ROYAL INSTITUTE
OF TECHNOLOGY

Affinity Arrays for Profiling Proteins and Autoantibody Repertoires

BURCU AYOGLU

Doctoral Thesis
KTH-Royal Institute of Technology
School of Biotechnology
Stockholm, Sweden 2014

Affinity Proteomics
Science for Life Laboratory
Division of Proteomics and Nanobiotechnology
School of Biotechnology
KTH–Royal Institute of Technology
Tomtebodavägen 23A
171 65 Solna, Sweden

ISBN 978-91-7595-209-3
TRITA-BIO Report 2014:12
ISSN 1654-2312

Printed by US-AB 2014.

Cover artwork by Brendan Monroe.
Other illustrations by Burcu Ayoglu, unless otherwise stated.

To my grandfather Hasan Yurdakul (1914–2006)
For leading me into intellectual pursuits;
To my grandmother Emine Yurdakul (1924–2013)
For teaching me the opposite of indolence.

Nicht Kunst und Wissenschaft allein,
Geduld will bei dem Werke sein.

Johann Wolfgang von Goethe
(Faust - Der Tragödie erster Teil)

–Not art and science serve alone; patience must in the work be shown.–

Abstract

Protein biomarkers hold promise to support diagnosis, prognosis and monitoring of diseases. The need for such biomarkers is escalating in a wide spectrum of diseases, ranging from autoimmune conditions to rare disorders. For discovery and verification of such biomarkers, tools and strategies are needed, which can facilitate a highly multiplex and high-throughput proteomic analysis of body fluid samples. This thesis presents approaches addressing this need by implementing affinity arrays for profiling proteins and autoantibody repertoires.

The common concept behind the investigations carried out in this thesis was to adopt hypothesis-driven, as well as hypothesis-free approaches to create arrays of antigens or antibodies produced within the Human Protein Atlas. Created arrays were used for multiplex profiling of proteins or autoantibody repertoires in body fluid samples. Five such investigations have been selected and included here, which demonstrate the application of these approaches.

In the first investigation, arrays of human protein fragments were implemented for profiling autoantibody repertoires. A multiple sclerosis-related plasma sample collection was profiled for autoantibody reactivity against 11,520 human protein fragments and 51 targets were proposed for further analysis. The second investigation focused on verifying these targets in an independent collection of over 2,000 plasma samples and identified an ion channel protein called anoctamin 2 (ANO2) as an autoimmune target candidate in multiple sclerosis. These two investigations demonstrated the utility of the human protein fragments generated within the Human Protein Atlas for profiling autoantibody repertoires, as well as the potential of the adopted hypothesis-free strategy for discovery of novel autoimmune targets. The third investigation aimed to develop an assay on bead-based antigen arrays for a parallel analysis of different antibody classes, as well as their antigen-specific complement activation property. The utility of this assay was demonstrated in the context of rheumatoid arthritis. The fourth investigation generated profiles for over 3,400 proteins in a multiple sclerosis-related plasma sample collection using bead-based antibody arrays. A selected set of these proteins and other literature-based targets were subsequently profiled in over 400 plasma samples, as well as in cerebrospinal fluid and brain tissue samples. Profiles for five proteins including a transcription factor called interferon regulatory factor 8 (IRF8) were highlighted, which now merit analyses in larger sample collections. In the fifth investigation, profiles for 384 pre-selected proteins were generated on bead-based antibody arrays in plasma and serum samples from muscular dystrophy patients. Analysis of 345 samples originating from three different countries allowed for identification of eleven biomarker candidates including muscle-specific proteins such as myosin light chain 3 (MYL3). These targets hold promise for development of new clinical tests to assist disease diagnosis and monitoring in muscular dystrophies.

Taken together, the investigations presented in this thesis contributed to extend the application range of affinity arrays in the context of different diseases including multiple sclerosis and muscular dystrophies. The presented protein and autoantibody profiling approaches include valuable concepts and strategies, which allow for an efficient discovery and verification of potential biomarker candidates in body fluids by means of affinity arrays.

Popular Scientific Summary

Proteins are the robot molecules of our bodies, doing all sorts of tasks they are programmed to do by our genes. They maintain our health by keeping our cells, tissues, organs and systems up and running. But, proteins are also responsible for and contain information about when the processes in the body do not take place as they should and lead to *diseases*. For example, in case of a disease starting to damage our muscle or brain tissue, a muscle- or brain-specific protein might leak into our *blood*, which would otherwise not be present at a high level in blood. In other cases, our immune system might get out of control and produce *antibodies* against its own proteins. Antibodies are the key proteins of the immune system, recognizing and eliminating foreign molecules originating from viruses and microbes. Yet, if the immune system mistakenly attacks body's own tissues and organs, we might expect blood to contain *autoantibodies* which have attacked the self-proteins.

In all these cases, knowing which proteins are at an altered level than usual in blood or knowing which self-proteins the autoantibodies are mistakenly recognizing can help to understand diseases on a molecular level. This information can be used to detect and monitor diseases by making use of such *biomarkers*. But, how can such biomarkers be discovered to develop tests, which would help doctors to diagnose diseases faster and earlier or to monitor the progress of diseases easier?

The work presented in this thesis can be considered as one of the ways to address this task. Here, the approach is to develop experiments on the so-called *affinity arrays* made of antibodies or proteins. These arrays are miniaturized arrangements of test sites either on microscopic glass slides ("chips") or microspheres ("beads"). They allow to simultaneously analyze hundreds or thousands of proteins or autoantibodies in blood or other body fluids. When this can be done in a comparative way by analyzing several hundreds of samples from patients with a certain disease and healthy individuals, proteins at an altered level or targets of autoantibodies which are not present in non-diseased individuals can be discovered. These can be eventually utilized as disease biomarkers.

In the presented studies, affinity arrays were generated by utilizing the great number of affinity reagents produced within the *Human Protein Atlas*. The Human Protein Atlas is a unique project dedicated to produce representative protein fragments for each human protein. These are also used to immunize rabbits to generate antibodies recognizing human proteins. In this thesis, arrays of human protein fragments were utilized for analysis of autoantibodies in plasma of patients with *multiple sclerosis*, a disabling disease of the central nervous system (**Article I-II**). After using over 11,000 protein fragments and analyzing over 2,500 samples, one interesting protein was identified as a potential target of autoantibodies in multiple sclerosis (**Article II**). Certainly, developing *assays* and experimenting with new assay concepts is closely engaged with application of affinity arrays. In line with this, **Article III** aimed to develop an assay workflow for a more detailed analysis of autoantibodies. In **Article IV** and **Article V**, arrays of several hundreds of antibodies were utilized both in the context of multiple sclerosis and the rare diseases called *muscular dystrophies*, leading to progressive muscle wasting. In both of these studies, a handful of proteins were identified to be differentially abundant in plasma of multiple sclerosis or muscular dystrophy patients.

These and similar discoveries made in the studies within this thesis provide a promising contribution to the understanding of the addressed diseases. These discoveries now require further and more dedicated efforts to establish them as biomarkers. Yet, they demonstrated the utility of affinity arrays for profiling the protein and autoantibody content of body fluid samples. Altogether, the studies presented in this thesis expand the utility of affinity arrays for biomarker discovery.

Populärvetenskaplig Sammanfattning

Proteiner är kroppens robotmolekyler som utför alla uppgifter som de programmeras till av våra gener. De upprätthåller vår hälsa genom att de ser till att våra celler, vävnader och organ fungerar. Då proteiner också är delaktiga när processerna i kroppen inte uppför sig som de ska så kan de ge information om vad det är som inte fungerar när kroppen drabbas av en sjukdom. När till exempel en *sjukdom* uppkommer som angriper vår muskel- eller hjärnvävnad kan det vara så att ett muskel- eller hjärn-specifikt protein som normalt inte finns i *blodet* ändå läcker ut dit. I andra sammanhang kan immunsystemet hamna i obalans och skapa *antikroppar* mot kroppens egna proteiner. Antikroppar är immunsystemets nyckelproteiner som känner igen och oskadliggör främmande ämnen från till exempel virus och bakterier. Om immunsystemet av misstag attackerar kroppens egna vävnader och organ, kan även dessa så kallade *autoantikroppar* finnas i blodet.

Om man vid olika sjukdomstillstånd vet vilka proteiner som det finns förändrade nivåer av i blodet eller om man vet vilka egna proteiner som autoantikropparna har känt igen av misstag, så kan det hjälpa till att förstå varför och hur sjukdomar uppstår. Denna kunskap kan i sin tur användas för att upptäcka och följa en sjukdoms utveckling. Hur kan då dessa så kallade *biomarkörer* upptäckas så att det så småningom finns möjligheter att utveckla blodprov som till exempel skulle kunna möjliggöra att en tidigare diagnos kan ställas?

Det är detta som arbetet som presenteras i denna avhandling handlar om. Mer specifikt om att utveckla metoder och tillvägagångssätt för att hitta nya sjukdomsmarkörer med hjälp av så kallade *affinitets-arrayer*. Dessa utgörs av stora samlingar av proteiner eller antikroppar som används till parallell analys av hundratals autoantikroppar eller proteiner i till exempel blodprov. När dessutom hundratals prover från olika individer med en viss sjukdom kan jämföras med lika många som inte har denna sjukdom på dessa affinitets-arrayer finns det möjlighet att upptäcka proteiner som skulle kunna fungera som markörer för denna sjukdom.

Affinitets-arrayerna som har använts här har möjliggjorts av tillgången till det mycket stora antal antikroppar och proteiner som har skapats inom det KTH-baserade *Human Protein Atlas* projektet. Arrayer med humana proteinfragment har använts för analys av autoantikroppar i plasma från patienter med *multipel skleros* (**Article I-II**). Efter att ha analyserat 11,000 proteiner och 2,500 blodprover kunde vi identifiera ett mycket intressant protein som skulle kunna vara ett viktigt målprotein för autoantikroppar inom multipel skleros och därmed inblandad i utvecklingen av sjukdomen. **Article III** handlar om ett arbete som syftar till att utveckla ett optimerat protokoll och arbetsflöde som möjliggör än mer detaljerad analys autoantikroppar. I **Article IV** och **Article V** har arrayer med flera hundra olika antikroppar använts för analys av blodprover både från patienter med multipel skleros samt *en ovanlig sjukdom där musklerna förtvinar*. I bägge dess två projekt så har en handfull proteiner hittats som man nu kommer att fortsätta att analysera för att se vilken roll de spelar för uppkomsten och utvecklingen av dessa två sjukdomar.

De upptäckter som presenteras i de fem artiklarna utgör ett lovande bidrag till att kunna öka förståelsen inom respektive sjukdom. Men det behövs fortsatta och utökade studier för att de förhoppningsvis ska kunna etableras som biomarkörer. Resultaten visar på den stora potential som finns med affinites-arrayer för storskalig analys av proteiner och autoantikroppar i blod för att upptäcka nya biomarkörer.

Doctoral Thesis Defense

This thesis will be defended on 5 September 2014 at 10:00 in lecture hall Inghesalen (Tomtebodavägen 18A-Karolinska Institute Solna Campus) for a degree of Doctor of Philosophy (PhD) in Biotechnology.

Respondent:
Burcu Ayoglu has a BSc degree in Chemical Engineering with a specialization in Biomolecular Engineering from Bogazici University, Istanbul (2007) and an MSc degree in Medical Biotechnology from KTH-Royal Institute of Technology, Stockholm (2009). Since December 2009, she has been carrying out her PhD studies and research within the Affinity Proteomics group at the Division of Proteomics & Nanobiotechnology at KTH School of Biotechnology, Science for Life Laboratory.

Faculty Opponent:
Paul J. Utz is a professor in Medicine at Stanford University School of Medicine. Following his MD degree from Stanford University School of Medicine and his post-doctoral training at Brigham and Women's Hospital of Harvard Medical School, he first joined Harvard Medical School Faculty and since 1999 he is a faculty member at Stanford University School of Medicine. Professor Utz is an expert in the study of human and murine autoantibodies and autoantigens, apoptosis signaling pathways, animal models of autoimmunity, proteomics and multiplexed assay development for biomarker discovery.

Evaluation Committee:
Marianne van Hage is a Professor of Clinical Immunology at Karolinska Institute, Stockholm and Senior Consultant Clinical Immunologist at the Department of Clinical Immunology, Karolinska University Hospital. She heads the Clinical Immunology and Allergy Unit at the Department of Medicine Solna, Karolinska Institute. The overall aim of Professor van Hage's research is to increase disease understanding in allergy and asthma, and thereby discover novel markers for disease and risk assessment and to develop new strategies for improved allergy vaccination. Novel laboratory analyses are combined with well-characterized clinical data.

Christer Wingren is a Professor at the Department of Immunotechnology at Lund University. Professor Wingren received his PhD degree in Biochemistry at Lund University. Following his post-doctoral training in structural biology at the Scripps Research Institute, La Jolla, he joined the Department of Immunotechnology at Lund University in 1999. He currently directs a research group working on the development of recombinant antibody microarrays for high-throughput disease proteomics with a particular focus on oncoproteomics and autoimmunity.

Ola Söderberg is a Docent at Department of Immunology, Genetics and Pathology (IPG) at Uppsala University. He received his PhD in pathology from Uppsala University and following his post-doctoral research at Linköping University, he joined the Uppsala University. Doctor Söderberg's research within the Molecular Tools group focuses on developing techniques to study protein modification and protein interactions on a single cell level, in particular the establishment and application of multiplex *in situ* proximity ligation assay (PLA) methods.

Chairman of the Thesis Defense:
Mathias Uhlén is a Professor in Microbiology at KTH School of Biotechnology. He received his PhD degree in chemistry at KTH in 1984 and following his post-doctoral research at EMBL, Heidelberg, he became a professor at KTH in 1988. Professor Uhlén, who founded the Human Protein Atlas in 2002, is an expert on the development and use of affinity reagents in biotechnology and biomedicine. He is currently the head of the Division of Proteomics and Nanobiotechnology and the director of the Human Protein Atlas and the Science for Life Laboratory.

Main Supervisor:
Peter Nilsson is a Professor in Proteomics at KTH School of Biotechnology at Science for Life Laboratory. He received his PhD degree in Biotechnology from KTH in 1998. Since 2003, he is the head of the Protein Microarray module within the Human Protein Atlas and since 2013 he is the director of the Affinity Proteomics national platform at Science for Life Laboratory. He is currently also the Vice Dean of KTH School of Biotechnology. Professor Nilsson's research focus is based on the development and utilization of various protein and peptide microarray technologies for biomarker discovery applications.

Co-Supervisor:
Jochen M. Schwenk is a Docent in Translational Proteomics at KTH Science for Life Laboratory. He received his PhD degree in Biochemistry from University of Tübingen in 2005 and continued with his post-doctoral research at KTH Department of Proteomics. He then became a principal investigator within Human Protein Atlas project and in 2010 he was appointed head of Biobank Profiling at KTH Science for Life Laboratory. His research focus is on assay development and protein profiling of body fluids using microarray technologies to translate findings from discovery driven efforts into clinical and biobank analysis.

List of Articles

This doctoral thesis includes the following articles, which can be found in the Appendix.

Article I
Burcu Ayoglu, Anna Häggmark, Mohsen Khademi, Tomas Olsson, Mathias Uhlén, Jochen M. Schwenk, Peter Nilsson (2013). Autoantibody profiling in multiple sclerosis using arrays of human protein fragments. *Molecular & Cellular Proteomics* 12(9): 2657-2672. doi: 10.1074/mcp.m112.026757

Article II
Burcu Ayoglu, Nicholas Mitsios, Ingrid Skelton Kockum, Mohsen Khademi, Ronald Sjöberg, Lars Alfredsson, Mathias Uhlén, Jan Mulder, Tomas Olsson, Jochen M. Schwenk, Peter Nilsson. The calcium-activated chloride channel anoctamin 2 as an autoimmune component of multiple sclerosis. *Manuscript.*

Article III
Burcu Ayoglu, Eszter Szarka, Krisztina Huber, Anita Orosz, Fruzsina Babos, Anna Magyar, Ferenc Hudecz, Bernadette Rojkovich, Tamás Gáti, György Nagy, Jochen M. Schwenk, Gabriella Sármay, József Prechl, Peter Nilsson, Krisztián Papp (2014). Bead arrays for antibody and complement profiling reveal joint contribution of antibody isotypes to C3 deposition. *PLoS ONE* 9(5): e96403. doi: 10.1371/journal.pone.0096403

Article IV
Sanna Byström*, ***Burcu Ayoglu****, Anna Häggmark, Nicholas Mitsios, Mun-Gwan Hong, Kimi Drobin, Björn Forsström, Claudia Fredolini, Mohsen Khademi, Sandra Amor, Mathias Uhlén, Tomas Olsson, Jan Mulder, Peter Nilsson, Jochen M. Schwenk. Affinity proteomic profiling of plasma, cerebrospinal fluid and brain tissue within multiple sclerosis. *Manuscript under review in Journal of Proteome Research.*

Article V
Burcu Ayoglu, Amina Chaouch, Hanns Lochmüller, Luisa Politano, Enrico Bertini, Pietro Spitali, Monika Hiller, Eric Niks, Francesca Gualandi, Fredrik Pontén, Kate Bushby, Annemieke Aartsma-Rus, Elena Schwartz, Yannick Le Priol, Volker Straub, Mathias Uhlén, Sebahattin Cirak, Peter A. C. ′t Hoen, Francesco Muntoni, Alessandra Ferlini, Jochen M. Schwenk, Peter Nilsson, Cristina Al-Khalili Szigyarto (2014). Affinity proteomics within rare diseases: A BIO-NMD study for blood biomarkers of muscular dystrophies. *EMBO Molecular Medicine.* 6(7):918-36. doi: 10.15252/emmm.201303724.

*Equal author contribution.

Author's Contribution to the Included Articles

The contribution of the author of this doctoral thesis to the included articles is as follows:

Article I
Study design, experimental planning and performing of all laboratory work, all data analysis and manuscript writing as the main contributor.

Article II
Study design, experimental planning and performing of all array-based laboratory work, all data analysis and manuscript writing as the main contributor.

Article III
Experimental planning and performing of the array-based laboratory work, parts of data analysis, and manuscript writing as the co-responsible author.

Article IV
Experimental planning and performing of parts of the array-based laboratory work, parts of data analysis, and manuscript writing as a co-responsible author.

Article V
Study design, experimental planning and performing of all laboratory work, all data analysis and manuscript writing as the main contributor.

Related Articles

The following articles have not been included in this doctoral thesis.

Anita N. Kremer, Judith C. van der Griendt, Edith D. van der Meijden, M. Willy Honders, ***Burcu Ayoglu***, Jochen M. Schwenk, Peter Nilsson, J.H. Frederik Falkenburg, Marieke Griffioen (2014). Development of a coordinated allo T-cell and auto B-cell response against autosomal PTK2B after allogeneic hematopoietic stem cell transplantation. *Haematologica* 99(2), 365-369. doi: 10.3324/haematol.2013.086652.

Anna Häggmark, Sanna Byström, ***Burcu Ayoglu***, Ulrika Qundos, Mathias Uhlén, Mohsen Khademi, Tomas Olsson, Jochen M. Schwenk, Peter Nilsson (2013). Antibody-based profiling of cerebrospinal fluid within multiple sclerosis. *Proteomics* 13(15), 2256-2267. doi:10.1002/pmic.201200580

Burcu Ayoglu, Anna Häggmark, Maja Neiman, Ulrika Igel, Mathias Uhlén, Jochen M. Schwenk, & Peter Nilsson (2011). Systematic antibody and antigen-based proteomic profiling with microarrays. *Expert Review of Molecular Diagnostics* 11(2), 219-234. doi:10.1586/erm.10.110

Brigitta Omazic*, J. Matthias Löhr*, ***Burcu Ayoglu****, Ralf Segersvärd, Caroline Verbeke, Isabelle Magaheas, Zuzana Potacova, Jonas Matsson, Alexei Terman, Sam Ghazi, Nils Albiin, Nicolas Kartalis, Peter Nilsson, Liu Zhenjiang, Rainer Heuchel, Jochen M. Schwenk, Johan Permert, Markus Maeurer, Olle Ringdén. Resection of pancreatic carcinoma combined with allogeneic hematopoietic cell transplantation results in long-term survival associated with responses to novel tumor-associated antigens. *Manuscript to be submitted to the New England Journal of Medicine.*

Burcu Ayoglu, Elin Birgersson, Anja Mezger, Josefine Jaxby, Mathias Uhlén, Mats Nilsson, Peter Nilsson, Jochen M. Schwenk. Sequential affinity capture for improved multiplexed protein profiling. *Manuscript in preparation.*

*Equal author contribution.

