancer invasiveness. Proteomics Clin Appl *3*, 30-40.

Kingsmore, S.F. (2006). Multiplexed protein measurement: technologies and applications of protein and antibody arrays. Nature reviews Drug discovery *5*, 310-320.

Knezevic, V., Leethanakul, C., Bichsel, V.E., Worth, J.M., Prabhu, V.V., Gutkind, J.S., Liotta, L.A., Munson, P.J., Petricoin, E.F., 3rd, and Krizman, D.B. (2001). Proteomic profiling of the cancer microenvironment by antibody arrays. Proteomics *1*, 1271-1278.

Kohler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. Nature *256*, 495-497.

Kolkman, A., Olsthoorn, M.M., Heeremans, C.E., Heck, A.J., and Slijper, M. (2005). Comparative proteome analysis of Saccharomyces cerevisiae grown in chemostat cultures limited for glucose or ethanol. Mol Cell Proteomics *4*, 1-11.

Kraemer, S., Vaught, J.D., Bock, C., Gold, L., Katilius, E., Keeney, T.R., Kim, N., Saccomano, N.A., Wilcox, S.K., Zichi, D., *et al.* (2011). From SOMAmer-based

biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. PLoS One *6*, e26332.

Krijgsveld, J., Gauci, S., Dormeyer, W., and Heck, A.J. (2006). In-gel isoelectric focusing of peptides as a tool for improved protein identification. J Proteome Res *5*, 1721-1730.

Kristensson, M., Olsson, K., Carlson, J., Wullt, B., Sturfelt, G., Borrebaeck, C.A.K, and Wingren, C. (2012). Design of Recombinant Antibody Microarrays for Urinary Proteomics. Proteomics-Clinical Applications *(In press)*.

Kruger, M., Moser, M., Ussar, S., Thievessen, I., Luber, C.A., Forner, F., Schmidt, S., Zanivan, S., Fassler, R., and Mann, M. (2008). SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell *134*, 353-364.

Kuzyk, M.A., Smith, D., Yang, J., Cross, T.J., Jackson, A.M., Hardie, D.B., Anderson, N.L., and Borchers, C.H. (2009). Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Mol Cell Proteomics *8*, 1860-1877.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lange, V., Malmstrom, J.A., Didion, J., King, N.L., Johansson, B.P., Schafer, J., Rameseder, J., Wong, C.H., Deutsch, E.W., Brusniak, M.Y., *et al.* (2008a). Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. Mol Cell Proteomics *7*, 1489-1500.

Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008b). Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol *4*, 222.

Lee, M.V., Topper, S.E., Hubler, S.L., Hose, J., Wenger, C.D., Coon, J.J., and Gasch, A.P. (2011). A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. Mol Syst Biol *7*, 514.

Leong, A.S., and Zhuang, Z. (2011). The changing role of pathology in breast cancer diagnosis and treatment. Pathobiology : journal of immunopathology, molecular and cellular biology *78*, 99-114.

Liu, T., Qian, W.J., Mottaz, H.M., Gritsenko, M.A., Norbeck, A.D., Moore, R.J., Purvine, S.O., Camp, D.G., 2nd, and Smith, R.D. (2006). Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. Mol Cell Proteomics *5*, 2167-2174.

Lossner, C., Warnken, U., Pscherer, A., and Schnolzer, M. (2011). Preventing arginine-to-proline conversion in a cell-line-independent manner during cell cultivation under stable isotope labeling by amino acids in cell culture (SILAC) conditions. Anal Biochem *412*, 123-125.

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E.M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol *25*, 117-124.

Lundberg, M., Thorsen, S.B., Assarsson, E., Villablanca, A., Tran, B., Gee, N., Knowles, M., Nielsen, B.S., Gonzalez Couto, E., Martin, R., *et al.* (2011). Multiplexed homogeneous proximity ligation assays for high-throughput protein biomarker research in serological material. Mol Cell Proteomics *10*, M110 004978.

MacBeath, G. (2002). Protein microarrays and proteomics. Nat Genet *32 Suppl*, 526-532.

Makarov, A. (2000). Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal Chem *72*, 1156-1162.

Mallick, P., and Kuster, B. (2010). Proteomics: a pragmatic perspective. Nat Biotechnol *28*, 695-709.