Preface

During my early attempts to contemplate the structure and content of this thesis, I soon realized that the *motive* for why I would be writing it was of most importance. And the motive was already hidden in the name of the academic degree I am pursuing for: Doctor of Philosophy.

Philosophia: Love of, or the search for, wisdom or knowledge.

I learned in high school how to love to read and write; got an engineering degree where I learned how to play with heavy calculations with love and during the recent years I discovered and practiced how to work with "beads and chips" with love. Writing this thesis was yet the expected next link of this chain: I wrote it because of my love to search for more knowledge about how, why and when affinity arrays were born. I wrote it because of my need to reflect upon how and why I use them; when and why do other researchers use them and what kind of research questions we can and are currently addressing within life-sciences by means of them.

While writing, I anticipated that this thesis might be read by colleagues with a clinical background who are relatively less familiar with the technical aspects; by researchers involved in technology development who are relatively less familiar with certain biological aspects; and by others who are neither clinically nor technically oriented. I therefore aimed to give an overview of the basic biological and technical concepts which I considered relevant within the context of this thesis work.

Believing that a good understanding of the present is built upon the awareness of past ideas and discoveries, I also frequently incorporated a historical perspective within the discussions. The breadth of the scope, in turn, forced me to dive into various areas where I discovered how little I knew. I truly enjoyed getting out of the comfort zone and attempting to explore such concepts which were less familiar to me. Yet, some incomplete presentations of technical or biological facts or aspects might have remained.

Although I payed utmost attention to identify and document seminal works within the field of affinity arrays and their applications, due to the rich scientific output in recent years and due to the limited space available, I certainly could not fully document all important contributions. Any such omission would have been made though in good faith.

Today I might know more than the day I started to write this thesis, yet I still know little. I however feel ready now to experiment and search for more knowledge at this crossroad of medicine, technology and biology. The path of identifying and verifying disease biomarkers might be, or might not be, the modern version of the myth of Sisyphus, who had to roll a big boulder up a mountain - only to watch it plunge to the bottom each time he would approach the summit and to repeat this again and again. Today I feel ready to continue to learn and explore more about new technological tools and biological concepts which I can apply - in order to find out whether the myth is indeed just a myth or not.

Burcu Ayoglu
Stockholm, 11 July 2014

Contents

Part I

Introduction

1 Proteins

Were it not for all the elegant mechanisms involving hundreds and thousands of *proteins* in our bodies, we would not be able to breathe. Or see, hear, walk, think, talk, read or write. Man's interest comes as no surprise to explore this molecular organization created by the proteins to establish and maintain *health*, as well as to understand the molecular chaos created by them in *disease*.

Today we know that the protein domain is the most ubiquitously affected molecular domain in human diseases. In this introductory chapter, I will give a very brief and mostly historical overview of the discoveries and accomplishments, which built up this realization and interest in exploring proteins with the aim to identify molecular signatures of diseases.

1.1 Proteins as Workforce Molecules of Life

> Cellular reality is more elaborate than the dreams of even the nucleus itself.
>
> *Anderson and Anderson (1998)*

Proteins are nature's robot molecules executing the biochemical tasks dictated by the information encrypted within the genetic code (Crick, 1970). As indicated by the origin of the term, *proteios*, meaning "primary" or "of first rank" in Greek (Hartley, 1951, Vickery, 1950), proteins are the primary molecules involved in nearly every structure and process of life: They serve as enzymes to facilitate biochemical reactions; they serve as hormones for endocrine signaling; they transport oxygen, nutrients and metabolic waste products; they provide structural framework at cellular and tissue level or they provide defense against pathogens.

Following the coining and adoption of the term *protein* by Jöns Jacob Berzelius (1779–1848) and Gerardus Johannes Mulder (1802–1880), respectively, Hermann Emil Fischer (1852–1919) was the one to propose the term *peptide* (Fischer, 1906), which was

later supported by Frederick Sanger (1918–2013), who determined the primary structure of insulin in 1953 and provided further evidence for the peptide nature of proteins (Sanger, 1945). Kaj Ulrik Linderstrøm-Lang (1896–1959) subsequently introduced the terminology which we still use to describe protein structure (Linderstrøm-Lang, 1952): The *primary structure* describing the linear sequence of amino acids; the *secondary structure* describing the hydrogen bond stabilized structures such as α-helices and β-sheets; the *tertiary structure* describing the folding of the polypeptide chain leading to a compact globular structure and the *quaternary structure* describing the arrangement of multiple protein or polypeptide subunits (Figure 1.1).

Figure 1.1: Waltz of the polypeptides.
A sculpture representing the tertiary structure of the B lymphocyte activating factor (BAFF or BLyS) protein playing a crucial role in maintaining normal immunity. Part of a 70 m. long mixed-media artwork by Mara Haseltine titled "Waltz of the Polypeptides" installed at Cold Spring Harbor Laboratory (CSHL) Long Island, NY. Photo taken by Burcu Ayoglu at CSHL, July 2012.

The sequence of the four different nucleotides in the genetic code encodes for twenty different amino acids (Crick et al., 1961), hundreds of which are combined in different ways to build the primary structures of proteins, which then become functional upon folding into their tertiary structures (Anfinsen, 1973). As of the date this thesis was written, there were 20,805 protein-encoding genes in the human genome, building the non-redundant collection of human proteins (Ensemble Database version 75.37, Flicek et al. (2014)). As a

result of a process called *alternative splicing* (Leff et al., 1986, Mironov et al., 1999), each of these protein-encoding genes can yield several protein variants in eukaryotes, or several *proteoforms*, in a recently introduced terminology by Smith et al. (2013). The diversity of human proteins is increased further by several *co- or post-translational modifications* (PTMs). These and other possible forms of variations, such as combinatorial variations generated by somatic deoxyribonucleic acid (DNA) rearrangements, increase the estimated number of all human proteins up to hundreds of thousands, if not few millions (Harrison et al., 2002, Uhlén and Ponten, 2005). Since Marc Wilkins's proposition of the term in late $1990's$, this entire complement of proteins expressed by a genome is described as the *proteome* (Wilkins et al., 1996).

The Human Genome Project (HGP), conceived in $1980's$ and launched in 1990, was announced as complete in April 2003, celebrating the 50^{th} anniversary of Francis Harry Crick (1916–2004) and James Dewey Watson's discovery of the structure of DNA (Watson and Crick, 1953). HGP delivered a finished human reference genome (Human Genome Sequencing Consortium, 2004) and this was considered as one of the most important scientific undertakings of all time. HGP was compared to the Apollo Moon Landing Project to indicate that both projects held the promise to provide the foundation for expanding man$'$s knowledge: one in space science and the other regarding the molecular basis of life. Yet, soon after this accomplishment, the inadequacy of the sequence of the human genome alone in elucidating physiological processes became clear and the attention turned to the next challenge, namely to "decipher" the human proteome. A similar effort aiming to characterize the human proteome was initiated in 2010 by the Human Proteome Project (HUPO-The Human Proteome Organization, 2010). Despite the enormously complex nature of the human proteome as opposed to the finite nature of the genome, a deep exploration of the proteome and its changes under the influence of biological perturbations, such as disease or drug treatment, is currently still one of the collective ambitions in life sciences.

1.2 Proteins as Disease Biomarkers

> Declare the past, diagnose the present, foretell the future; practice these acts.
>
> *Hippocrates*

Although the first use of the term *biomarker* in a published report dates back to 1980 (Paone et al., 1980), the first biomarker, Bence-Jones protein, was in fact discovered in 1847 (Jones, 1848) by the "father of clinical chemistry" Henry Bence Jones (1813–1873). More than 100 years later, it was characterized to be a tumor-produced free antibody light

chain (Korngold and Lipari, 1956) and was approved by the Food and Drug Administration (FDA) in 1998 as a diagnostic biomarker for multiple myeloma.

Yet, a widely accepted definition of a biomarker, namely a *biological marker*, was not provided until the end of 1990′*s* (National Institute of Health-Definitions Working Group, 2000). As suggested by the National Institutes of Health (NIH), a biomarker is "a characteristic that is objectively measured and evaluated as an indication of a normal biologic process, a pathogenic process or a pharmacologic response to a therapeutic intervention". Although this definition engulfs physical features (e.g. change in skin color) or physiological metrics (e.g. body temperature), the term is currently used rather as a shorthand for *molecular biomarker*, as I will also refer to in the following parts of the thesis.

Biomarkers can be categorized in several ways and can take several forms regarding the type of biomolecule being used as an indicator (Aronson, 2005). While the discussion in this thesis will rather focus on protein biomarkers, there are other forms of molecular biomarkers such as DNA, RNA or metabolite biomarkers, which can be present or excreted by the body and can be measured in biological media such as body fluids, tissues or cell lines. Biomarkers are needed in various clinical settings for different purposes: *diagnostic biomarkers* for detection or staging of a disease; *prognostic biomarkers* for forecasting the likely course of a disease and *response biomarkers*, or sometimes referred to as *predictive biomarkers*, for identification of sub-populations of patients which might respond to a given treatment (Chatterjee and Zetter, 2005).

Certainly, the greatest area of benefit, both for individuals and the health care system, relates to the detection of early stage of diseases, rather than to the treatment of diseases at a more advanced stage (Etzioni et al., 2003). Such protein biomarkers are therefore anticipated as the potential enablers of a more preventive, more stratified and even personalized medicine, which are concepts being envisioned for almost two decades (Hanash, 2003). Due to high mortality rates, the quest for such biomarkers for various cancer types has been certainly dominating the efforts within the field of protein biomarker discovery. However, the need, interest and investments have escalated in a broad spectrum of other conditions with continuously increasing prevalence rates, ranging from neurodegenerative disorders such as Alzheimer's disease to autoimmune diseases such as type 1 diabetes. The investigations presented in this thesis aimed to identify such biomarkers, for instance, in the context of multiple sclerosis and muscular dystrophies. In later chapters and as the main subject of this thesis, I will discuss some of the available tools and strategies which can facilitiate the quest for disease biomarkers.

2 Antibodies

Affinity can be defined as the degree of the attractive force between molecules, which causes them to interact with each other. The complex machinery within all living organisms, including man, is based on affinity. Yet, affinity is not only a natural phenomenon but also a powerful and versatile tool in the toolbox of the life-scientist.

The main subject of this chapter will be *antibodies*, which can be considered as proteins specialized to function as affinity proteins. I will first describe antibodies as a part of the functional immune system and then as assassins serving the "dark-side" of a dysregulated immune response, leading to autoimmune conditions. I will finally describe antibodies as *affinity reagents* and highlight their utility as a research tool offering possibilities to explore the human proteome.

2.1 Antibodies as Key Molecules of the Immune System

> Studying the immune system is like peeling an onion: beneath each layer we find another; chopping the onion will bring tears . . . only during peeling does it speak the truth.
>
> *Lakkis (2012)*

The immune system is like an orchestra consisting of various types of tissues, cells and molecules, where no player alone can create a symphony on its own, yet each of them and their interactions are needed for the symphony. This orchestra of the immune system allows for recognition and elimination of a diverse range of pathogens and toxins to protect an organism, while sparing its own constituents.

Immunological mechanisms can take place at different levels, which can be categorized under the two main forms of immunity: the *innate* and the *adaptive immunity* (Figure 2.1). Innate immunity is the "inborn" first-line of defense, not providing an antigen-specific recognition, and is found in all metazoan organisms. The more sophisticated adaptive or

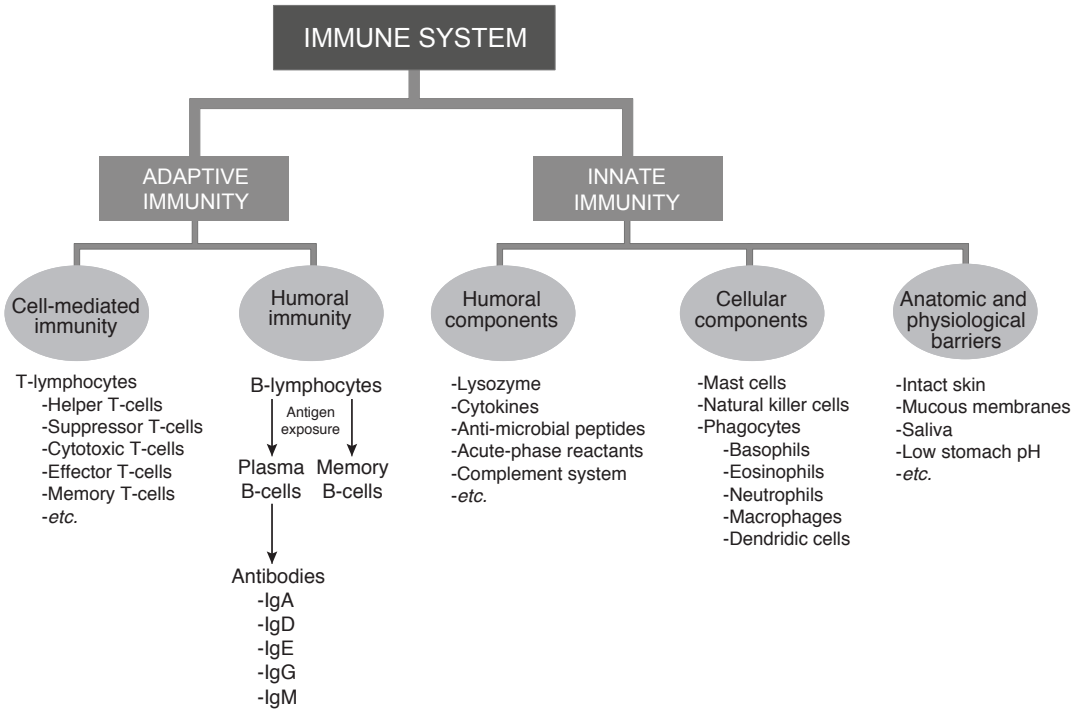

Figure 2.1: A simplified view of the immune system.
An overview of different levels and different types of cells of the immune system. A functional immune system can be viewed to consist of anatomic and physiological barriers and innate immune mechanism providing a first-line defense, and the adaptive immune mechanism providing a second-line defense. The cell-mediated adaptive immunity is regulated by T-lymphocytes and the humoral adaptive immunity is mediated by different classes of antibodies.

"acquired" immunity, found only in vertebrates, displays a high degree of specificity and memory. The components of the innate immune system include various anatomical and physiological barriers, as well as various types of cells, such as neutrophils and macrophages and several secretory molecules, such as proteins of the complement system. Similarly, the adaptive immunity has two branches: the *cell-mediated immunity*, regulated by T-cells and the *humoral immunity* mediated by antibodies.

The antibody production pathway begins when the antigen-binding receptors of B-cells recognize and bind an antigen circulating in the *humor* – a medieval physiology term for *body fluid*. This initiates the recruitment of helper T-cells secreting cytokines, which in turn help the activation and proliferation of B-cells into antibody-secreting plasma cells. The secreted antibodies have two essential roles - the first is to bind to the *epitope* of the antigen via their antigen-binding domains residing at the tip of the arms and the second is to flag the antigen and ensure its removal through various effector functions such as activation of complement pathways or phagocytosis via their effector domain residing in the tail (Elgert, 2009, Mak et al., 2013, Parkin and Cohen, 2001).

The fundamentals of the innate immunity were laid by Elie Metchnikoff (1845–1916), who discovered and coined the terms *phagocytes* and *phagocytosis* in 1894. In 1900, Paul Ehrlich (1854–1915) proposed the side-chain theory of antibody formation, laying the principles of adaptive immunity. During the beginning of the 20^{th} century, the theories of innate and adaptive immunity created an epic debate, or rather a battle, and the field of immunology witnessed a dichotomous period regarding whether the source of an immune response is innate or adaptive (Kaufmann, 2008, Silverstein, 1991). Yet today, the importance of and the close and complex interplay between both innate and adaptive arms of the immune systems is acknowledged (Fearon, 1997, Fearon and Locksley, 1996).

2.1.1 History, Anatomy and Functions of the Antibody Molecule

> Antibodies can be likened to ready-made suits. The antigen is a buyer who decides to pick a number of different suits that fit more or less well, rather than instruct a tailor to make one suit to fit him. To be satisfied, the buyer must patronize a store with a very large stock of suits in a great variety of sizes and styles. The immune system is like a store with an almost unlimited stock, one ready to please any possible customer.
>
> *Edelman (1970)*

Our comprehension of antibodies derives from a long series of classical observations related to early studies of vaccination by Edward Jenner (1749–1823), Louis Pasteur (1822–1895) and Robert Herman Koch (1843–1910), the fathers of immunology. The term *Antikörper*, the German word for *antibody*, was first used in a work of Paul Ehrlich in 1891 (Ehrlich, 1891), although a year earlier Adolf Emil von Behring (1854-1917) and Kitasato Shibasaburō (1853–1931) had described them as *antitoxins* (Behring and Kitasato, 1890), setting the grounds for the humoral theory of immunity. Following the works of these architects of *immunochemistry*, as coined by Svante August Arrhenius (1859–1927), subsequent technological advancements such as the invention of ultracentrifugation by Theodor Svedberg (1884–1971) (Svedberg and Rinde, 1924) and electrophoresis by Arne Wilhelm Tiselius (1902–1971) (Tiselius, 1937) led to the finding by Tiselius and Elvin Abraham Kabat (1914–2000), that antibodies are found in the gamma-globulin fraction of serum (Tiselius and Kabat, 1939), providing the origin of the term *immunoglobulin* (Ig). The establishment of the anatomy of antibody molecule was enabled mainly by the

"molecular surgery" works of Gerald Maurice Edelman (1929–2014), Rodney Robert Porter (1917–1985) and Alfred Nisonoff (1923–2001), who, using different strategies, found that enzymatic cleavage or reduction of the antibody molecule results in stable fragments of different sizes (Edelman, 1959, Nisonoff et al., 1960, Porter, 1959).

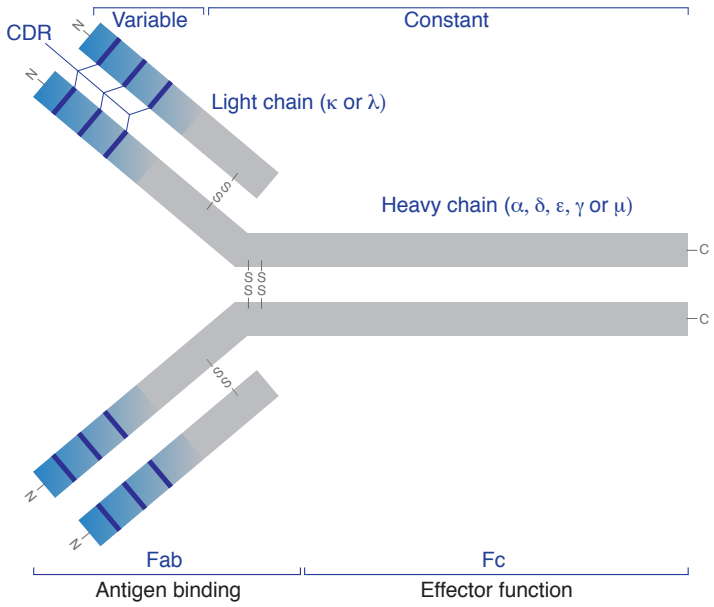

Figure 2.2: Structure of an immunoglobulin.
An antibody molecule is made of four polypeptide chains: two heavy chains and two light chains, each consisting of a variable region containing complementarity-determining regions (CDRs) and a constant region. The heavy and light chains are linked by inter- and intra-chain disulfide bonds. The arms of the antibody molecule are termed the Fab fragments and the constant regions of the heavy chain make the Fc fragment, which has several effector functions. There are two types of light chains, κ and λ, with no known functional differences, whereas the five different types of heavy chains, α, δ, ϵ, γ and μ, determine whether an antibody is of IgA, IgD, IgE, IgG or IgM class.

Owing to all the remarkable work by all these, and undoubtedly several other contributors not mentioned above, today we know that the structural unit of antibodies is a glycoprotein consisting of about 1,300 amino acids and four polypeptide chains, two identical light and two identical heavy chains (Figure 2.2). The *heavy chains*, each of a molecular weight (MW) of $\sim$ 50,000 Daltons (Da), and the *light chains*, each of a MW of $\sim$ 25,000 Da, are stabilized and linked by inter- and intra-chain disulfide bonds. The regions of the heavy and light chains involved in antigen binding are called *variable regions* and the remaining portions are called *constant regions* of each chain. The identical "arms" of the antibody molecule are termed *Fab fragments* (fragment, antigen-binding), whereas the constant regions of the heavy chain make the *Fc fragment* (fragment, crystalline), which has many effector functions such as binding to various proteins of the complement system or

interacting with cell surface receptors called *Fc receptors*, expressed e.g. on macrophages and monocytes. The variable regions of the heavy and light chains each contain three regions with extremely variable amino acid sequences, the *hyper variable regions*, each 5-10 amino acids in length. These regions, called *complementarity-determining regions* (CDRs) constitute the actual binding surface to the antigen, the so-called *paratope*, and also determine the *specificity* of the antibody (Elgert, 2009, Kindt et al., 2007). The antigenic determinants recognized by the antibodies are called the *epitopes* of the antigen, which can be either *continuous* or *discontinuous**. *Antibody affinity* can be considered as the strength of interaction between an epitope of an antigen and the paratope of an antibody. The affinity of antibody-antigen interaction is defined by the same thermodynamic principles which applies to any reversible biochemical interaction. It is therefore often expressed by the equilibrium dissociation constant K_d. *Avidity*, on the other hand, can be considered as the "functional affinity" as it describes the overall strength of an antigen-antibody complex which is determined not only by the affinity, but also on the *valency* of both the antigen and the antibody (Karush, 1978).

There are two types of light chains present in immunoglobulin molecules of humans, which are denoted κ and λ. Although there are yet no known functional differences between κ and λ-containing antibodies, the five different types of heavy chains denoted α, δ, ϵ, γ and μ, determine whether an immunoglobulin is of IgA, IgD, IgE, IgG or IgM class, respectively, each with closely related but different structural and functional characteristics (Figure 2.3). IgA, IgD and IgM antibodies have also two subclasses, whereas IgG has four subclasses.

The main functions of these five antibody classes, building the antibody armory of the humoral adaptive immunity, can be briefly outlined as follows: IgE is found in trace amounts in blood serum but it is bound to basophils and mast cells throughout the body and is associated with hypersensitivity and allergic reactions, as well as anti-parasitic activity. IgD antibodies remain membrane-bound and they function as an antigen receptor on B-cells which have not been exposed to antigens yet. IgA antibodies are found in several mucosal surfaces in the body, such as the gut, urogenital tract and respiratory tract, where they provide localized protection against pathogens. They have a half-life of six-eight days and they constitute 10-15% of the total serum immunoglobulin. IgM constitutes 5-10% of the total serum immunoglobulin. It is the first immunoglobulin expressed during a primary immune response against an antigen, whereas IgG is the predominant antibody in serum during secondary immune response, namely upon subsequent encounter with the same antigen. The Fc portion of IgG antibody can bind to phagocytes, thus it enhances

*In fact, the nomenclature used to describe epitope classes is still ambiguous. One nomenclature, originally proposed by Sela et al. (1967), classifies epitopes as *sequential vs. conformational*. Another nomenclature, proposed by Hurrell et al. (1977), distinguishes between *linear vs. spatial* epitopes. In this thesis, I have preferred to use the nomenclature proposed by Atassi and Smith (1978), who classified epitopes as *continuous vs. discontinuous*.

phagocytosis. IgG can also cross the placenta and provide passive immunity to fetus. Antibodies of IgG class have the longest half-life in serum (23 days) and they constitute 80% of the total serum immunoglobulin. Both IgM and IgG can lead to opsonization (coating) of an antigen for destruction and can activate the complement system, as I will discuss in more detail in the next section (Elgert, 2009, Janeway et al., 2001, Wood, 2006).

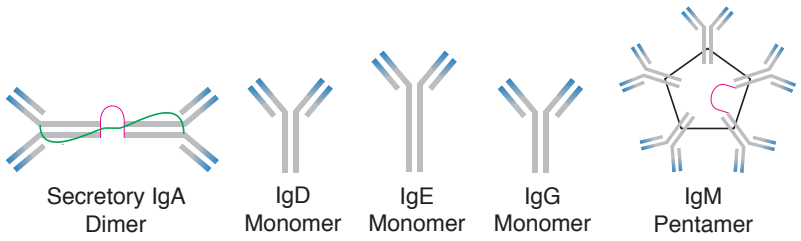

Figure 2.3: Classes of immunoglobulins.
Representation of the five classes of immunoglobulins in humans. IgD, IgE and IgG are present as monomers. IgA molecules are produced as monomers but plasma cells secrete them as dimers composed of two monomeric IgA molecules covalently linked through the J chain (colored in red), a 137 amino acid long polypeptide, and a secretory component (colored in green). IgM is mostly a pentameric molecule consisting of five covalently linked monomers and a J chain (colored in red), although it also exists in a hexameric form without a J chain. IgA monomers, IgD, IgG and IgE have two antigen binding sites, thus they are bivalent, whereas IgM pentamers are decavalent.

2.1.2 Antibodies as Activators of the Complement System

> Complement has truly come out of hiding and shown fascinating connections we had never before imagined, and these hidden connections might indeed be stronger than the original, obvious ones.
>
> *Ricklin et al. (2010)*

While Metchnikoff proposed the role of phagocytes in the host immune response against microbial pathogens, other microbiologists, such as Hans Ernst Buchner (1850–1902) and Jules Vincent Bordet (1870–1961) became fascinated about the humoral response to microbial pathogens and discovered mysterious properties of the blood (Ferenčík, 1993). In 1890, Buchner described a "substance" in cell-free blood serum which was capable to lyse bacteria and he named this substance *Alexin* (Buchner, 1889). In 1894, Bordet demonstrated that if serum was stored at room temperature or heated at 55° C for 30 minutes, the bacteria could still be agglutinated but serum's ability to lyse bacterial cells

was lost, which however could be restored by addition of fresh, unheated serum. Bordet subsequently postulated that serum's ability to lyse bacteria and red blood cells depended on two components: a stable and specific substance (namely the agglutinating *antibodies* as discovered by Behring in 1890) and a heat-sensitive, non-specific *Alexin* responsible for lysis of cells or bacteria (Bordet and Gengou, 1901). Paul Ehrlich and Julius Morgenroth (1871–1924), who simultaneously conducted similar studies and obtained resembling results, named this heat-sensitive component *das Komplement* as it "complemented" the antibody response, yet as opposed to Bordet, Ehrlich proposed that each antibody would be not only specific for an antigen but also for a complement molecule (Ehrlich and Morgenroth, 1899). Later, Bordet's concept of a rather non-specific complement proved to be correct, although Ehrlich's nomenclature survived (Silverstein, 2001b, 2009) and in 1907, Adolfo Ferrata (1880–1946) was the first to propose that complement is not a single entity but rather a complex of proteins present in normal serum (Ferrata, 1907).

The *complement system* has an "ancient" origin (Nonaka, 2001). It is the major humoral component of the innate immunity and underlies one of the main effector functions of antibody-mediated immunity. As recently reviewed by Ricklin et al. (2010), it consists of over 30 different serum proteins and cell surface receptors, where the nomenclature of them ("C" followed by a number or "Factor" followed by a letter) originated from the order of their discovery. The majority of the complement system proteins are *zymogens*, namely inactive proenzymes, which become activated upon proteolysis. Once activated, they become proteases, which cleave other complement proteins to activate them in turn. Thus, the proteins of the complement system work in an amplifying cascade, similarly to other humoral cascade system such as coagulation (Janeway et al., 2001, Walport, 2001). The majority of the complement system proteins such as C3, the pivotal protein of the complement system (Erdei et al., 1991), are produced and secreted by the hepatocytes (liver cells) (Alper et al., 1969). However other proteins such as C1q, C7 and properdin have been shown to be derived e.g. from monocytes and tissue macrophages as reviewed by Morgan and Gasque (1997).

The complement system can be activated through three different biochemical pathways, as reviewed e.g. by Dunkelberger and Song (2010) and outlined in Figure 2.4. Although these pathways differ in the mechanisms initiating them, they all result in the generation of homologous enzyme complexes, namely C3 convertases, that can cleave the most abundant complement protein C3. Cleavage of C3 is followed by a terminal stage involving the cleavage of C5, which is common for all the complement activation pathways and involves creation of the membrane attack complex leading to direct osmotic lysis of cells (Müller-Eberhard, 1986), that can be cleared by specialized macrophages such as Kupffer cells. Importantly, complement activation is regulated to minimize damage to host cells, which is achieved by a series of both soluble and membrane-bound complement regulatory proteins (Kim and Song, 2006, Zipfel and Skerka, 2009) .

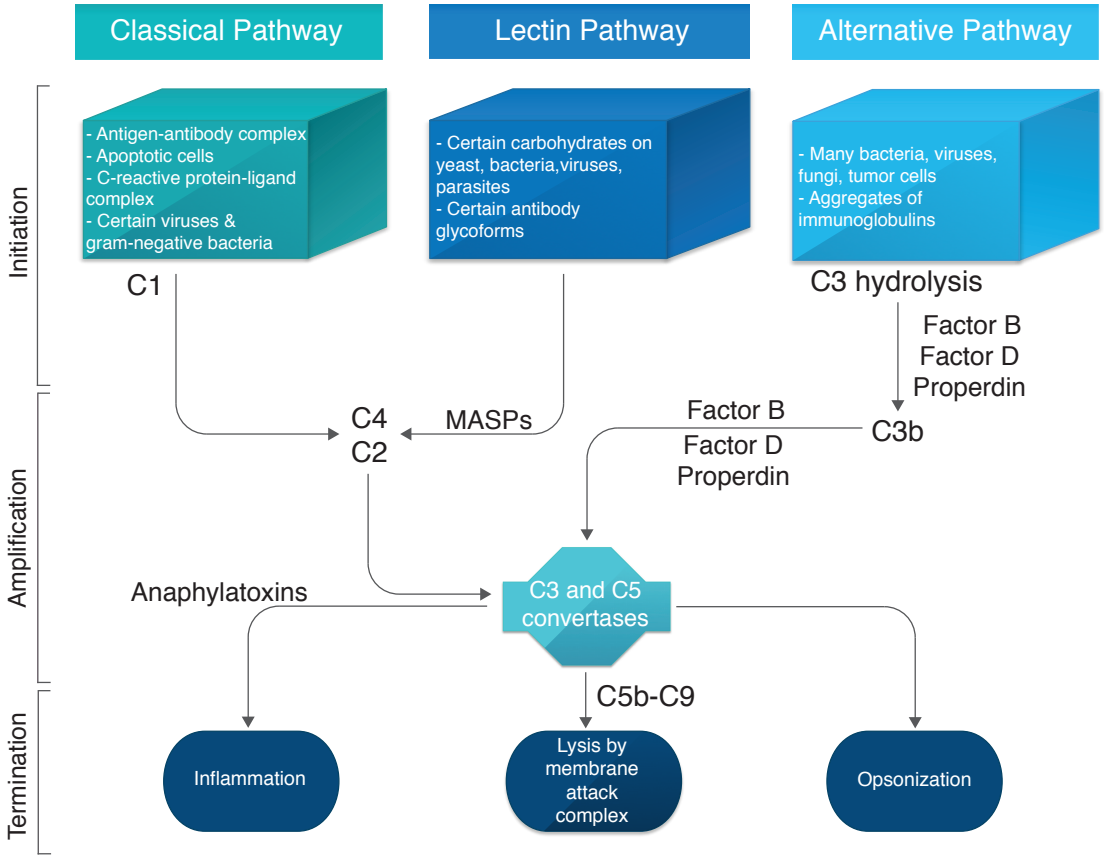

Figure 2.4: A simplified overview of the complement activation pathways.
The complement system can be activated via three major pathways: classical, lectin and alternative pathway. The alternative pathway is activated by spontaneous and continuous hydrolysis of complement component 3 (C3) molecule in the presence of Factors B and D, leading to formation of C3 and C5 convertases stabilized by properdin. The classical pathway is induced mainly by C1 binding to antibody-antigen complexes, resulting in the cleavage C4 and C2. The lectin pathway is activated when mannose-binding lectin or ficolins encounter conserved pathogenic carbohydrate motifs, upon which several mannose-associated serine proteases (MASPs) get activated, cleaving again C4 and C2. Cleavage products of C4 and C2 build the C3 convertase of the classical and lectin pathways, which cleaves C3 into C3a and C3b, where the latter forms the C5 convertase of the classical and lectin pathways. All the three pathways converge in the formation of the convertases, which in turn produce the three main effectors of the complement system: The membrane attack complex (MAC), opsonins (e.g. C3b) and anaphylatoxins (C3a, C4a and C5a). C3b induces phagocytosis of opsonized antigens; MAC, which is an assembly of complement proteins C5b through C9, enables direct osmotic lysis of cells and the anaphylatoxins are potent, cytokine-like polypeptides mediating inflammation.

The pathway that was discovered and characterized first is the *classical pathway*, which is typically initiated by IgG or IgM antibodies in complex with antigens under the presence of C1, C4 and C2 molecules and calcium and magnesium cations (Ca^{2+} and Mg^{2+}) (Müller-Eberhard, 1988). IgM antibodies are the most efficient antibodies to activate the complement system. Among the IgG subclasses, IgG1 and IgG3 are the most effective, followed by IgG2, where IgG4 is not able to activate the complement system. The *alternative pathway*, which in fact is evolutionarily older than the classical pathway, is not dependent on antigen-antibody complexes. It can be activated upon spontaneous hydrolysis of C3 to C3b protein and formation of a C3 convertase directly on pathogenic cell surfaces in the presence of Mg^{2+}, properdin, Factor B and Factor D in normal serum (Müller-Eberhard and Schreiber, 1980). The more recently discovered *lectin pathway* (Ikeda et al., 1987) is activated upon binding of mannose-binding lectin or ficolins and other mannose-associated serine proteases to carbohydrates or acetyl groups on the cell surface of pathogens (Fujita et al., 2004). Accordingly, the lectin and alternative pathways are the components of the innate immune response, whereas the classical pathway is a crucial mechanism of the antibody-mediated adaptive humoral response.

2.2 Antibodies as Hallmarks of Autoimmune Diseases

> The conception that antibodies, which should protect against disease, are also responsible for disease, sounds at first absurd.
>
> *Clemens Peter Freiherr von Pirquet*

In 1901, Paul Ehrlich postulated his theory of *Horror Autotoxicus* (Ehrlich and Morgenroth, 1901), which has been mostly -and incompletely- interpreted over decades to claim that immune response can be directed against the non-self antigens but not against the self-antigens. As clarified retrospectively (Silverstein, 2001a), Ehrlich's Horror Autotoxicus theory did in fact not entirely prohibit the formation of *autoantibodies* and did not discount the possibility of *autoimmunity*, but rather suggested that even such autoantibodies, namely *autotoxins*, would be formed, they would be prevented by certain (and by that time still unknown) mechanisms from exerting any destructive effect in the body.

Ehrlich's theory, or at least the perception it created, received its first challenge when Julius Donath (1870–1950) and Karl Landsteiner (1868–1943) provided the first solid evidence that autoantibodies might exist and be the cause of human diseases, in this case the disease called paroxysmal cold hemoglobunaria (Donath and Landsteiner, 1904). Yet interestingly,

due to Ehrlich's influential theory, the likelihood of autoantibodies being present and causing a destructive autoimmune reaction was scarcely contemplated, let alone accepted until the $1960's$ (Jennette and Falk, 2010, Silverstein, 2009).

In 1957, Frank Macfarlane Burnet's (1899–1985) landmark *clonal selection* theory of antibody formation provided the possible mechanisms for Ehrlich's anticipation regarding how the immune reactivity against self-antigens could be prevented from exerting any destructive effect (Burnet, 1957). Incorporated into Burnet's clonal selection theory –namely that each B-cell has a specificity for a particular antigen and binding to this antigen results in proliferation and production of increased amount of antibody– was also the hints of a *clonal deletion* theory, clarified later by Joshua Lederberg (1925–2008) (Lederberg, 1959). The clonal deletion theory claimed that B-cells recognizing any self-antigens would be "deleted" from the immune repertoire before achieving functional maturity, with the exception of the *sequestered antigens* expressed in sites of no contact with immune system, such as lens, brain or sperm (Dighiero and Rose, 1999). Ehrlich's aversion to the concept of autoantibodies and autoimmunity, later inherited by Burnet and incorporated into a comprehensive theory, was challenged again 50 years later, when thyroid-specific autoantibodies could be identified in the blood serum of patients with Hashimoto's disease (Doniach and Roitt, 1957), the first disease to be recognized as an autoimmune disease. Subsequent discoveries in other disease contexts about the presence of autoantibodies provided the stimulus for a further understanding of autoimmune diseases and the mechanisms behind them.

2.2.1 Autoimmunity as a Cause of Disease

> Everything is autoimmune until proven otherwise.
>
> *Shoenfeld (2013)*

The potential antigen recognition diversity of the immune system in terms of its B-cell and T-cell receptors (BCRs and TCRs) is achieved via somatic gene modifications in BCR and TCR genes in the central lymphoid tissues (bone marrow and thymus) and peripheral lymphoid tissues (spleen, tonsils, lymph nodes) and is estimated to be $2.5\text{x}10^7$ for each cell type (Janeway et al., 2001). During this process of random antigen receptor production, between 20 to 60% of BCRs and TCRs inevitably display self-reactivity, as shown for instance for immature B-cells (Wardemann, 2003).

This large fraction of self-reactive early BCRs and TCRs provides evidence for control mechanisms preventing B- and T-cells from attacking almost every other protein in one's own body tissues. These control mechanisms, referred to as *tolerance*, develop

in the bone marrow for B-cells and in the thymus for T-cells by means of *apoptosis* (programmed cell death) and *clonal deletion* –as originally envisaged by Burnet– but also by means of *receptor editing*, which allows editing of the self-reactive receptors by further gene modifications to display not self-reactive receptors. Self-reactive B- and T-cells escaping these *central tolerance mechanisms* in the central lymphoid tissues and migrating to peripheral lymphoid tissues can still be kept in an *anergic* state (i.e. functionally inactivated or unresponsive). This can occur via e.g. limited supply of essential growth factors or via mechanism involving active suppression by regulatory T-cells, as reviewed by Schwartz (1989), Sakaguchi (2000) and Goodnow et al. (2005).