Malmstrom, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E.W., and Aebersold, R. (2009). Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. Nature *460*, 762-765.

Mann, M. (2008). Can proteomics retire the western blot? J Proteome Res *7*, 3065.

Mann, M., Hojrup, P., and Roepstorff, P. (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biological mass spectrometry *22*, 338-345.

Mann, M., and Wilm, M. (1994). Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem *66*, 4390-4399.

Marroni, F., Aretini, P., D'Andrea, E., Caligo, M.A., Cortesi, L., Viel, A., Ricevuto, E., Montagna, M., Cipollini, G., Federico, M., *et al.* (2004). Penetrances of breast and ovarian cancer in a large series of families tested for BRCA1/2 mutations. European journal of human genetics : EJHG *12*, 899-906.

Maupetit, J., Derreumaux, P., and Tuffery, P. (2009). PEP-FOLD: an online resource for de novo peptide structure prediction. Nucleic Acids Res *37*, W498-503.

McCafferty, J., Griffiths, A.D., Winter, G., and Chiswell, D.J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. Nature *348*, 552-554.

Medzihradszky, K.F., Campbell, J.M., Baldwin, M.A., Falick, A.M., Juhasz, P., Vestal, M.L., and Burlingame, A.L. (2000). The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. Anal Chem *72*, 552-558.

Michaud, G.A., Salcius, M., Zhou, F., Bangham, R., Bonin, J., Guo, H., Snyder, M., Predki, P.F., and Schweitzer, B.I. (2003). Analyzing antibody specificity with whole proteome microarrays. Nat Biotechnol *21*, 1509-1512.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., *et al.* (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science *266*, 66-71.

Milstein, C. (2000). With the benefit of hindsight. Immunology today *21*, 359-364.

Mischak, H., Apweiler, R., Banks, R.E., Conaway, M., Coon, J., Dominiczak, A., Ehrich, J.H., Fliser, D., Girolami, M., Hermjakob, H., *et al.* (2007). Clinical proteomics: A need to define the field and to begin to set adequate standards. Proteomics Clin Appl *1*, 148-156.

Nagaraj, N., Kulak, N.A., Cox, J., Neuhaus, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2011a). Systems-wide perturbation analysis with near complete coverage of the yeast proteome by single-shot UHPLC runs on a bench-top Orbitrap. Mol Cell Proteomics.

Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., and Mann, M. (2011b). Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol *7*, 548.

Narod, S.A., and Foulkes, W.D. (2004). BRCA1 and BRCA2: 1994 and beyond. Nat Rev Cancer *4*, 665-676.

Nilsson, P., Paavilainen, L., Larsson, K., Odling, J., Sundberg, M., Andersson, A.C., Kampf, C., Persson, A., Al-Khalili Szigyarto, C., Ottosson, J., *et al.* (2005). Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. Proteomics *5*, 4327-4337.

Nilsson, T., Mann, M., Aebersold, R., Yates, J.R., 3rd, Bairoch, A., and Bergeron, J.J. (2010). Mass spectrometry in high-throughput proteomics: ready for the big time. Nat Methods *7*, 681-685.

Nord, K., Gunneriusson, E., Ringdahl, J., Stahl, S., Uhlen, M., and Nygren, P.A. (1997). Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. Nat Biotechnol *15*, 772-777.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. J Biol Chem *250*, 4007-4021.

Olsen, J.V., Ong, S.E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics *3*, 608-614.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics *1*, 376-386.

Pandey, A., and Mann, M. (2000). Proteomics to study genes and genomes. Nature *405*, 837-846.

Pappin, D.J., Hojrup, P., and Bleasby, A.J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. Current biology : CB *3*, 327-332.

Patterson, S.D., and Aebersold, R.H. (2003). Proteomics: the first decade and beyond. Nat Genet *33 Suppl*, 311-323.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551-3567.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., *et al.* (2000). Molecular portraits of human breast tumours. Nature *406*, 747-752.

Persson, J., Backstrom, M., Johansson, H., Jirstrom, K., Hansson, G.C., and Ohlin, M. (2009). Molecular evolution of specific human antibody against MUC1 mucin results in improved recognition of the antigen on tumor cells. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine *30*, 221-231.

Petrak, J., Ivanek, R., Toman, O., Cmejla, R., Cmejlova, J., Vyoral, D., Zivny, J., and Vulpe, C.D. (2008). Deja vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins. Proteomics *8*, 1744-1749.