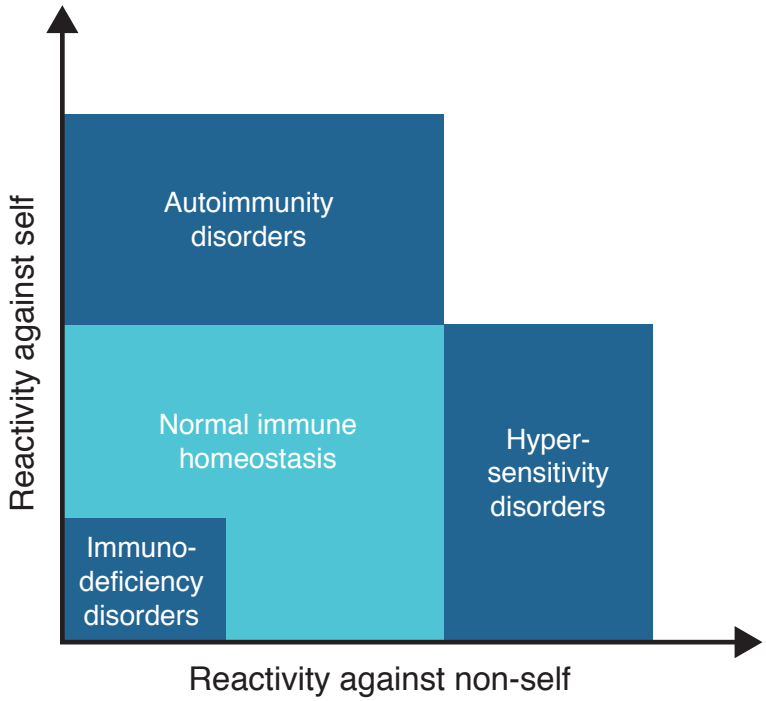

Figure 2.5: Conceptual representation of the self and non-self immune reactivity. A balanced level of immune reactivity between self and non-self is crucial for the maintenance of a normal immune homeostasis. A dysregulated and elevated self-reactivity results in autoimmune diseases. Similarly, a dysregulated and either diminished or elevated non-self reactivity results in allergy and hypersensitivity diseases or immuno-deficiencies, respectively.

Dysregulation of these checkpoint mechanisms results in disruption of the homeostatic balance between self- and non-self reactivity and facilitates the development of autoimmunity. Contemporary and revised views acknowledge that a certain degree of autoimmunity is inevitable and indeed physiologically needed (Schwartz and Cohen, 2000). A balance between self- and non-self reactivity is believed to be crucial for the maintenance of a normal immune homeostasis (Avrameas and Selmi, 2013). Thus, as illustrated in Figure 2.5, autoimmune diseases arise due to a dysregulated and elevated self-reactivity, where similarly, a dysregulated and either diminished or elevated non-self reactivity can

lead to other immunolopathological conditions, such as immuno-deficiencies or allergy and hypersensitivity diseases (Jennette and Falk, 2010). Interestingly, these conditions can even concomitantly occur in the same patients (Bartůnková et al., 2009, Etzioni, 2003).

The possible mechanisms underlying this dysregulation of self-tolerance processes are still poorly understood, however the onset of autoimmune diseases is believed to be multi-factorial, involving complex interactions between genetic susceptibility, exogenous factors, hormonal factors and environmental triggers as reviewed by Christen and Herrath (2004), Atassi and Casali (2008) and Shoenfeld et al. (2008b). Regarding the genetic factors, strong associations have been established for over fifty years between genetic susceptibility to autoimmunity and particular alleles of the *major histocompatibility complex* (MHC) class II genes* (Klein and Sato, 2000). MHC molecules play a crucial role in antigen presentation to T-cells. Certain MHC polymorphisms can therefore potentially influence the susceptibility to a given autoimmune condition by enhancing the presentation of antigens in the periphery. This, in turn, might results in increased T-cell activation, or by ineffective presentation of self-antigens during central tolerance in the thymus, which can cause e.g. more aggressive T-cells or fewer numbers of regulatory T-cells (Christen and Herrath, 2004).

Infectious agents, such as the Epstein-Barr virus (Pender, 2004), are considered among exogenous factors with a potential to trigger different mechanisms by which infection can lead to onset of an autoimmune response (Kivity et al., 2009, Pordeus et al., 2008, Wucherpfennig, 2001). For instance, a usual immune response against an infectious agent with sequence or structural similarities to a self-antigen might result in cell clones recognizing self-antigens, a mechanism referred to as *antigen mimicry* (Blank et al., 2007). Similarly, a so-called *bystander activation* can non-specifically wake anergic cells with a self-reactive potential and lead to enhanced processing and presentation of self-antigens. This can further stimulate an expansion or spreading of immune response towards other self-antigens, a mechanism referred to as *epitope spreading* (Vanderlugt and Miller, 2002). Besides infectious agents, other environmental factors, such as prolonged exposures to certain toxic chemicals (Yoshida and Gershwin, 1993) or drugs (D'Cruz, 2000) might play a role in onset of autoimmune diseases.

Currently, there are over 80 diseases recognized as or suspected to be autoimmune conditions (Hayter and Cook, 2012). They cumulatively affect 5-8% of the population in Europe and USA (Eaton et al., 2007, Shoenfeld et al., 2008a), with around 75% of the patients being female (Gleicher and Barad, 2007, Whitacre, 2001). Autoimmune diseases constitute the third most common disease group in USA after cancer and cardiovascular diseases (National Institute of Health-The Autoimmune Disease Coordinating Committee,

*In humans, the locus of genes on chromosome 6 encoding for MHC is referred to as *human leukocyte antigen* (HLA) system.

2005). Majority of the autoimmune diseases are life-long, chronic and often disabling conditions manifesting themselves at a relatively young age. Thus, they have a disproportionate impact on health care system, for instance with an estimated yearly direct cost of over $100 billion in USA alone (Shoenfeld et al., 2008a). Autoimmune diseases are also a very diverse disease group. There is an autoimmune disease affecting nearly every organ in the body. Clinically, they are categorized as either organ-specific (e.g. multiple sclerosis, type I diabetes, Hashimoto's thyroiditis) or systemic, namely affecting multiple organs and tissues (e.g. rheumatoid arthritis, systemic lupus erythematosus, primary Sjögren's syndrome) as reviewed by Davidson and Diamond (2001) and Marrack et al. (2001). A shared feature of autoimmune diseases is though the presence of pathogenic autoantibodies and the contribution of both humoral and cellular immune responses to tissue injury.

2.2.2 Pathogenic *vs.* Non-Pathogenic Autoantibodies

> The repertoire of target autoantigens is a *Wunderkammer* – a collection of curiosities – of molecules with no obvious linking principle.
>
> *Plotz (2003)*

Although counterintuitive, immune cell repertoire contains many self-reactive immune cells and it is an empirical fact that *natural autoantibodies* can be found in the blood serum of healthy individuals. Circulating natural autoantibodies are polyreactive antibodies produced at tightly regulated levels. They have weak to moderate affinity, undergo minimal or no somatic mutation, and are often of IgM, but also of IgG or IgA classes (Avrameas, 1991, George et al., 1997, Grönwall et al., 2012). These low-affinity antibodies, fulfilling the definition of autoantibodies but being also able to react with non-self antigens, in particular target nuclear proteins and DNA. Natural autoantibodies have been shown to have important and physiologically protective roles in several biological processes. This includes, but is not limited to the participation in the fist-line defense against toxins and infectious agents; antigen presentation to T-cells, B-cells and antigen-presenting cells; removal of apoptotic cells; clearance of cellular debris and subcellular particles and clearance of altered or senescent cells or tumors (Coutinho et al., 1995, Shoenfeld and Toubi, 2005). Thus, autoimmunity can be regarded as a context-dependent phenomenon: Self-reactive immune cells may elicit either a protective or a destructive immune reaction, mostly also depending on the tissue context (Schwartz and Cohen, 2000).

Pathogenic autoantibodies leading to autoimmune diseases are believed to be of higher titer in serum. They have often undergone class switching to IgG and somatic mutations and

have higher affinity towards self-antigens (Elkon and Casali, 2008). Two main models have been proposed to explain the production of pathogenic antibodies in autoimmune diseases. One model suggests that pathogenic autoantibodies are the consequence of an uncontrolled polyclonal stimulation of normal auto-reactive B-cells yielding elevated levels of otherwise natural autoantibodies (Klinman and Steinberg, 1987). The other model, which has been acknowledged more (Dighiero and Rose, 1999), suggests that pathological autoantibodies are the consequence of an antigen-driven selection of auto-reactive B-cells, which, under selective pressure of the self-antigen undergo further somatic mutations leading to the production of pathological autoantibodies (Shlomchik et al., 1987).

Certain characteristics have been attributed as common features for the target self-antigens of autoantibodies. As reviewed by Reeves and Satoh (1996) and by Plotz (2003), these include sequence-related and structural properties (e.g. containing coiled-coil structures, repetitive elements, evolutionarily conserved epitopes or having highly charged surfaces); catabolic properties (e.g. being cleaved by nucleases, caspases or granzyme B) or post-translational modifications. Besides, it is also suggested that autoantibodies targeting cell surface antigens (e.g. hormone receptors), membrane-associated antigens, extracellular antigens or circulating antigens have in general a greater potential to be pathogenic autoantibodies than the ones targeting intracellular components, which can cause clinical manifestations only if the antigen would migrate to the cell surface (Naparstek and Plotz, 1993, Racanelli et al., 2011). Some good examples of known target antigens in autoimmune diseases are nuclear antigens (DNA, histones, ribonucleoproteins) in systemic lupus erythematosus; thyroid-stimulating hormone receptor in Graves diseases; citrullinated cyclic peptides in rheumatoid arthritis and myelin basic protein and myelin oligodendrocyte glycoprotein in multiple sclerosis (two extensive lists have been provided by Lernmark (2001) and Hayter and Cook (2012)).

Despite the fact that circulating autoantibodies are not exclusive for autoimmune diseases and that they are found in healthy individuals and can be associated with other clinical conditions (such as cancer, as reviewed by Tan and Zhang (2008) and Zaenker and Ziman (2013)), the detection of circulating autoantibodies has a central role in the diagnosis and classification of autoimmune diseases. The value of autoantibodies as disease biomarkers has been recognized in some autoimmune conditions but it still remains underutilized regarding several other ones, where disease-related target self-antigens are not known yet. As recently reviewed by Maecker et al. (2012) and Georgiou et al. (2014), there are various emerging discovery tools, such as high-throughput next-generation DNA sequencing, which allows to identify the targets of autoantibodies by sequencing the B-cell receptors. In the following chapters and as one of the main subjects of this thesis, I will discuss other alternative high-throughput technologies offering the possibility to profile the autoantibody repertoire in body fluids and to identify autoantibodies as potential biomarkers in the context of diseases with an autoimmune character.

2.3 Antibodies as Affinity Reagents

> All that we seem to have acquired is the potential ability to select from an animal any of the antibodies of his repertoire. It is somewhat like selecting individual dishes out of a very elaborate menu: antibodies "á la carte". But surely our "immunological gourmandizing" cannot be satisfied by the menu that the animals are offering us [...]. A gastronomer [...] cannot be satisfied by selecting what somebody else has cooked. He wants to experiment with new ingredients, new combinations [...], to invent new dishes [...]. I am sure that our next step will be to move from the dining table [...] to the kitchen, where we shall attempt to mess them up!
>
> *Milstein (1981)*

Antibodies play an important role not only *in vivo* but also on the laboratory bench as *affinity reagents*. The first reported application and description of antibodies as a "life-science tool" was by Yalow and Berson (1959) for detection of human insulin in a radioimmunoassay (RIA). Since then, antibodies gradually became the workhorse reagents in life-science research due to their natural ability to bind a wide variety of molecules with high specificity and affinity, as well as their intrinsic stability. As reviewed e.g. by Uhlén (2008), antibodies and other affinity reagents are needed and utilized in a very broad spectrum of applications, such as Western blot, immunocytochemistry, immunohistochemistry and immunoassays, with a purpose of both *in vivo* or *in vitro* protein detection, purification, localization and quantification in body fluids, cells and tissues, as well as for clinical diagnostics and even therapy. Antibodies for such purposes and applications can be generated by three main strategies: Animal immunization, hybridoma technology, or recombinant methods, where the latter allows for generating fragments of antibodies.

As reviewed in detail by Hanly et al. (1995), *polyclonal antibodies* are harvested from the sera of host animals, usually rabbits, mice or goats, which are injected with an *immunogen* of interest. The host animal's immune system produces antibodies as a natural immune response to the injected antigen, which is usually co-administered with an adjuvant as an immune response enhancer to increase the antibody titer in host serum. Polyclonal

antibodies are a mixture of different antibodies against different epitopes of the same immunogen as they originate from a mixture of B-cell clones. In addition to their heterogeneous nature, polyclonal antibodies are also susceptible to "animal-to-animal", namely "batch-to-batch" variations in affinity and cross-reactivity, as polyclonal antibodies obtained from two different host animals would *per se* not be identical. The challenges regarding heterogeneity in antibody quality and sustainability of the supply are addressed by the *hybridoma technology*, which initiated the *monoclonal antibody* revolution since its development and first description in year 1975 by Georges Jean Franz Köhler (1946–1995) and César Milstein (1927–2002). Köhler and Milstein elegantly demonstrated the feasibility of producing immortalized antibody-producing cells by fusing immortal mouse myeloma cells with a single clone B-cells secreting antibodies with an exquisite specificity (Köhler and Milstein, 1975).

A continuous culture of B-cell hybridomas offers in principle a sustainable and reproducible supply of an antibody with pre-defined single specificity. Yet, compared to polyclonal antibodies, monoclonal antibodies are at a disadvantage regarding the required amount of workload and expertise. A polyclonal anti-serum might be obtained within a timeframe of 4-8 weeks, whereas it can take any time from 3-6 months up to a year or longer to produce a new monoclonal antibody (Leenaars and Hendriksen, 2005, Smith, 2004). Polyclonal antibodies, potentially recognizing both continuous and discontinuous epitopes of the same antigen might also be beneficial in certain application contexts, where the detection of an antigen with a monoclonal antibody via a single epitope might be compromised due to conformational changes introduced e.g. by pH or heat (Lipman et al., 2005). Purification of a polyclonal anti-serum using the very same immunogen as a ligand offers the alternative of *mono-specific antibodies*, where antibody species not selective for the immunogen can be excluded while still retaining the multi-epitope nature of the purified polyclonal antibody mixture (Hjelm et al., 2011). The feasibility of producing such affinity-purified mono-specific antibodies on a proteome-scale has been demonstrated in recent years by the Human Protein Atlas project, which I will discuss further in the next section.

Antibody generation through animal immunization does not allow to generate antibodies against toxic molecules or unstable antigens, and most importantly, it does not allow to change or improve the binding properties of an antibody, or to reduce its immunogenicity for a therapeutic application in human (Bradbury et al., 2011, Karu et al., 1995). These and similar needs paved the way for generating antibodies recombinantly, namely without the need for animal immunization (Dübel et al., 2010). The progress in *recombinant antibody* technology was allowed by combinatorial technologies to display antibody libraries constructed by random combination of antibody fragment genes, as well as technologies for *in vitro* screening and selection from such combinatorial libraries. Several platforms are currently in use, where phage display (Clackson et al., 1991, McCafferty et al., 1990) is

the most widely used platform applied for high-throughput antibody generation (Schofield et al., 2007), followed by ribosomal display (Hanes and Plückthun, 1997) and cell surface display on yeast (Boder and Wittrup, 1997) or bacteria surface (Daugherty et al., 1998). As reviewed and compared by Bradbury et al. (2003) and Hoogenboom (2005), these different selection platforms offer different maximum library sizes, selection scopes and compatibility with the antibody fragment formats. The antibody fragment called *single-chain variable fragment* (scFv), which is compatible with all three aforementioned platforms, is a widely utilized recombinant antibody format. An alternative to scFv is the Fab fragment, namely the entire *antigen-binding fragment*. Fab and scFv fragments have also been engineered into diabodies, triabodies, and tetrabodies for increased multi-specificity, as reviewed e.g. by Holliger and Hudson (2005).

Although "natural" polyclonal or monoclonal antibodies are still the most established and widely utilized affinity reagents, there is an expanding range of alternative, namely non-immunoglobulin affinity reagents. These antibody-mimetics mainly include *protein scaffolds* and *nucleic acid aptamers*, as well as small molecules, such as peptides and organics. Examples of small protein scaffolds, which in general demonstrate an engineered stability and specificity (Löfblom et al., 2011), include ankyrin repeats, e.g. designed ankyrin repeat proteins (DARPins, Binz et al. (2003)); fibronectin type III domain (Koide et al., 1998); lipocalins (Skerra, 2000), and affibodies (Nygren, 2008) among others. Scaffolds based on single-stranded nucleic acids, namely DNA and RNA aptamers, were introduced around the same time as protein scaffolds (Ellington and Szostak, 1990, Tuerk and Gold, 1990). Incorporation of modified nucleotides, as well as introduction of selection based on slow dissociation has led to a new generation of aptamers, the so-called slow off-rate modified aptamers (SOMAmers), as recently reviewed by Lollo et al. (2014).

Despite the availability of modern recombinant technologies and a rich spectrum of emerging alternative affinity reagents with versatile characteristics, as well as the realized demand for such renewable reagents for proteome-wide applications, their overall application range is currently more restricted than antibodies. This might be in part due to intellectual property rights, or at least the perception they created, which have in general hampered the dissemination of the recombinant display and selection technologies. It might also be due to the fact that most of these technologies are still in "development phase", aiming currently for more powerful and specialized, rather than simpler and standardized methods (Sidhu, 2012). Although not yet, "in less than 10 years from now, we all might have a desktop robot in our laboratory that is capable of generating a recombinant antibody within a couple of days", as claimed by Dübel et al. (2010). Until then and for the foreseeable future, antibodies will continue to be indispensable affinity reagents needed on the laboratory bench, which in turn calls for efforts to generate antibodies on a proteome-scale, as I will discuss next.

2.3.1 Mission Possible: Proteome-Wide Generation of Affinity Reagents

> The exceptionally impressive throughput in generation of monospecific polyclonals [...] in the ongoing Human Protein Atlas project is an excellent illustration of what can be achieved in affinity proteomics with "classical" methods.
>
> *Stoevesandt and Taussig (2012a)*

Despite the availability of several technologies and strategies for generating affinity reagents, comprehensive, high-quality and uniform sets of affinity reagents are still scarce. This is one of the major reasons behind the sub-optimal use and impact of systematic analysis of protein abundances and expression patterns in body fluids, cells and tissues in "health" and "disease". This motivation and an increasingly systematic characterization of the human genome has lately encouraged the attempts to generate affinity reagents on a large-scale, envisioned to be proteome-scale. As reviewed by Stoevesandt and Taussig (2012b), there are already a handful of initiatives developing and implementing roadmaps for generating and cataloging high-quality and proteome-wide collections of affinity reagents both in Europe and USA. Yet, the scale and coverage to be achieved by these attempts is very closely related to their conceptual starting points and the definition of "proteome-wide".

The two main starting points for generating proteome-wide affinity reagents are through a *gene-centric* or a *protein-centric* approach. As I will describe next in more detail, the gene-centric approach has been implemented by the Human Protein Atlas, where the approximately 20,800 protein-encoding genes in the human genome are the starting points to define each protein target to generate antibodies for. As of the date this thesis was written, the Human Protein Atlas project has achieved a coverage about 80% of the gene-centric proteome and is expected to reach a full coverage of the proteome, as so defined, in 2015. Defining a similar finish-line for a protein-centric approach is though almost impossible due to the enormous diversity of the proteome introduced by alternative splicing, mutation and post-translational modification events.

A proteome-centric approach would in theory provide a coverage for all possible modifications and splice variants for each protein target. Although on a less-inclusive rather than a global scale, this has been successfully exercised within few finished and ongoing projects, where as a derivation of a protein-centric approach, a *function-centric* approach, focusing on protein sets related to particular functions, was chosen. The ongoing EU project AFFINOMICS, for instance, a follow-up project carried out by

the ProteomeBinders consortium (Taussig et al., 2007), focuses currently on systematic production of affinity reagents against a set of 1,000 proteins. This includes all known protein kinases, protein tyrosine phosphatases and SH2 (Src Homology 2) domain containing proteins involved in signal transduction and cell regulation pathways, their known post-translational variants and cancer-related variants. More than 1,000 recombinant affinity reagents targeting twenty SH2-domain containing proteins were generated within a preceding project of AFFINOMICS. Similarly, different types of recombinant affinity reagents were generated against proteins involved in TGF-β and MAPK pathways within the AffinityProteome project, all demonstrating on a pilot-scale the possibility of concerted efforts to systematically generate function- and application-specific affinity reagent sets (Colwill et al., 2011). Similar efforts, as recently reviewed by Stoevesandt and Taussig (2012a) and Taussig et al. (2013), include the ongoing NIH Protein Capture Reagents Program, focusing on human transcription factors. Additional approaches have also been defined and implemented, such as the *chromosome-centric* approach adopted by the chromosome-centric Human Proteome Project to catalog proteins encoded in each chromosome using tools including affinity-reagents (Legrain et al., 2011) and a *disease-centric* approach adopted by the National Cancer Institute (NCI) aiming to produce and characterize monoclonal antibodies against cancer-associated targets (Haab et al., 2006).

As important as the efforts in generating affinity reagents is to build databases and portals to catalog the available affinity reagents from commercial vendors or academic providers, as well as quality control data for each reagent. One such open-access, community-based portal is the Antibodypedia (http://www.antibodypedia.com/), which was originally developed within the Human Protein Atlas (Björling and Uhlén, 2008). Antibodypedia lists, as of the date this thesis was written, over 1.3 million antibodies from 49 providers reviewed based on over 5.5 million experiments and covering gene-products encoded by 19,137 genes (approximately 92% of all human genes). This, and similar efforts, such as the Antibody Registry (http://antibodyregistry.org/) are becoming increasingly convenient tools for researchers allowing to compare and choose the right affinity reagent for the right application among several suppliers.

The Human Protein Atlas

The *Human Protein Atlas* is a large-scale effort and resource initiated in year 2003, which since then has been generating antibodies towards human proteins. The aim of the Human Protein Atlas is to create a repository of minimum two antibodies against a representative product of each protein-coding human gene and with a vision to build a comprehensive "atlas" of the human proteome across cells, tissues and organs (Uhlén et al., 2005).

The starting point of the Human Protein Atlas pipeline is the *in-silico* selection of antigens, the so-called *protein epitope signature tags* (PrESTs) consisting of approximately 50-150

amino acids of each protein target sequence. Using an interactive tool (Lindskog et al., 2005), the PrESTs are chosen based on several criteria such as <60% sequence homology to the rest of the human protein-encoding genes and exclusion of transmembrane regions and signal peptides (Berglund et al., 2008a,b). As originally described by Agaton et al. (2003), upon design of primer pairs and amplification of the selected gene fragments from a human total RNA template pool, the fragments are cloned and expressed in *Escherichia coli* (Tegel et al., 2009, 2010) as recombinant fusion protein fragments with an N-terminal dual affinity tag (His_6-ABP) consisting of a hexahistidyl (His_6) tag allowing for a one-step purification on nickel columns, and an albumin binding protein (ABP) for an increased immunogenicity (Sjölander et al., 1997) and solubility. Upon sequence verification of the purified PrESTs using mass-spectrometry, the PrESTs are utilized first as immunogens for immunization of rabbits and then as ligands to prepare affinity columns for purification. The harvested polyclonal rabbit antisera are purified on the affinity columns prepared with the respective PrESTs utilized for immunization (Nilsson et al., 2005).

The eluted affinity-purified polyclonal antibodies, also referred to as *mono-specific antibodies*, are characterized for their binding specificity using first an antigen array-based and then a Western blot-based analysis. Array batches, each produced with different and random sets 384 PrESTs including the target PrEST of interest are generated in a systematic manner. These arrays are utilized to verify the specificity of the antibodies to the respective PrEST and to check for off-target binding to all the other PrESTs in a high-throughout manner (Nilsson et al., 2005). Antibodies passing these criteria are further analyzed in Western blot using lysates of two human cell lines, depleted blood plasma, as well as whole tissue lysates of tonsil and liver (Älgenäs et al., 2014).

Antibodies passing these check points are then utilized for a subsequent tissue expression profiling by immunohistochemistry analysis (IHC) on tissue microarrays (TMAs). They are also used for protein localization on sub-cellular level by immunofluorescence-based confocal microscopy analysis on cell lines. The TMAs for tissue expression profiling include specimen from 44 different normal human tissue types, including e.g. breast, pancreas, kidney, liver and heart, as well as 20 different cancer tissue types from e.g. melanoma, glioma, lymphoma or breast cancer (Kampf et al., 2005). In parallel, cell microarrays in a TMA format are utilized for IHC analysis on 46 different, well-characterized human cell lines in order to complement for the cell types not present in the TMAs (Andersson et al., 2006). For a subsequent sub-cellular localization analysis, three cell-lines with a different origin (glioblastoma, osteosarcoma and epithelial carcinoma) are utilized to annotate the protein location to sub-cellular organelles (Barbe et al., 2007). All high-resolution images and annotations regarding the tissue expression and sub-cellular localization in all represented tissue and cell types, as well as antigen microarray results confirming antigen specificity and Western blot images confirming the presence of bands of predicted molecular weight, are made publicly available through the open-access portal of Human Protein Atlas

(http://www.proteinatlas.org/) (Uhlén et al., 2010). Recently, the antibody-based tissue expression profiles on Human Protein Atlas portal have been also supplemented with transcript profiles, where the transcriptome of 27 different human organs and tissues with specialized functions were profiled by RNA-seq (Fagerberg et al., 2014).

The current version of the Human Protein Atlas, released in December 2013, contains 21,984 antibodies targeting proteins from 16,621 human genes, corresponding to approximately 80% of the human protein-coding genes. It represents the paradigm for the hypothesis-free, gene-centric approach, which demonstrates the possibility to generate affinity reagents with a broad coverage of the human proteome. The publicly available portal of the Human Protein Atlas offers several types of valuable tissue expression and sub-cellular localization information, creating new or extending the existing knowledge about these protein targets (Ponten et al., 2011). Equally importantly, the availability of such a comprehensive set of antibodies, as well as antigens (PrESTs) within the Human Protein Atlas offers a unique possibility to explore protein profiles and autoantibody repertoires in body fluids. The aim of the latter approach is to identify potential biomarker candidates in a broad spectrum of diseases by developing and using high-throughput approaches for analysis of body fluid samples, which refers to the main concept behind the work presented in this thesis.

3 Mining the Proteome for Biomarkers

Exploring the human proteome for biomarkers is akin to searching for a needle in a haystack; an analogy often made to express the challenging nature of the biomarker discovery efforts. In this chapter, I will first briefly discuss this challenging aspect of the protein biomarker discovery process. I will then continue the discussion with the biological challenges related to protein biomarker discovery, namely the complex nature of the proteomes of body fluids, such as blood, which become the haystack to dig in in order to identify disease-related biomarkers. Within this discussion, I will also touch upon the importance of standardization of procedures in collection and handling of biological fluid specimen and underline the role of biobanking efforts as an indispensable part of the protein biomarker discovery field. Finally, I will provide an overview of the main strategies which can be utilized to mine the proteome in the quest towards protein biomarkers.

3.1 Biomarkers *vs.* Potential Biomarker Candidates

> The pessimist sees only the tunnel; the optimist sees the light at the end of the tunnel; the realist sees the tunnel and the light - and the next tunnel.
>
> *Sydney J. Harris*

Despite the widespread interest of researchers and diagnostic companies, as well as of granting agencies (Ptolemy and Rifai, 2010), the expectations that new protein biomarkers could be identified and easily implemented into clinical practice has not been fulfilled during the last decade (Anderson and Anderson, 2002, Gutman and Kessler, 2006). Regarding cancer, for instance, the few well-known cancer biomarkers which are already used in the clinic despite their limitations, such as the prostate specific antigen (PSA) for prostate cancer, could not be supported with several new, let alone better ones (Diamandis, 2014). Although this initially created a premature frustration, soon later it has nurtured several constructive discussions identifying the current limitations and pitfalls in the quest towards clinically useful protein biomarkers (Ioannidis, 2013, Kern, 2012, Pavlou et al.,

2013, Ransohoff, 2005), as well as suggesting strategies and guidelines to address them (Füzery et al., 2013, Kulasingam and Diamandis, 2008, Mischak et al., 2007, Paulovich et al., 2008, Rifai et al., 2006, Zolg, 2006).

The number of FDA-cleared or -approved proteins which are assayed in clinical settings is approximately 200 (Anderson, 2010). The number of reports containing the term *protein biomarker* in their abstract and published between 1996 – the year when the term *proteome* appeared in a report for the first time (Wilkins et al., 1996) – and 2013 is 433,662*. In recent years, this trivial comparison has become a way to illustrate the imbalance between the great efforts put into the quest for protein biomarkers and the small number of biomarkers which made it to the clinic. This is indeed a reality, however, instead of only creating doubts and disbelief about the value of all biomarker discovery efforts, this comparison should increase the awareness of two other important aspects I would like to underline.

The classical biomarker discovery pipeline, as outlined for instance by Rifai et al. (2006), starts with the identification of one or mostly several *potential biomarker candidates* in representative sample collections of preferably blood, but also tissue or proximal fluids. Next, these candidates need to be technically verified and prioritized further based on criteria such as presence in blood. Assays are then to be developed further to establish analytical performance characteristics in extended sample collections. Importantly, this step might require also the process of generating suitable binders, such as monoclonal antibodies, in order to detect these *biomarker candidates* in a clinical assay format. Upon the accomplishment of these pre-clinical feasibility steps, candidates need to be validated in an appropriate context and in –ideally– thousands of samples. This latter step potentially requires to join forces with commercial and diagnostic partners. If successful, these clinically validated biomarker assays might become a part of the clinical assay portfolio upon their filing to and clearance and approval by FDA. This multi-stage process starting with the initial discovery of a biomarker candidate and ending with an implementation of it into clinical practice takes a great amount of time and it requires extensive and long-term collaboration between researchers and commercial partners. Following the example of PSA again, the first immunoassay detecting PSA in serum was reported in 1980 (Kuriyama et al., 1980, Papsidero et al., 1980), yet it was not before 1986 when it acquired FDA clearance for monitoring and not before 1994 when it acquired FDA approval for prostate cancer screening. It should therefore be noted that the biomarker discovery and development path is most likely not a linear one.

As indicated by the title of this section and by the nomenclature used in the previous paragraph, there is also a need to realize the general overuse of the term *biomarker*

*This number has been retrieved from the PubMed database by querying *((((protein OR proteome OR proteomic OR proteomics[Title/Abstract])) AND (biomarker OR biomarkers[Title/Abstract])) AND ("1996"[Date - Publication] : "2013"[Date - Publication]))*.

in publications. This unintentional misuse, causing an accumulation of "biomarker discoveries" in the literature, is partially responsible for the judgement of the value of biomarker discovery efforts as questionable. This could be addressed by maturing the nomenclature and using the term *potential biomarker candidates* to describe proteins which are identified to be differentially abundant in diseased conditions and as outcome of initial discovery efforts analyzing representative number of samples. The term *biomarker candidate* can be used to describe those targets which can be verified further in extended sample collections using orthogonal technologies and/or methods. Clearly, a biomarker candidate is still several steps away from becoming a *biomarker* to be used in the clinic. However, the path leading to these biomarkers starts inevitably by identification of potential candidates. Considering the continuously decreasing chance of success of a potential biomarker candidate along the pipeline, identification and verification of as many candidates as possible is an important aspect, as long as the recent lessons learned in the biomarker discovery field are taken into consideration.

In the later sections and as the main theme of this thesis, I will discuss some of the available tools, different technologies, strategies and methodologies enabling identification and verification of protein biomarker candidates, that might feed the pipeline for the translation into clinically useful biomarker tests. However, before shifting the focus to the tools, technology and their applications, I first would like to discuss the natural sample sources of the human body, offering possibilities to access and explore the proteome in order to identify potential biomarker candidates.

3.2 Sources to Mine for Biomarkers

> All of the vital mechanisms, however varied they may be, have always one goal, to maintain the uniformity of the conditions of life in the internal environment [the milieu intérieur] [...] The stability of the internal environment is the condition for the free and independent life.
>
> *Bernard (1974)*

Fluids of the body constitute approximately 60% of the adult human body weight and they are life's multi-purpose environments. Around two-thirds of the human body fluid is the *intracellular* and one-third is the *extracellular fluid*. The extracellular fluid compartment can be divided further into the following "virtual" but physiologically meaningful compartments: The *interstitial compartment*, the *transcellular compartment*

and the *intravascular compartment*. The interstitial compartment surrounds and provides a micro-environment for all the cells in tissues and provides the link between the intracellular and intravascular compartment. Blood plasma, of an approximate volume of 3.2 liters, forms the intravascular compartment, whereas the transcellular compartment, of an approximate total volume of 1 liter, represents the other body fluids contained within epithelial lined spaces, such as cerebrospinal, synovial, intra-ocular, peritoneal fluids and digestive secretions (Ahn and Simpson, 2007, Guyton and Hall, 2006).

Almost 150 years ago, Claude Bernard (1813–1878) elucidated the concept of *milieu intérieur*, the internal environment, envisioned to surround and bath all the anatomical elements of tissues. He also proposed that the milieu intérieur would include the liquid portion of blood, namely the plasma. Bernard's notion about milieu intérieur and the importance of maintaining its constancy still holds today. In fact, exploring the proteomes of the fluids of the human body in order to identify the perturbations of this milieu intérieur in disease state is one of the main ambitions within the field of life sciences.

In the quest towards potential biomarker candidates, body fluids can be considered as the primary sample source utilized by the contemporary protein profiling approaches. However, the domain to search for potential biomarker candidates is certainly not limited to body fluids. Samples derived from cells and tissues of the human body, as well as of animal models mimicking human diseases, represent equally valuable sources for biomarker discovery. In fact, human tissue-derived samples collected by means of surgical biopsies would offer the most direct insight into a disease affecting the given tissue. Establishment of methods such as laser capture microdissection (Emmert-Buck et al., 1996), allowing for isolation of diseased and non-diseased proximal tissue of the same but often heterogenous tissue environment, render tissues as valuable sources for biomarker discovery, especially in the field of cancer proteomics. However, body fluids in general offer certain advantages over tissue-derived samples due to their accessibility, relatively less cost of retrieval and possibility for multiple sampling due to relatively minor invasiveness of the sampling procedures. In the following discussion, I will restrict the frame to body fluids as sources to mine for potential biomarker candidates and the investigations carried out within this thesis work have also involved body fluid samples as their starting points.

3.2.1 Blood

> Blood is a fluid of rarest quality.
>
> *Johann Wolfgang von Goethe*

The human body contains almost 100,000 kilometers of veins, arteries and capillaries (Wilson, 2000), where almost 6 liters of blood continuously flows throughout the body.

Over the course of a human's life time, the heart beats over 2.5 billion times and pumps almost 250 million liters of blood (Spellman, 2007), which would fill the interior space of a skyscraper. Blood has several physiological functions within the body: It transports the cells of the immune system to sites of infection; acts as a stabilizing medium to maintain bodily pH, temperature and electrolyte balance and acts as a medium to supply cells with oxygen, nutrients, hormones and enzymes and to handle their waste products and carbon dioxide (Guyton and Hall, 2006). Blood is the macro-environment of the body, continuously accessible for all different tissue micro-environments. It is therefore in an intimate relation with and under continuous exposure from cells aside the blood stream. Indeed, no cell in the body is more than 100 µm away from a capillary of the circulation system (Raven et al., 2010). If "eyes are the windows to the soul", then blood is the window to the body's state of health, containing signatures of every process taking place at any part of the body.

The liquid portion of blood is called *plasma* and it forms around 55% of the total blood volume. The cellular components suspended in the plasma include *red blood cells* (erythrocytes), *white blood cells* (leukocytes) and *platelets* (thrombocytes). The major function of red blood cells is to transport hemoglobin, which in turn carries oxygen from the lungs to the tissues. The white blood cells include lymphocytes of the adaptive immunity and basophils, eosinophils, monocytes and neutrophils acting mainly as phagocytes within the cellular arm of innate immunity (Figure 2.1). The platelets play a very important role in hemostasis, namely prevention of blood loss. They form platelet plugs for closing of minute vascular raptures occurring several thousand times each day and they also play an important role in the blood clotting mechanism, which involves a complex cascade of chemical reactions involving more than a dozen blood coagulation factors. As a net outcome in response to a blood vessel rupture, a complex called prothrombin activator is formed, which catalyzes the conversion of prothrombin into thrombin, which in turn catalyzes the conversion of fibrinogen into fibrin fibers that enmesh platelets, blood cells, and plasma to form the clot (Guyton and Hall, 2006).

Due to its intimate relation with all parts of the body, plasma has an overwhelmingly diverse and dynamic protein content. This intrinsically rich but complex nature of the plasma proteome, however, is also what renders the exploration of the plasma proteome a big analytical challenge. The major issue regarding the nature of plasma proteome is its large concentration range, which is often quoted to be over ten orders of magnitude (Anderson and Anderson, 2002). A handful of proteins, which represent less than 0.1% of the total number of plasma proteins constitute almost 99% of the bulk mass of proteins. Albumin, at a concentration of 35-50 mg/ml, alone constitutes around 60% of the total plasma protein content of 60-80 mg/ml, followed by immunoglobulins at a concentration of 5-18 mg/ml and other abundant proteins such as transferrin and fibrinogen (Figure 3.1). These abundant proteins mask the remaining content, which is very interesting and rich

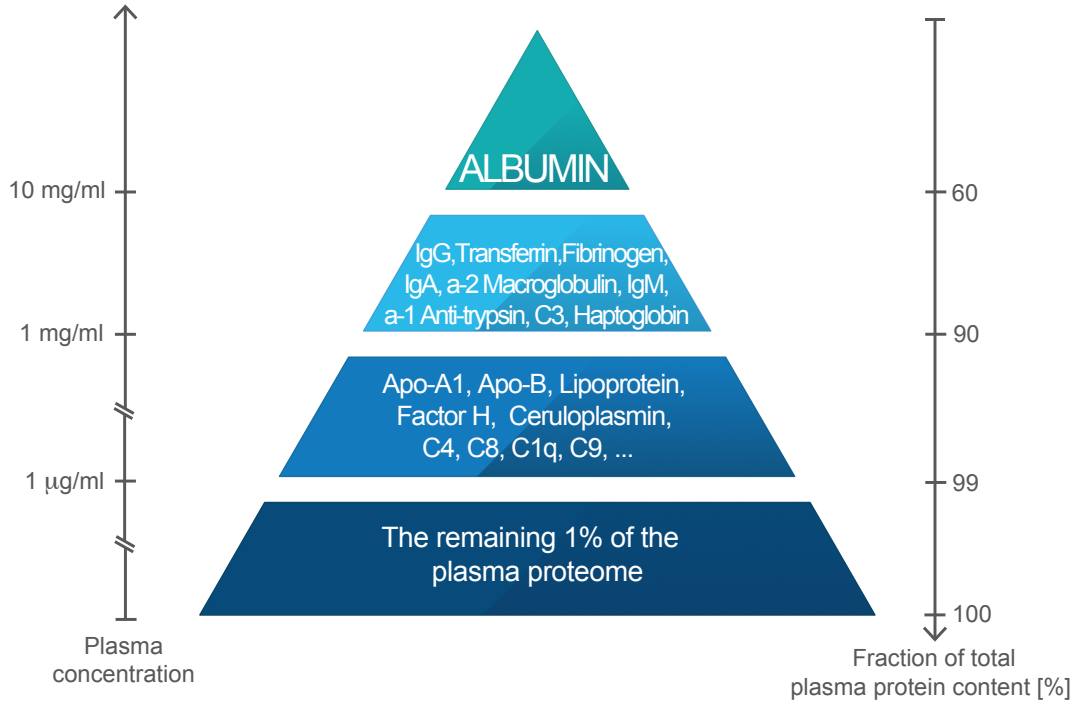

Figure 3.1: A simplified overview of the most abundant plasma proteins, accounting for 99% of the content of plasma proteome.
The 10 most abundant "classical" proteins in the plasma constitute almost 90% and the 22 most abundant proteins constitute almost 99% of the plasma proteome. The remaining 1% is both the richest and the most challenging portion of the plasma proteome to explore, containing e.g. all tissue leakage products and interleukins and cytokines (Anderson and Anderson, 2002).

in number of proteins, including e.g. tissue leakage products or interleukins, but all at a much lower concentration range, such as interleukin 6 at a concentration around 5 pg/ml (Anderson and Anderson, 2002, Schiess et al., 2009).