Phelan, M.L., and Nock, S. (2003). Generation of bioreagents for protein chips. Proteomics *3*, 2123-2134.

Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B., and Aebersold, R. (2009). Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. Cell *138*, 795-806.

Planatscher, H., Supper, J., Poetz, O., Stoll, D., Joos, T., Templin, M.F., and Zell, A. (2010). Optimal selection of epitopes for TXP-immunoaffinity mass spectrometry. Algorithms Mol Biol *5*, 28.

Poetz, O., Hoeppe, S., Templin, M.F., Stoll, D., and Joos, T.O. (2009). Proteome wide screening using peptide affinity capture. Proteomics *9*, 1518-1523.

Polanski, M., and Anderson, N.L. (2006). A list of candidate cancer biomarkers for targeted proteomics. Biomark Insights *1*, 1-48.

Ponten, F., Jirstrom, K., and Uhlen, M. (2008). The Human Protein Atlas--a tool for pathology. The Journal of pathology *216*, 387-393.

Ponten, F., Schwenk, J.M., Asplund, A., and Edqvist, P.H. (2011). The Human Protein Atlas as a proteomic resource for biomarker discovery. Journal of internal medicine *270*, 428-446.

Pope, M.E., Soste, M.V., Eyford, B.A., Anderson, N.L., and Pearson, T.W. (2009). Anti-peptide antibody screening: selection of high affinity monoclonal reagents by a refined surface plasmon resonance technique. J Immunol Methods *341*, 86-96.

Rae, J.M., Johnson, M.D., Scheys, J.O., Cordero, K.E., Larios, J.M., and Lippman, M.E. (2005). GREB 1 is a critical regulator of hormone dependent breast cancer growth. Breast cancer research and treatment *92*, 141-149.

Ramachandran, N., Srivastava, S., and Labaer, J. (2008). Applications of protein microarrays for biomarker discovery. Proteomics Clin Appl *2*, 1444-1459.

Razavi, M., Pope, M.E., Soste, M.V., Eyford, B.A., Jackson, A.M., Anderson, N.L., and Pearson, T.W. (2011). MALDI immunoscreening (MiSCREEN): a method for selection of anti-peptide monoclonal antibodies for use in immunoproteomics. J Immunol Methods *364*, 50-64.

Renberg, B., Nordin, J., Merca, A., Uhlen, M., Feldwisch, J., Nygren, P.A., and Karlstrom, A.E. (2007). Affibody molecules in protein capture microarrays: evaluation of multidomain ligands and different detection formats. J Proteome Res *6*, 171-179.

Righetti, P.G., Boschetti, E., Lomas, L., and Citterio, A. (2006). Protein Equalizer Technology : the quest for a "democratic proteome". Proteomics *6*, 3980-3992.

Ringner, M., Fredlund, E., Hakkinen, J., Borg, A., and Staaf, J. (2011). GOBO: gene expression-based outcome for breast cancer online. PLoS One *6*, e17911.

Robbins, P., Pinder, S., de Klerk, N., Dawkins, H., Harvey, J., Sterrett, G., Ellis, I., and Elston, C. (1995). Histological grading of breast carcinomas: a study of interobserver agreement. Human pathology *26*, 873-879.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., *et al.* (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics *3*, 1154-1169.

Roux-Dalvai, F., Gonzalez de Peredo, A., Simo, C., Guerrier, L., Bouyssie, D., Zanella, A., Citterio, A., Burlet-Schiltz, O., Boschetti, E., Righetti, P.G., *et al.* (2008). Extensive analysis of the cytoplasmic proteome of human erythrocytes using the peptide ligand library technology and advanced mass spectrometry. Mol Cell Proteomics *7*, 2254-2269.

Roychowdhury, S., Iyer, M.K., Robinson, D.R., Lonigro, R.J., Wu, Y.M., Cao, X., Kalyana-Sundaram, S., Sam, L., Balbin, O.A., Quist, M.J., *et al.* (2011). Personalized oncology through integrative high-throughput sequencing: a pilot study. Science translational medicine *3*, 111ra121.

Saerens, D., Ghassabeh, G.H., and Muyldermans, S. (2008). Antibody technology in proteomics. Briefings in functional genomics & proteomics *7*, 275-282.