As discussed in the most frequently cited article in plasma proteomics by Anderson and Anderson (2002) and other subsequent work (Anderson et al., 2004b, Shen et al., 2005), the plasma proteome contains proteins of various functional groups. The first group to be mentioned is certainly the "classical plasma proteins" including e.g. albumin and transferrin, which are mainly secreted by liver. These proteins function as "molecular sponges", binding to and transporting other proteins of lower molecular weight, as well as lipids, metals and hormones. As I have discussed in the previous chapter, immunoglobulins are the key mediators of immune response and human plasma contains an estimated number of 10 million immunoglobulin species of unique sequence. In line with this, it also contains the proteins originating from pathogens and infectious organisms. Another functional group consists of hormones - the "long distance" receptor ligands, and cytokines - the "short distance" receptor ligands. There are also "temporary passenger" proteins, traveling through plasma temporarily from the site of synthesis or secretion to the site of

primary function, such as lysosomal proteins. A very important group of proteins from a biomarker discovery perspective are the secreted proteins from tumors and other diseased tissues, which are otherwise not present in plasma. An equally important group with a biomarker potential is formed by tissue leakage products, such as cardiac troponins, which are released into plasma upon cell damage or death.

The number and identity of the proteins at the top of the plasma proteome pyramid (Figure 3.1), namely the classical plasma proteins are well-defined. However, the top of the plasma proteome pyramid resembles the tip of a giant iceberg, where there is no accepted estimate of how many protein species can be found hidden below it. The initial attempts to characterize the human plasma proteome using two-dimensional electrophoresis date back to 1970′s (Anderson and Anderson, 1977), yet still in the beginning of 1990′s only dozens of plasma proteins could be identified (Anderson and Anderson, 1991). Soon after, advancements in the mass spectrometric methods contributed significantly to this ambition and characterization of the plasma proteome has since then become a marathon without a finish line, namely a definite number of protein species present in human plasma. In year 2002, a Human Plasma Proteome Project (HPPP) was initiated as part of the Human Proteome Organization (HUPO) and a first multi-center approach (Omenn, 2004) initially identified 3,020 protein species in plasma (Omenn et al., 2005). This list was later condensed to a high-confidence list of 889 plasma proteins by States et al. (2006) and of 697 proteins excluding immunoglobulins by Schenk et al. (2008). Meanwhile, similar efforts have been directed towards identification of post-translational modifications of plasma proteins (Carrascal et al., 2010, Zhou et al., 2009). As part of the second phase of HPPP, in 2011 Farrah et al. (2011) reported an extended, non-redundant list of 1,929 plasma proteins after a meta-analysis of 91 experiments from various laboratories. Very recently, Farrah et al. (2014) shared the news that this list has been extended even further to 3,553 non-redundant proteins as outcome of a meta-analysis of 127 experiments, which is currently the most updated and largest non-redundant list of plasma proteins. However, as wisely expressed by Anderson (2014), "there are plenty of plasma proteins, and we will never see the "last" splice variant, post-translational modification or mutation".

Blood as a sample source can be processed as *plasma* upon withdrawal by venipuncture in the presence of anti-coagulants, followed by centrifugation to remove cellular elements. The most commonly used anti-coagulants include ethylenediamine tetraacetic acid (EDTA), heparin and sodium citrate. Both EDTA and sodium citrate prevent coagulation by chelating calcium ions, which play a central role in the coagulation cascade, whereas heparin prevents coagulation by activating anti-thrombin. In the absence of an anti-coagulant agent, blood is prepared as *serum* upon removal of the fibrin clot by centrifugation (Jambunathan and Galande, 2014). The absence of the large portion of fibrinogen content in serum, as well as the absence of several coagulation factors, makes plasma and serum qualitatively different from each other. Regardless of the method

utilized for a downstream protein profiling analysis, sample preparation type systematically affects the detectability of the blood proteins, as previously demonstrated in several studies (Ayache et al., 2006b, Haab et al., 2005, Hsieh et al., 2006, Qundos et al., 2013, Schwenk et al., 2010a), as well as in **Article V** of this thesis work. Furthermore, detectability of certain protein families, such as metalloproteinases, is in particular compromised in plasma preparations with chelating agents, such as EDTA (Mannello, 2008, O'Neal et al., 2014). Thus, although EDTA-plasma has been generally suggested as a preferred sample preparation type (Omenn, 2007, Rai et al., 2005), the endpoint measurement of interest should also be taken into consideration regarding the decision for the most suitable sample preparation type. Most importantly, the inherent differences in the protoemic content of different blood preparations makes them generally incompatible sample sources to be interchangeably utilized within the same study. It is therefore important to use a consistent sample preparation type for both descriptive but especially for comparative protein profiling studies.

In this section, I have so far focused the discussion on the cell-free portion of blood, namely to plasma or serum, as a target sample source for identification of biomarker candidates. However, I would like to underline that the proteomic analysis of red blood cells, white blood cells and platelets offers an additional potential to dissect the mechanisms involved in a wide spectrum of hematologic diseases, as reviewed by D'Alessandro et al. (2010), Liumbruno et al. (2010) and Prudent et al. (2011). I would like to provide now a brief discussion about other fluids of the human body, which provide alternative sample sources for biomarker discovery purposes.

3.2.2 Proximal Body Fluids

> The body of man has in itself blood, phlegm, yellow bile and black bile; these make up the nature of this body, and through these he feels pain or enjoys health. Now he enjoys the most perfect health when these elements are duly proportioned to one another in respect of compounding, power and bulk, and when they are perfectly mingled.
>
> *Hippocrates*

Proximal body fluids are secreted by or gather in direct vicinity of the tissue of interest affected by a disease. As summarized in Figure 3.2, there are several types of proximal fluids in the human body. While sampling of naturally secreted or excreted body fluids, such as urine, stool, seminal fluid, tear and saliva, is non-invasive, collection of others,

such as CSF, pancreatic juice or bile are highly-invasive and require endoscopic or surgical methods. Compared to sampling of blood, which is often regarded as a minimally invasive procedure, sampling procedures for synovial fluid, bronchoalveolar lavage fluid, ascites fluid and amniotic fluid can also be considered as invasive (Ahn and Simpson, 2007).

Proteins, either secreted or leaking from the diseased tissue microenvironment are at a higher regional concentration in proximal fluids, which otherwise are diluted in the systemic blood. This has been demonstrated in the case of cancer antigen 125 (CA125), where average levels of CA125 in ascites and cyst fluid were 27-fold and 64-fold higher than the average of levels in serum of ovarian cancer patients (Sedlaczek et al., 2002). The enriched concentration of potentially disease-related proteins in the proximal fluids therefore renders these as presumably rich sources for identification of potential protein biomarker candidates.

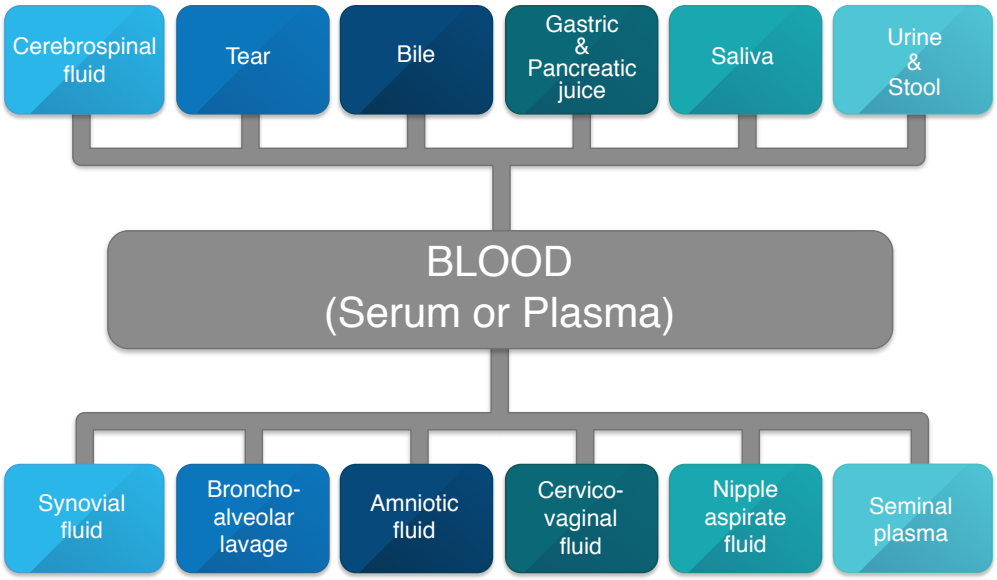

Figure 3.2: Proximal fluids of the human body.
Blood is the systemic fluid, circulating to serve the body's cells, tissues and organs. Proximal fluids are located close to the various tissues or organs of the body, thus representing the closest environment to the site of diseases and malignancies and providing a potentially rich source of potential disease biomarker candidates.

Characterization of the proteome of each proximal fluid would enable a unique advantage to study the related diseases. In line with this, there have been several attempts to characterize the proteome of proximal fluids, including pancreatic juice and gastric juice for gastrointestinal cancer types (reviewed by Farina (2014)); synovial fluid for joint-related diseases such as rheumatoid arthritis and osteoarthritis (Cretu et al., 2013, Gibson and Rooney, 2007); bronchoalveolar lavage fluid for interstitial lung diseases (Magi et al., 2006); seminal plasma for disorders of the male reproductive system and prostate cancer

(Drabovich et al., 2014); tears for eye diseases (Zhou et al., 2012b); stool for diseases of colon and rectum (de Wit et al., 2013); saliva for oral cancer types (Amado et al., 2013) and CSF for neurological diseases and central nervous system tumors (Roche et al., 2008). In the last decade, these and several other similar efforts generated extensive lists of potential biomarker candidates for each proximal fluid and in the context of different diseases. While providing a catalog of these protein or autoantibody candidate biomarkers falls beyond the scope of this thesis, there are comprehensive reviews listing the growing number of such biomarker candidates and highlighting the wide range of information generated by proteomic approaches to explore proximal fluids (Hu et al., 2006, Teng et al., 2010, Thongboonkerd, 2007).

As demonstrated by Yan et al. (2009) for saliva, by Schutzer et al. (2010) for CSF, and by Farrah et al. (2014) for urine, systematic comparisons of the proteome content of plasma or serum with proximal fluids offer an additional possibility to identify proteins which are specific for a given proximal fluid and thus might exhibit an increased potential as disease-related protein biomarkers. Well-characterized collections of *matched samples*, namely proximal fluid and blood samples, collected from same individuals during the same visit to the clinic, provide a unique possibility. Co-profiling of these sample collections on high-throughput platforms allows to co-evaluate the biomarker potential of identified proteins in both body fluids, as also demonstrated in **Article IV** of this thesis work.

Various biochemical properties and the proteomic content and total protein concentration is unique to each body fluid. The pH, for instance, ranges from 1.7 in gastric juice to 8.2 in bile, and the average total protein concentration ranges from 10-250 µg/ml in nipple aspirate fluid (Klein et al., 2001) to 60-80 mg/ml in serum. Regardless of the utilized analytical platform, this requires the chosen tools and protocols to be adjusted for the analysis of each proximal fluid. This is not only for descriptive proteome characterization purposes, but more importantly, also for biomarker discovery-oriented comparative protein profiling studies in sample collections of diseased and non-diseased populations.

As in the case of CSF, the related organ such as the brain is otherwise very difficult for tissue sampling. This renders proximal fluids as valuable sources for biomarker discovery. However, in case of such proximal fluids collected by invasive or highly-invasive sampling procedures, the scarcity of sample material especially from healthy individuals might pose a challenge when building patient and matched control groups of reasonable sample sizes. This applies for discovery but especially for verification purposes and might eventually disqualify such body fluid types as a source for biomarker detection. This consideration underlines the general fact that the choice of the body fluid, namely either blood or a proximal fluid, is an important decision to be made when designing biomarker discovery studies. Each body fluid offers particular advantages while simultaneously posing its own set of challenges regarding study and experimental design.

As I will briefly touch upon next, attempts to standardize sampling and storage protocols and to study the impact of pre-analytical factors on the downstream analyses of protein content of a majority of proximal fluids have so far not been reported as detailed as for plasma and serum. Thus, standardized protocols for sample collection, handling and storage, as well as criteria to assess the integrity of each biospecimen type will allow to gain full benefit from proximal fluids within the biomarker discovery pipeline.

3.2.3 Standardization and Banking of Body Fluid Collections

> Biobanking: old activity or young discipline?
>
> *Morente et al. (2008)*

As reviewed in several reports over the recent years, variation in pre-analytical sample conditions might have an underestimated impact on the downstream analysis of body fluid collections for both characterization and biomarker discovery purposes (Ferguson et al., 2007, Jackson and Banks, 2010, Rai and Vitzthum, 2006, Schrohl et al., 2008). The immune response profiles against viral antigens (Neumann and Bonistalli, 2009) or self-antigens (Jarius and Wildemann, 2011, Zhao et al., 2012) in blood-derived samples have been demonstrated to be robust, however protein profiles are *per se* more susceptible to variations in sample processing or storage conditions. Although parameters such as delay time between centrifugation and freezing have been shown not to affect the profiles for a majority of proteins in plasma or serum samples (Hsieh et al., 2006, Ostroff et al., 2010, Qundos et al., 2013), levels of particular protein families, such as growth factors, chemokines and interleukins are indeed affected by variations in pre-analytical sample conditions (Ayache et al., 2006a). In this regard, the so-called *biobanks* play a cardinal role to assure standardization of samples and to control and track the possible pre-analytical variability.

Biobanks operate in adherence to standard operating procedures, so that procedures and conditions are consistent throughout the pre-analytical stages of sample collection and processing. This includes sample withdrawal procedures, time delay between withdrawal and processing, tube types, centrifugation speed and time, temperature at processing, as well as storage temperature and other related conditions (Jackson and Banks, 2010). Furthermore, contemporary biobank infrastructures involving automated liquid handling instruments allow for a direct aliquotation of samples upon sample collection, allowing to avoid repeated freeze-thaw cycles as much as possible, which otherwise might also compromise sample integrity (Mitchell et al., 2005, Rai et al., 2005).

Biobanks are well-organized resources of biological samples from a population or a subset of a population and their associated metadata, and they make these resources accessible

to scientific investigations. Although the concept of biobank is almost two decades old, biobanking can be considered still as a young discipline, continuously evolving in line with the development of new technologies and new research questions and demands. It is estimated that there are globally over 1,500 biobank initiatives, where the majority (70%) of the biobanks are cancer-oriented (Puchois et al., 2013). There are also several initiatives and organizations, such as Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) sponsored by the European Commission (Yuille et al., 2007) or the International Society for Biological and Environmental Repositories (ISBER), which are devoted to mature the biobanking discipline by disseminating reporting guidelines and best-practice reference documents (Marko-Varga et al., 2012, Wichmann et al., 2011).

Biobanks can range in capacity, infrastructure, extend of expertise and operational model: They can be within single local units, such as an academic or hospital laboratory; or across organizations on multiple sites on a national or international level initiated for specific projects or grant programs; or they can be in the form of very large public repositories (LaBaer, 2012, Marko-Varga et al., 2012), where the cost to establish the biobank but also the quality and utility of samples is dictated by the biobank type and capacity. Similarly, in terms of their scope, biobanks can range from population-based archives to disease-oriented epidemiological or disease-oriented general biobanks (e.g. tumor biobanks or brain biobanks) (Riegman et al., 2008). Sample collections of biobanks can be categorized further as cohort study *vs.* case/control study collections or prospective *vs.* retrospective collections. Cohort study collections are usually longitudinal samples collected prospectively, where case and control samples are collected under relatively more identical conditions. These offer an opportunity to evaluate samples prior to clinical presentation. Such collections though require large populations, take more time and are more expensive to establish than case/control collections. The latter are comparatively easier to establish, but in turn, they are more prone to possible biases in sample collection and processing (LaBaer, 2012). Certainly, depending on the research question and the context of disease, both types of collections are needed and equally valuable for the biomarker discovery efforts.

The concept of biobanking certainly does not only involve standardized collection, processing and storage of samples but also an equally standardized and organized collection of metadata from the sample donors. Such metadata should include information such as age, gender, weight, height, menopausal status, concomitant infection, clinical and treatment history, medications and life-style factors such as exercising and smoking. Furthermore, biobanks do not only serve to the purpose of establishing collections of well-characterized samples of standardized quality; especially for certain diseases, biobanking efforts are the only enabling factor allowing to carry out studies involving statistically powerful number of samples. Biomarker discovery efforts in rare diseases, for instance, affecting one person in every several thousands or millions, suffer by the

scarcity of enough number of samples. As demonstrated in **Article V** of this thesis work, such diseases call even for multi-center biobank collections on an international level and a wide geographic span. Similarly, proximal fluids collected by invasive procedures require collaborative and multi-center biobanking efforts to assemble sample collections of reasonable sizes for biomarker discovery studies. These considerations, in turn, underline the need to adopt internationally accepted standards to minimize variation between biobanking activities, as well as during transport to the analysis site. Adherence to such standards and the increasing awareness that the outcome of biomarker discovery efforts are directly dependent on the sample quality will decrease the gap between the identification of potential biomarker candidates and their intended implementation in clinical settings.

3.3 Profiling of Proteins and Proteomes

> When does "use of antibodies" become "affinity proteomics"?
>
> *Stoevesandt and Taussig (2012a)*

Proteomics is an umbrella term covering systematic studies that investigate the various aspects of the *proteome*. These include localizing proteins on a tissue or at a sub-cellular level, mapping of protein-protein interactions, identification of post-translational modifications and relative or absolute quantification of protein composition and alterations thereof, namely profiling the differential expression in health and disease.

Traditionally, proteomics has been divided into two methodological categories: *gel-based proteomics* and *gel-free proteomics*. This was mainly due to the impact of *two-dimensional polyacrylamide gel electrophoresis* (2D-PAGE, or shortly 2-DE), which allows to separate proteins in complex mixtures first by their differences in isoelectric point and then in molecular weight. Indeed, the concept of analyzing the entire set of proteins being produced by a cell arose and was first demonstrated on 2-DE by O'Farrell (1975) and Klose (1975). As recently reviewed by Oliveira et al. (2014), 2-DE and its variant 2-DIGE allowing for a differential gel electrophoresis via fluorescent labeling (Unlü et al., 1997) had a central role within proteomics. However, the improvements in *mass spectrometry*-based methods led to the retirement of 2-DE as a mainstream proteomics tool and replaced it. Although the investigations carried out within the frame of this thesis work have not directly employed mass spectrometry-based methods, in this section I first would like to provide a discussion regarding mass spectrometry-based protein profiling approaches.

The birth of *mass spectrometer* goes back to the year 1897, when Sir Joseph John Thomson (1856 –1940) developed a cathode ray tube with an electric field, which he used to measure atomic weight of elements (Thomson, 1897). In the following era, during

and after the World War II, mass spectrometers were used mainly in the petroleum industry to characterize jet fuels, crude oils and gasoline (Yates, 2011). The attempts to use mass spectrometers to characterize peptides and proteins started at the end of 1950′s by Biemann et al. (1959). During 1970′s and 1980′s, mass spectrometers evolved rapidly with the development of the *tandem mass spectrometry* approach by Hunt et al. (1981) and the introduction of two important ionization methods suited for biomolecules, namely *electrospray ionization* (ESI) by Fenn et al. (1989) and *matrix-assisted laser desorption ionization* (MALDI) by Karas and Hillenkamp (1988). Since then, the menu of mass spectrometers has been enriched, with a current availability of several hybrid instruments combining different mass analyzers, such as quadrupole/time-of-flights, linear-ion-trap/Orbitraps or quadrupole/Orbitraps, where the engineering goals regarding resolution, scanning speed and mass accuracy have been pushed a step further by each release of a new instrument.

To a layperson, a mass spectrometer can be described as a big and fancy balance to weigh ions. In principle, it indeed measures the mass-to-charge (m/z) ratios of ions, from which molecular weight can be determined. Regardless of its type and mode, a mass spectrometer consists of three fundamental components: An ionization source, where molecules are converted to gas-phase ions, mostly by MALDI or ESI; a mass analyzer, where ions are separated by their m/z values via electric and/or magnetic fields; and a detection system, where the separated ions and the abundance of each species with a particular m/z ratio are detected (de Hoffmann and Vincent Stroobant, 2007).

As reviewed by Steen and Mann (2004) and Domon and Aebersold (2010), a typical tandem mass spectrometry experiment for a global, or *shotgun* protein profiling begins with digestion of the protein mixture into a peptide mixture by a sequence-specific protease, such as trypsin. The generated peptide mixture is injected to a capillary liquid chromatography (LC) column, where peptide species elute e.g. in order of their hydrophobicity and subsequently get ionized e.g. by ESI at the end of the capillary column. Electrosprayed peptide ions enter the mass analyzer, where they are separated based on their m/z ratio. The isolated precursor ions in the peptide ion spectrum are fragmented further into product ions to generate a fragment ion spectrum. Finally, the peptide-sequence information contained in the fragment ion spectrum is searched against databases where the experimental fragment ion spectra is matched to *in-silico* predicted spectra to reveal the protein identity, as well as its post-translational modifications (Mann and Jensen, 2003). This described approach, also known as the *bottom-up approach* is widely implemented within proteomics and allows for a global exploration of proteins present in biological samples. It also enables for a differential analysis across samples from different biological conditions, if either a label-free or a label-based peptide quantification is performed. The latter includes isotope-coded affinity tagging (ICAT, Gygi et al. (1999)), stable-isotope labeling by amino acids (SILAC, Ong et al. (2002)) or isobaric tags for relative and absolute

quantification (iTRAQ, Ross et al. (2004)). Shotgun mass spectrometry is a very powerful explorative methodology. For instance, a vast majority of the studies I have referred to in the previous sections 3.2.1 and 3.2.2 utilized this approach for characterization of the protein content of plasma and other body fluids. A more inclusive overview of the diverse application range of the shotgun approach, as done e.g. by Cravatt et al. (2007), Walther and Mann (2010) and Mann et al. (2013), goes beyond the scope of this thesis. Yet, two very recent independent studies by Wilhelm et al. (2014) and Kim et al. (2014), where a first and mass spectrometry-based drafts of the human proteome has been reported, should highlight the importance of this methodology within the field of proteomics.

In *shotgun mass spectrometry*, several approaches can be adopted to increase the coverage of the proteome, thus to increase the number of protein identifications in complex samples, such as blood plasma. As recently reviewed by Pernemalm and Lehtiö (2014), these approaches include depletion of high-abundant proteins and pre-fractionation of the samples, either on protein level, such as in gel-based size separation (Lundby and Olsen, 2011), or on peptide level, such as in multi-step fractionation (Washburn et al., 2001) or by isoelectic focusing on immobilized pH gradient gels (Pernemalm et al., 2009). Nevertheless, shotgun approaches offer an analytical sensitivity often reported in the µg/ml range. *Targeted mass spectrometry* methods, mainly multiple reaction monitoring (MRM), or emerging alternatives such as SWATH (Gillet et al., 2012), offer a possibility for a more sensitive analysis, with reported detection sensitivities ranging between 10-1000 ng/ml for plasma (Boja and Rodriguez, 2012, Gallien et al., 2011).

As reviewed by Weiß et al. (2014), the advantage of combining mass spectrometry with affinity-reagent based enrichment, either on protein level, as demonstrated by Nicol et al. (2008) or on peptide level has been recognized. The SISCAPA method (Anderson et al., 2004a), for instance, offers an affinity enrichment prior to MRM analysis by employing antibodies specific for peptides from the proteins of interest. Wingren et al. (2009) and Poetz et al. (2009) alternatively proposed and later demonstrated the utility of multi-specific affinity reagents generated against short peptide motifs shared by groups of proteins, which can be similarly applied for a more generic affinity-enrichment (Hoeppe et al., 2011, Olsson et al., 2011, Volk et al., 2012).

Besides allowing for a reduced sample complexity and improved analytical sensitivity in targeted mass spectrometry analysis, affinity-reagent based purification coupled with shotgun mass spectrometry offers a possibility for a comprehensive identification of interaction partners of proteins in order to reveal functionally relevant protein associations (Miteva et al., 2013). Another hybrid and recent strategy combining the utility of antibodies, labeled with metal isotopes in this case, with mass spectrometry-based methodology is *mass cytometry* as described by Bandura et al. (2009) and utilized by Bendall et al. (2011) for single cell analysis. It circumvents the multiplexing limitations of conventional fluorescence-activated cell sorting (FACS) methods and represents a new

tool suited for a comprehensive characterization of immune cells in the context of diseases such as autoimmune conditions.

Contemporarily, mass spectrometry plays a major role within the field of proteomics. Despite the limited sample throughput capacity, it is a very powerful technology. It allows for the unambiguous identification of the proteins present in a biological sample, and their modifications. As I previously discussed, one of the main limitations of mass spectrometry-based approaches, namely the analytical sensitivity when analyzing complex body fluids such as blood, is being addressed by incorporation of affinity reagents, which are themselves the principal tools of *affinity-based proteomics*.

Affinity reagents can be used to examine various aspects of the human proteome, or the sub-proteomes thereof. Several methods, including immunohistochemistry (IHC), immunofluorescence (IF), immunoblotting, flow cytometry, immunoprecipitation and immunoassays can be considered as classical methods making use of affinity reagents. Alone or combined, these various tools serve for several end-purposes: IHC analysis allows to investigate the expression profiles in tissues; confocal microscopy-based IF analysis reveals the sub-cellular localization of proteins; flow cytometry allows for an analysis of proteins displayed on the surface of cells: immunoprecipitation can be utilized to isolate protein complexes for downstream characterization and immunoassays can be performed for a qualitative or quantitative analysis of protein composition in body fluid samples.

While these methods using affinity reagents can be considered as classical methods for analysis of single or a limited number of proteins, performing e.g. immunoassays for a parallel analysis of hundreds or thousands of proteins or autoantibodies in an equal number of clinical samples can be regarded as a different realm. This realm of *affinity-based proteomics* gets stimulated by the ongoing efforts generating various types of affinity reagents on proteome-scale, as I have discussed in section 2.3.1. Availability of such comprehensive collections of antibodies and antigens opens up possibilities to implement high-throughput strategies for parallel analysis of proteins and autoantibodies in body fluids, where a remarkable fraction of the proteome can be explored. This aspect renders *affinity-based profiling* approaches a valuable alternative and the *affinity array* technology a crucial tool for profiling of proteins and autoantibody repertories.

In the following discussion, I will provide a more detailed overview of affinity arrays as high-throughout tools of the affinity-based proteomics field, which allow to analyze the protein content of body fluids and to identify targets of autoantibodies present in body fluids in the quest towards biomarkers.

4 Affinity Arrays for Profiling Proteins and Autoantibody Repertoires

An *array* is literally defined as an ordered arrangement of quantities or symbols, usually in rows and columns. Simply put, *affinity arrays* can be considered as ordered arrangements of test sites, where each site would host a certain, immobilized affinity reagent. They are miniaturized architectures of density, allowing for an analysis of proteins or antibodies present within small volumes of complex protein mixtures, such as body fluids.

In this final chapter of introduction, I will discuss available technologies for generating affinity arrays, with a focus on antibody and antigen arrays, which served as the main technical tools in the investigations carried out within this thesis work. In addition to providing an overview of various features of affinity array technologies, I will also touch upon important aspects which pose analytical or practical challenges. Finally I will give an overview of the attempts made so far regarding the application of affinity arrays for protein and autoantibody profiling, mostly geared towards identification of potential biomarker candidates. I will also provide a discussion about methodological and statistical considerations applicable for studies employing affinity arrays as their main discovery tools.

Yet, before embarking into this discussion, I would like define certain terms within the context they will be utilized throughout this chapter. One such important term is *multiplex* assay, which refers to the parallel detection of more than one target protein present in the same undivided volume of a sample. *Throughput*, which is not a synonym for multiplexing capacity, refers to the number of samples which can be analyzed within a single experiment. The analytical *sensitivity* of an assay refers to its ability to discriminate between different concentrations of a target, whereas the more explicit term lower *limit of detection* (LOD) of an assay refers to the concentration at which the assay read-out signal for a given target can be discriminated from the background signal at a given level of statistical significance* (Ekins and Edwards, 1997, Wilson, 2013). The *dynamic range* refers to the concentration range between the lower LOD and the concentration at which signal saturation occurs,

*A common practice defines lower LOD as the target concentration corresponding to a signal that is three standard deviations above the average background signal, namely the concentration at which the target has a probability of α =0.015 of not being detected (Wilson, 2013).

namely the upper LOD. The analytical *specificity*, or in recently more encouraged terms the *selectivity*⋆, of an assay refers to the degree of how effectively a given set of affinity reagents can produce measurable signals for their intended targets in a complex mixture without interference from other components in the mixture. Selectivity therefore closely depends on the degree of *cross-reactivity* of affinity reagents to unintended targets present in a given sample context, namely to the degree of their *off-target binding*, as well as on the degree of cross-reactivity of affinity reagents with each other. And finally, analytical sensitivity and selectivity (specificity) are distinct terms from *diagnostic sensitivity* and *diagnostic specificity*. These refer to the percentage of case samples truly identified by the assay as positive for the tested analyte and to the percentage of control samples truly identified by the assay as negative for the tested analyte, respectively (Saah and Hoover, 1997).

4.1 Affinity Array Technologies and Platforms

> Like the younger child who has had an older sibling to "soften up" his or her parents and make life a little easier, protein microarrays have benefited from lessons learned during the noisy adolescence of DNA microarrays.
>
> *Eisenstein (2006)*

Before discussing affinity arrays and the types and applications of assays on affinity arrays, I first would like to provide a brief historical perspective about *immunoassays*. As I previously mentioned, in a groundbreaking work Rosalyn Sussman Yalow (1921–2011) and Solomon Aaron Berson (1918–1972) described the concept of competitive radioactive isotope-labeled antigen detection method and utilized the developed *radioimmunoassay* (RIA) for detection of insulin in plasma (Yalow and Berson, 1959). Owing to their refusal to patent the method (Glick, 2011), RIA soon became widely applied for detection of hormones and immunoglobulins. Miles and Hales (1968) later described an immuno-radiometric assay, where radioactive isotope-labeled antibodies were utilized for detection of insulin in plasma. Wide et al. (1967) described a similar assay to detect allergen-specific antibodies.

⋆It has been previously suggested that when describing an analytical method, the use of the term *selectivity* should be preferred over *specificity* (Vessman et al., 2001). Although the term *specificity* still dominates the practical jargon, as well as the literature, in the subsequent discussion I will prefer to utilize the term *selectivity*.

The concerns with regard to the use of radioactive substances led to search for non-radioactive immunoassay alternatives with comparable analytical sensitivity. Despite the initial skepticism about the alternative of using enzymes as reporter molecules (Lequin, 2005), Peter Perlmann (1919–2005) and Eva Engvall demonstrated the feasibility of quantitative measurement of IgG in serum using alkaline phosphatase as the reporter molecule (Engvall and Perlmann, 1971). They named this assay *enzyme-linked immunosorbent assay* (*ELISA*), which soon became the most widely used method for protein detection. Subsequent to the introduction of the hybridoma technology for production of monoclonal antibodies (Köhler and Milstein, 1975), the first "sandwich" ELISA concept was proposed by Uotila et al. (1981). In this first report of *sandwich ELISA*, two monoclonal antibodies with different specificities, one as a capture antibody and another enzyme-labeled one as a detection antibody were utilized for detection of alpha-fetoprotein.

Despite some early attempts of microspot-based immunoassays by Feinberg (1961), it was not until the end of 1980′s that the theoretical principles and the concept of microspot-based immunoassays was put forward: Roger Ekins proposed that low amounts of capture molecules immobilized as microspots would capture only a minor fraction of the analyte and this fraction would reflect the analyte concentration (Ekins, 1989). An assay performed under such conditions would be independent of both the capture molecule concentration and the sample volume. The so-called *ambient analyte immunoassays* would therefore result in maximum fractional occupancy of the capture molecules, thus higher signal intensities per area and optimal signal-to-noise ratios (Ekins, 1989, Ekins and Chu, 1991). Ekins and Chu (1994) emphasized further that such a miniaturized microspot-based assay format would open up the possibility for a multi-analyte, namely multiplex, analysis. As envisioned in the latter report by Ekins and Chu (1994), the concept of measuring hundreds of analytes in parallel in a miniaturized assay setup has been widely adopted within the affinity array field. In the following discussion, I will give a more detailed overview regarding the available affinity array formats and different assay types.

4.1.1 Affinity Array Formats

Planar Arrays

The conceptual foundation for producing an immunoassay in a "microspot" fashion was first introduced by George Feinberg, when he demonstrated the utility of depositing microspots of an antigen solution with a fine capillary tube on an antiserum-agar coated microscope glass slide (Feinberg, 1961), which he later used for detection of autoantibodies against thyroglobulin in serum samples of autoimmune thyroiditis patients (Feinberg and Wheeler, 1963). Two decades later, Roger Ekins proposed the theoretical principles of microspot

assays. It was however not until 1990′s that the contemporary planar, or "spotted" microarray concept took its current shape, the DNA microarray technology hit its peak and the concept of a simultaneous analysis of several thousand probes spotted on solid supports, such as microscopic glass slides, proved itself as a highly valuable tool within the genomic and transcriptomic research field (Chee et al., 1996, Schena et al., 1995). Towards the end of 1990′s, the emerging proteomics field started to witness an equally fast growing need to establish an analogous concept for the simultaneous analysis of hundreds to thousands of proteins. In the beginning of 2000, the ready availability of robotic printing devices and scanners within the DNA microarray world led to the adaptation of equipments and procedures and the pioneering proof-of-concept studies demonstrating the feasibility of "printing proteins" followed (Haab et al., 2001, MacBeath and Schreiber, 2000). Soon after, the potential of spotted protein microarrays for the analysis of entire proteomes was demonstrated (Zhu et al., 2001, 2000) and highlighted (Phizicky et al., 2003, Zhu and Snyder, 2003).

Planar arrays are generated by immobilization of a large number of different capture or affinity reagents, such as antibodies, in microspots, generally at a spatial density of up to 2,000 per cm^2 or more (Wingren and Borrebaeck, 2007). The capture reagents are immobilized on the solid support by means of either contact printing devices using e.g. pins that touch the surface of the solid support (MacBeath and Schreiber, 2000, Zhu et al., 2000), or non-contact printing devices that involve jetting systems forming and propelling droplets onto the solid surface (Delehanty, 2004, Delehanty and Ligler, 2003). Only minute amounts of capture reagent volume are spotted, usually in the range of 50-500 pl, resulting in spot sizes of 100-300 µm depending on the utilized solid support properties (Espina et al., 2003).

A large number of different solid support options are available (Guilleaume et al., 2005), offering different surface chemistry choices for planar arrays (Kusnezow and Hoheisel, 2003, Zhu and Snyder, 2003). As reviewed by Rusmini et al. (2007), Sutandy et al. (2013), Tomizaki et al. (2005), surface immobilization strategies can in general be categorized as physical, covalent or affinity-based immobilization. Surfaces coated with nitrocellulose or gel pad allow for a passive adsorption resulting in non-covalent and non-specific attachment of proteins, whereas e.g. epoxy-activated surfaces allow for a covalent attachment of proteins through their accessible amine ($–NH_2$), hydroxy (–OH) or thiol (–SH) groups and e.g. avidin-coated surfaces allow for an affinity-based, non-covalent but specific attachment of proteins carrying a biotin tag. There is no universal solid support type with a particular surface chemistry, which is suitable for several, let alone all applications. On the contrary, before exploiting the full utility of planar affinity arrays for profiling of samples, usually a considerable amount of time has to be spent to develop and find the right surface chemistry and to fine-tune the assay protocols accordingly for the individual purpose. Here, several aspects need to be considered, such as whether the printed proteins have to retain their

functional stability without undergoing denaturation. Usually, a similar effort has to be spent for identifying the optimal buffer composition to block the array surface prior to application of samples.

These and similar considerations apply certainly more when home-brewed protocols need to be developed and adjusted to fabricate planar affinity arrays. There are however also various commercially available planar array options providing access to research labs not familiar with all aspects of the microarray fabrication technologies. One of the widely distributed planar protein array platforms is the ProtoArray® Human Protein Microarray, containing in its current version 5.0 over 9,000 full-length human proteins per slide. There are also various commercially available planar antibody array platforms, mostly consisting of several dozens to few hundreds of antibodies and targeting specific set of proteins such as phosphorylated proteins, cytokines or kinases, as exemplified by the RayBio® Human RTK Phosphorylation Antibody Array. A more extensive list of commercially available planar array formats has been provided e.g. by Yu et al. (2010) and Sutandy et al. (2013).

To date, signal generation on planar affinity arrays mostly relies on fluorescence-based methods. This is simply due to the fact that the laser scanners compatible with fluorescent dyes, developed and widely used for DNA microarrays, were directly implemented into the protein microarray field at its birth. Such scanners typically provide a fluorophore excitation at wavelengths of 532 and 633 nm, allowing for a detection of organic fluorescent dyes such as Cyanine and Alexa Fluor or protein-pigment complexes such as R-phycoerythrin (R-PE). As a side note, although read-out by fluorescence has been and still is by far the most widespread signal generation method within planar (as well as bead-based) affinity array applications, there is indeed a rich spectrum of more sophisticated alternatives. This includes both label-based methods, such as quantum dots, gold nanoparticles or surface enhanced Raman scattering and label-free methods, such as carbon nanotubes, nanowires and microcantilevers. A very comprehensive overview of such emerging signal generation methods, as well as a comparison regarding the detection sensitivities they offer, has been provided by Chandra et al. (2011) and Ray et al. (2010).

While practically being restricted solely to the availability of affinity reagents, planar array platform theoretically offers the possibility to spot tens of thousands of capture reagents within one glass slide for a highly multiplex analysis. This can be considered as the main advantage of planar affinity arrays, suiting the analysis of binding against thousands of proteins, antibodies or complex mixtures of samples at once. Certainly, the more features are spotted on the array surface increasing the multiplexing capacity of the assay, the less possibility to create sub-arrays, where individual samples can be analyzed on the same array surface, thus a lower sample throughput. Planar array platform can therefore be considered as a suitable initial discovery platform suited for identification of potential biomarker candidates by analyzing sample collections in the range of lower hundreds. These can be verified further on technological platforms offering higher sample

throughput capacity, such as the bead-based arrays discussed next. This strategy has been demonstrated within **Article I** and **Article II**, where planar arrays of human protein fragments, routinely generated within the Human Protein Atlas project based on the previously described protocols (Nilsson et al., 2005, Sjöberg et al., 2012), were implemented for a large-scale autoantibody profiling approach, followed by verification of findings using bead-based arrays.