Sanchez-Carbayo, M., Socci, N.D., Lozano, J.J., Haab, B.B., and Cordon-Cardo, C. (2006). Profiling bladder cancer using targeted antibody arrays. The American journal of pathology *168*, 93-103.

Sandin, M., Krogh, M., Hansson, K., and Levander, F. (2011). Generic workflow for quality assessment of quantitative label-free LC-MS analysis. Proteomics *11*, 1114-1124.

Santoni, V., Molloy, M., and Rabilloud, T. (2000). Membrane proteins and proteomics: un amour impossible? Electrophoresis *21*, 1054-1070.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.

Schwenk, J.M., Gry, M., Rimini, R., Uhlen, M., and Nilsson, P. (2008). Antibody suspension bead arrays within serum proteomics. J Proteome Res *7*, 3168-3179.

Schwenk, J.M., Lindberg, J., Sundberg, M., Uhlen, M., and Nilsson, P. (2007). Determination of binding specificities in highly multiplexed bead-based assays for antibody proteomics. Mol Cell Proteomics *6*, 125-132.

Scigelova, M., and Makarov, A. (2006). Orbitrap mass analyzer--overview and applications in proteomics. Proteomics *6 Suppl 2*, 16-21.

Slamon, D.J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., *et al.* (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. The New England journal of medicine *344*, 783-792.

Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science *228*, 1315-1317.

Soderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N., Alexeiva, V., Aberg, A.M., Nilsson, A., Jansson, B., Ohlin, M., Wingren, C., *et al.* (2000). Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. Nat Biotechnol *18*, 852-856.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A *98*, 10869-10874.

Steinhauer, C., Wingren, C., Hager, A.C., and Borrebaeck, C.A. (2002). Single framework recombinant antibody fragments designed for protein chip applications. BioTechniques *Suppl*, 38-45.

Stella, R., Biancotto, G., Krogh, M., Angeletti, R., Pozza, G., Sorgato, M.C., James, P., and Andrighetto, I. (2011). Protein expression changes in skeletal muscle in response to growth promoter abuse in beef cattle. J Proteome Res *10*, 2744-2757.

Stoevesandt, O., and Taussig, M.J. (2007). Affinity reagent resources for human proteome detection: initiatives and perspectives. Proteomics *7*, 2738-2750.

Strande, V., Canelle, L., Tastet, C., Burlet-Schiltz, O., Monsarrat, B., and Hondermarck, H. (2009). The proteome of the human breast cancer cell line MDA-MB-231: Analysis by LTQ-Orbitrap mass spectrometry. Proteomics Clin Appl *3*, 41-50.

Sury, M.D., Chen, J.X., and Selbach, M. (2010). The SILAC fly allows for accurate protein quantification in vivo. Mol Cell Proteomics *9*, 2173-2183.

Tabb, D.L., Vega-Montoto, L., Rudnick, P.A., Variyath, A.M., Ham, A.J., Bunk, D.M., Kilpatrick, L.E., Billheimer, D.D., Blackman, R.K., Cardasis, H.L., *et al.* (2010). Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. J Proteome Res *9*, 761-776.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. (1988). Protein and polymer analyses up to *m/z* 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom *2*, 151-153.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., *et al.* (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods *6*, 377-382.

Taylor, C.F., Paton, N.W., Lilley, K.S., Binz, P.A., Julian, R.K., Jr., Jones, A.R., Zhu, W., Apweiler, R., Aebersold, R., Deutsch, E.W., *et al.* (2007). The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol *25*, 887-893.

Templin, M.F., Stoll, D., Bachmann, J., and Joos, T.O. (2004). Protein microarrays and multiplexed sandwich immunoassays: what beats the beads? Combinatorial chemistry & high throughput screening *7*, 223-229.

Tu, C., Rudnick, P.A., Martinez, M.Y., Cheek, K.L., Stein, S.E., Slebos, R.J., and Liebler, D.C. (2010). Depletion of abundant plasma proteins and limitations of plasma proteomics. J Proteome Res *9*, 4982-4991.

Tuttle, T.M., Abbott, A., Arrington, A., and Rueth, N. (2010). The increasing use of prophylactic mastectomy in the prevention of breast cancer. Current oncology reports *12*, 16-21.

Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, C.A., Amini, B., Andersen, E., Andersson, A.C., Angelidou, P., Asplund, A., Asplund, C., *et al.* (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics *4*, 1920-1932.

Uhlen, M., and Ponten, F. (2005). Antibody-based proteomics for human tissue profiling. Mol Cell Proteomics *4*, 384-393.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., *et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530-536.

Washburn, M.P., Wolters, D., and Yates, J.R., 3rd (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol *19*, 242-247.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Whiteaker, J.R., Lin, C., Kennedy, J., Hou, L., Trute, M., Sokal, I., Yan, P., Schoenherr, R.M., Zhao, L., Voytovich, U.J., *et al.* (2011a). A targeted proteomics-based pipeline for verification of biomarkers in plasma. Nat Biotechnol *29*, 625-634.

Whiteaker, J.R., Zhao, L., Abbatiello, S.E., Burgess, M., Kuhn, E., Lin, C., Pope, M.E., Razavi, M., Anderson, N.L., Pearson, T.W., *et al.* (2011b). Evaluation of large scale

quantitative proteomic assay development using peptide affinity-based mass spectrometry. Mol Cell Proteomics *10*, M110 005645.

Whiteaker, J.R., Zhao, L., Anderson, L., and Paulovich, A.G. (2010). An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. Mol Cell Proteomics *9*, 184-196.

Whiteaker, J.R., Zhao, L., Lin, C., Yan, P., Wang, P., and Paulovich, A.G. (2011c). Sequential multiplexed analyte quantification using peptide immunoaffinity enrichment coupled to mass spectrometry. Mol Cell Proteomics. M111.015347

Wilkins, M.R., Appel, R.D., Van Eyk, J.E., Chung, M.C., Gorg, A., Hecker, M., Huber, L.A., Langen, H., Link, A.J., Paik, Y.K., *et al.* (2006). Guidelines for the next 10 years of proteomics. Proteomics *6*, 4-8.

Wilkins, M.R., Gasteiger, E., Sanchez, J.C., Bairoch, A., and Hochstrasser, D.F. (1998). Two-dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number. Electrophoresis *19*, 1501-1505.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., *et al.* (1996). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology (N Y) *14*, 61-65.

Wingren, C., and Borrebaeck, C.A.K. (2004). High-throughput proteomics using antibody microarrays. Expert Rev Proteomics *1*, 355-364.

Wingren, C., and Borrebaeck, C.A.K. (2006). Antibody microarrays: current status and key technological advances. Omics : a journal of integrative biology *10*, 411-427.

Wingren, C., Ingvarsson, J., Dexlin, L., Szul, D., and Borrebaeck, C.A.K. (2007). Design of recombinant antibody microarrays for complex proteome analysis: choice of sample labeling-tag and solid support. Proteomics *7*, 3055-3065.

Wingren, C., James, P., and Borrebaeck, C.A.K. (2009). Strategy for surveying the proteome using affinity proteomics and mass spectrometry. Proteomics *9*, 1511-1517.

Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., Schutz, F., *et al.* (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast cancer research : BCR *10*, R65.

Wolters, D.A., Washburn, M.P., and Yates, J.R., 3rd (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem *73*, 5683-5690.

Wooster, R., Neuhausen, S.L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D., *et al.* (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science *265*, 2088-2090.

Yang, W.S., Moon, H.G., Kim, H.S., Choi, E.J., Yu, M.H., Noh, D.Y., and Lee, C. (2012). Proteomic Approach Reveals FKBP4 and S100A9 as Potential Prediction Markers of Therapeutic Response to Neoadjuvant Chemotherapy in Patients with Breast Cancer. J Proteome Res *11*, 1078-1088.

Yates, J.R., 3rd, Speicher, S., Griffin, P.R., and Hunkapiller, T. (1993). Peptide mass maps: a highly informative approach to protein identification. Anal Biochem *214*, 397-408.

Yates, J.R., Cociorva, D., Liao, L., and Zabrouskov, V. (2006). Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. Anal Chem *78*, 493-500.

Yu, X., Schneiderhan-Marra, N., and Joos, T.O. (2010). Protein microarrays for personalized medicine. Clin Chem *56*, 376-387.

Zichi, D., Eaton, B., Singer, B., and Gold, L. (2008). Proteomics and diagnostics: Let's Get Specific, again. Curr Opin Chem Biol *12*, 78-85.