Bead-Based Arrays

The principle of bead-based arrays, or the so called *suspension bead arrays*, relies on immobilization of capture reagents on distinguishable microsphere sets as solid supports and detection of the captured targets on each microsphere set by means of a flow cytometric read-out system. Microspheres utilized as solid supports carry functional groups, such as carboxyl (–COOH) or thiol (–SH) groups, facilitating the immobilization of proteins and antibodies. As demonstrated in **Article III**, microspheres pre-coupled with neutravidin can also be utilized to immobilize biotin-containing capture reagents, which offers an alternative for immobilization of e.g. synthetic, biotin-containing peptides to provide adequate space from the microsphere surface.

The concept of utilizing microspheres as a solid support is in fact not as new as it is generally anticipated (Kellar, 2002). Almost 40 years ago, the potential use of microspheres for immobilization of antigens to identify antigen-specific antibodies in serum was described by Horan and Wheeless (1977). A decade later, McHugh et al. (1988) demonstrated this possibility by simultaneously detecting serum antibodies against herpes simplex and cytomegalovirus antigens, which were immobilized on microsphere sets with different diameters. Towards the end of 1990′s, embedding spectrally different fluorophores into microspheres to prepare 64 spectrally distinct microsphere populations was demonstrated (Fulton et al., 1997, Kettman et al., 1998). The letter work lay the foundation for the multiplex microsphere array technology provided today by Luminex Corporation under the name of xMAP® technology (standing for *multi-analyte profiling*).

Curently, the xMAP® technology is the most widely utilized bead-based platform. It is built on the use of polystyrene microsphere sets, which are internally dyed with precise amounts of two or three spectrally different fluorophores with distinct emission profiles. Since each microsphere set can be distinguished by their spectral address, microsphere sets, each linked to a distinct capture reagent can be combined within a single array, allowing for a multiplex measurement. The first generation of these microspheres contained two internal fluorescent dyes allowing for a multiplexing of up to 100 analytes, whereas the next generation microspheres, introduced in 2010, contain three internal dyes allowing for a multiplexing of up to 500 analytes. In addition to commercially available technologies, it is also feasible to develop home-brewed protocols to generate spectrally distinguishable

microspheres allowing to create a microsphere array as demonstrated by Wu et al. (2009), where maleimide-derivatives of fluorescent dyes were used to generate 1,152 different microsphere sets.

The Luminex 100/200™ and FLEXMAP 3D® analyzers are compatible with 100-plex or 500-plex bead arrays and offer a throughput of 96 or 384 samples in micro-titer plates, respectively. In the flow cytometer-like analyzers, each microshpere is subjected to a 635 nm laser beam, which excites the internal fluorescent dyes in each microsphere and allows for the deconvolution of the spectral address. A 532 nm laser beam excites the reporter molecule (preferably R-phycoerythrin) bound to the analyte and enables for read-out in terms of median fluorescence intensity (MFI) across each distinct microsphere set (Kellar and Oliver, 2004). As demonstrated by Jacobson et al. (2006) and later by Won et al. (2012), it is also possible to use the fluorescence measurements of individual microspheres within each set and to perform more customized statistical analysis of individual binding events.

The Microplex® microspheres, with a diameter of 5.6 µm are the standard non-magnetic microspheres (Figure 4.1), whereas the MagPlex® microspheres of 6.5 µm diameter are embedded with superparamagnetic particles (Angeloni et al., 2013), allowing for an automated plate washing or magnetic bead transfer, which are very important aspects contributing for a good recovery of microspheres and good assay reproducibility and a reduced hands-on time and human error. Both non-magnetic and magnetic microspheres are functionalized with carboxyl groups, allowing for covalent immobilization of proteins or antibodies via their primary amine groups. As reviewed by Hsu et al. (2009), there are several commercially available ready-to-use kits with current clinical diagnostic utility for single- or multiplex analysis of e.g. cytokines and growth factors. Similarly, there are ready-to-use kits for analysis of autoantibodies, such as anti-nuclear antibodies (ANAs), where some kits for the latter obtained approval from US Food and Drug Administration (FDA) (Shovman et al., 2005).

Several studies have evaluated and compared the performance of bead-based multiplex sandwich assays to sandwich ELISA and reported usually very high correlations and similar or even better detection sensitivities for multiplex measurement of cytokines (De Jager et al., 2003, duPont et al., 2005, Elshal and McCoy, 2006) or human antibodies (Dasso et al., 2002). Similar studies have been performed to evaluate the agreement between measured serum autoantibody levels, such as ANAs (Martins et al., 2004, Venner et al., 2013), or levels of antibodies against viral antigens e.g. from Epstein-Barr virus (Martins et al., 2008), using ELISA and multiplex bead-based assays which reported equally concordant performances. Offering less consumption of crude sample volume, less sample processing time and multiplexing capacity while not compromising analytical sensitivity and accuracy makes bead-based array platforms highly versatile tools with large number of applications.

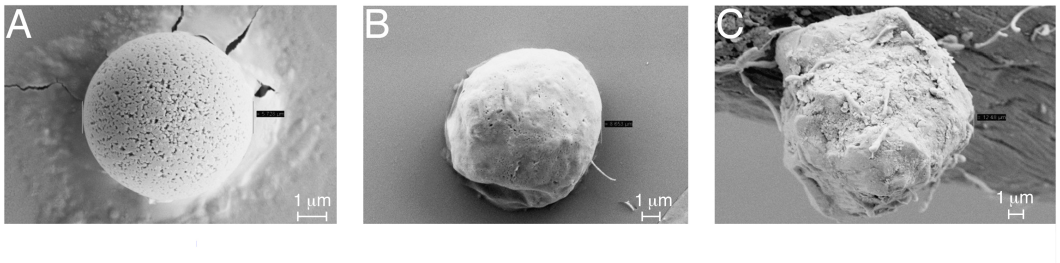

Figure 4.1: Microplex® beads under the microscope.
Ultra-high resolution field emission scanning electron microscope micrographs of Microplex® beads prior to (A) and after conjugation to a protein (B) and upon addition of a biotin-labeled antibody and streptavidin-phycoerythrin (SAPE) for read-out (C), with microsphere diameters of approximately 5.6 µm, 8.6 µm and 12.5 µm at each respective condition. Image adapted from work by Liu et al. (2013).

Bead-based arrays offer also certain advantages over planar arrays, including no requirement for laborious image analysis, direct collection of data, greater sample throughout and a more flexible and customizable array content. An array facility to generate and utilize planar arrays would require scanners, image analysis softwares and sophisticated printing robots which can usually be operated by experienced users, whereas a facility based on bead-based arrays might be established in less time-frames due to the more user-friendly nature of the currently available bead-based technologies. However, as indicated before, planar arrays offer a possibility to generate arrays with hundreds to few thousands of proteins or antibodies per cm^2, thus theoretically allowing for a greater degree of multiplexing than offered by the most widely used bead-based platforms. As demonstrated in one of the large studies included in this thesis (**Article I-II**), these two array platforms indeed complement each other when efficiently utilized at different phases of biomarker discovery studies, as well as when each platform is utilized to technically verify the findings revealed by the other, as demonstrated in **Article I** and by Rimini et al. (2009).

The investigations presented in this thesis all employed bead-based arrays either for profiling proteins or antibody reactivities in body fluid samples. The protocols for the immobilization of antibodies or antigens on beads were originally described by Schwenk et al. (2008, 2007), and as later summarized by Drobin et al. (2013), a semi-automated workflow for a parallel immobilization of up to 384 antibodies (or antigens) in micro-titer plates has been established. As demonstrated in all the studies included in the thesis, the open architecture offered by the bead-based arrays offers the possibility to develop and continuously fine-tune home-brewed protocols to create arrays with a fully customized content of proteins, peptides or antibodies. It enables to carry out different proteome profiling tasks by addressing particular sets of targets in the context of various diseases.

4.1.2 Assay Formats on Affinity Arrays

Although the nomenclatures are not very frequently used in practice, assays performed on affinity arrays can be categorized in different ways. One categorization has been made between *forward-phase assays* and *reverse-phase assays* (Caiazzo Jr et al., 2009, Hartmann et al., 2009, Yu et al., 2010); another one has been proposed as *functional assays*, *analytical assays* and *reverse-phase assays* (Hall et al., 2007, Kodadek, 2001, Wolf-Yadlin et al., 2009).

Forward-phase assays involve the immobilization of capture reagents, such as antibodies on solid supports, either planar or bead-based surfaces, to capture and detect specific proteins from a complex mixture of proteins. Forward-phase assays thereby enable to obtain the protein profile of each sample across hundreds to thousands of capture reagents in parallel. Forward-phase assay strategy is mainly applied when utilizing antibody or other affinity reagent arrays for protein profiling purposes in complex samples, which I will discuss in more detail in the following section. Assays carried out on antigen arrays are also regarded as a special type of forward-phase assay, where the immobilized antigens on a solid support serve as bait reagents, so that the antibodies present in a complex sample, or interaction partners of the immobilized antigens, can bind to the antigens and be detected in parallel. In the following sections, I will also describe in more detail antigen arrays and their applications, with a focus on profiling autoantibody repertoires.

Regarding the second categorization of assay formats, in fact *forward-phase assays* assays carried out either on antibody and antigen microarrays with the aim to profile complex mixture of protein samples can be considered as *analytical assays*. *Functional assays*, on the other hand, would mainly refer to the use of arrays of antigens, generally full-length proteins or protein domains, for studying their biological activities by the analysis of protein-protein (Zhu et al., 2000), protein-DNA (Boutell et al., 2004), protein-lipid (Zhu et al., 2001), protein-small molecule or enzyme-substrate (MacBeath and Schreiber, 2000) interactions.

Before limiting the discussion in the following sections to the forward-phase assays –or analytical assays– on antibody and antigen arrays, I would like to provide a brief overview of the *reverse-phase assay* format and its applications.

In reverse-phase assays, instead of the capture or bait reagents, minute amount of the complex samples to be analyzed are themselves immobilized on solid supports. This is mostly achieved by means of adsorption on nitrocellulose coated glass microscope slides (Speer et al., 2007). Subsequently, the immobilized complex samples are interrogated in parallel, using a single reagent specific for a target of interest. The term reverse-phase array was coined within the seminal work by Paweletz et al. (2001), where the feasibility of parallel analysis of lysates obtained from micro-dissected malignant prostate lesions was demonstrated to elucidate cell signaling events. Several signal pathway profiling studies,

including also the time course analysis of post-translational phosphorylation events by employing phospho-specific antibodies have adopted the reverse-phase assay strategy for parallel analysis of lysates collected mostly from cancer lesions in the context of various cancer types (Grubb et al., 2003, Sheehan et al., 2005, Wulfkuhle et al., 2003). Development of commercial reverse-phase assay platforms, such as the ZeptoMARK platform (Pawlak et al., 2002), offered improved analytical sensitivity within tissue lysate applications. It also broadened the application context of reverse-phase assays to other diseases, such as for analysis of muscle tissue lysates within muscular dytrophies (Escher et al., 2010). In addition to serving as a miniaturized equivalent to the classical immunohistochemistry assays, the utility of reverse-phase assays has been also demonstrated for protein or phospho-protein profiling in various studies analyzing different types of cancer cell line lysates (Boyd et al., 2008, Chan et al., 2004, Chruscinski et al., 2013, Luckert et al., 2012, Nishizuka et al., 2003, Sevecka et al., 2011, Ummanni et al., 2014), as well as intact cells (Schwenk et al., 2002).

Not least of all, the reverse-phase assay concept has been also implemented to develop *serum arrays*, allowing for a parallel analysis of several hundreds of serum samples (Aguilar-Mahecha et al., 2009, Grote et al., 2008, Janzi et al., 2009a, 2005, 2009b). However, due to the simple fact that only minute volumes of samples are spotted on the array surfaces, there is an intrinsic analytical challenge regarding the analytical sensitivity of reverse-phase assays, unless additional signal amplification strategies, such as application of fluorescent vesicles (Bally et al., 2011), are employed. Thus, the targets of interest in the serum applications have mainly been mid- or high-abundant serum proteins, such as immunoglobulins or C3 in the context of immunodeficiencies. These allowed to analyze targets present in serum at a concentration range exceeding the LOD of the reverse-phase assay format, which is reported to be in the 500-700 ng/ml range (Aguilar-Mahecha et al., 2009, Janzi et al., 2009b).

Application of reverse-phase assays for analysis of proximal body fluids has been generally very restricted, with less than a handful of attempts such as by Gyorgy et al. (2010) for analysis of CSF. General technical and analytical challenges and considerations regarding the reverse-phase assay format, as well as a broader overview of its application range has been discussed by Speer et al. (2007), Vanmeter et al. (2007), Korf et al. (2009) and Mueller et al. (2010). Although the studies carried out within the frame of this thesis work have not directly employed the reverse-phase assay format, I finally would like to underline its potential for a simultaneous analysis of up to thousands of samples for an abundant target of interest. This aspect renders the reverse-phase assay format a very valuable strategy in the context of verification of identified potential biomarker candidates in extended sample collections of thousands of body fluid samples.

4.2 Antibody Arrays for Protein Profiling of Body Fluids

> Whereas DNA chips use the natural power of hybridization, protein microarrays commonly harness Nature's own protein capturing agents – antibodies.
>
> *Mitchell (2002)*

Within section 2.3, I provided an overview of various types of affinity reagents, including antibodies, recombinantly produced antibody fragments such as single-chain variable fragment (scFv), as well as alternative, namely non-immunoglobulin affinity reagents such as nucleic acid aptamers. These and similar affinity reagents can be utilized to build affinity arrays, either in planar or bead-based format, which can be implemented for the multiplex protein profiling of body fluids. Using such multiplex affinity arrays, and exploiting their sample throughput capacities, comparative studies searching after differential protein profiles in disease *vs.* health can be carried out.

Assay formats utilizing planar or bead-based affinity reagent arrays can be divided into two categories: the *single-binder assay* format and the *sandwich assay* format. The single-binder assay is based on interrogation of an affinity reagent array with a labeled sample, followed by the detection of the labeled target proteins captured from the given sample. As reviewed by Wingren and Borrebaeck (2008), samples can be labeled either directly with fluorescent dyes, such as Cyanine dyes in a *single-color* or *dual-color* approach and be subjected to assay read-out; or alternatively they can be labeled e.g. with the small molecule *biotin* and be detected using *streptavidin* conjugated to a fluorescent dye by exploiting the robust and very high affinity biotin-streptavidin interaction*. The pioneering and early studies demonstrating the utility of single-binder antibody arrays employed sample labeling strategies based on fluorescent dyes (Haab et al., 2001, Miller et al., 2003, Sreekumar et al., 2001). These studies followed the tradition of DNA microarrays and adopted a *dual-color* approach, where two different samples, one of them usually applied as a reference sample for internal normalization (Andersson et al., 2005), could be mixed and simultaneously analyzed per array. This approach is indeed still applied (Alhamdani et al., 2010, Schröder et al., 2010, Srinivasan et al., 2014). Other studies, such as by Knezevic et al. (2001), Gao et al. (2005) and Orchekowski et al. (2005) utilized biotin-labeled samples. Later, using antibody and recombinant scFv fragment arrays, respectively, Kusnezow et al. (2007) and Wingren et al. (2007) demonstrated

*As a side note, the dissociation constant K_d for the biotin-streptavidin interaction is $\sim 10^{-15}$, about 10^{5}-10^{8} times lower than the K_d for antigen-antibody interactions which is typically in the range of 10^{-7}-10^{-10} (Diamandis and Christopoulos, 1991).

that a single-color approach employing biotin labeling outperforms a dual-color approach employing fluorescent dyes. An assay protocol for single-binder bead-based assays has been described by Schwenk et al. (2008), where biotin-labeled samples could be applied without the removal of excess biotin, followed by a read-out using R-phycoerythrin conjugated streptavidin, which has also been the adopted strategy within **Article IV** and **Article V** of this thesis work.

The concept of multiplex sandwich assay format relies on interrogation of an affinity reagent array with a sample, followed by detection of the captured targets using a cocktail of second affinity reagents, where each detection affinity reagent is matched to one of the capture affinity reagents. These assays therefore represent a multiplex version of standard ELISA immunoassays. For signal generation, either each detection antibody can be fluorescently labeled; a fluorescently labeled tertiary antibody specific for the species of the detection antibodies can be utilized; or biotin-labeled detection antibodies can be detected by fluorophore conjugated streptavidin. Several early examples of multiplex sandwich assays in a planar array format have been described (Huang, 2001, Nielsen et al., 2003, Tam et al., 2002, Wiese et al., 2001), demonstrating their utility for a multiplex detection of various cytokines, where mostly less than a dozen of different targets were analyzed. Multiplex sandwich assays have also been established on bead-based arrays (De Jager et al., 2003, Hsu et al., 2008) and as mentioned previously, this array platform became commercially mature for the analysis of cytokines and chemokines in a multiplex sandwich assay format, with the availability of several ready-to-use kits.

As signal generation in sandwich assays depends on two antibodies binding their intended target, rather than one, the selectivity of sandwich assays outperforms the selectivity of single-binder assays. Similarly, sandwich assays in general offer a better analytical sensitivity due to reduced background (Haab, 2005). However, sandwich assays are more cost- and labor-intensive to develop. First, matching pairs of antibodies need to be found. While antibodies with a multi-epitope nature, such as polyclonal, or affinity purified mono-specific antibodies might be utilized both as capture and detection reagents, establishing an assay based on monoclonal antibodies would require two different reagents generated against different epitopes of each target of interest. Second, the detection reagents need to be labeled, either with biotin or a fluorescent dye. Although protocols for semi-automated and parallel labeling of affinity reagents are merging (Dezfouli et al., 2014), accomplishing this step is still a very labor-intensive stage if labeling is performed for dozens of affinity reagents. Finally, possible cross-reactive interactions between and across detection and capture reagents need to be evaluated in detail and the array content needs to adjusted prior to application on biological samples. Because of these practical and technical reason, sandwich assays are usually quoted to have a multiplexing capacity of lower than 50 targets (Haab, 2005), and the multiplexing limit of commercially available assays do not even extend beyond 30 (Tighe et al., 2013). In contrast, single-binder assays

in theory offer an unlimited multiplexing capacity as also underlined by (Haab, 2006, Hoheisel et al., 2013), although their practical multiplexing capacity is certainly equally restricted by the availability of affinity reagents.

As demonstrated in **Article IV** and **Article V**, single-binder assay strategies on affinity reagent arrays can be regarded as highly suitable tools for flexible protein profiling approaches geared towards discovery of potential biomarker candidates. With the growing number of available affinity reagents, an increasingly larger fraction of the human proteome can be explored using single-binder assay strategies. However, once potential biomarker candidates are identified, more focused assay development efforts should be spent to develop sandwich assays, which can offer improved assay performance characteristics and thereby facilitate the introduction of identified biomarker candidates into clinical practice.

4.2.1 Technical Challenges and Considerations

> Much fishing for biomarkers has been tried in the shallow waters of the more abundant proteins without much to show for, further underlining the need to now probe the depths.
>
> *Landegren et al. (2012)*

In this section I would like to address technical and analytical aspects relevant for assays carried out on affinity arrays. As I already discussed in section 3.2.1, one challenge when analyzing especially blood-derived samples on multiplex affinity arrays is the vast dynamic range of target concentrations. The fact that several potentially disease-related proteins might be present at very low concentrations renders the analytical sensitivity as a cardinal aspect. In recent years, "divide and conquer" strategies of the mass spectrometry field have been adopted to address this challenge within affinity array field. This includes reduction of the sample complexity by means of pre-fractionation (Ingvarsson et al., 2006, Slaastad et al., 2011) or depletion of high abundant proteins, such as albumin or immunoglobulins (De Jager et al., 2005). While several studies demonstrated the utility of depletion strategies, others have proposed that affinity array-based methods do not necessarily benefit from depletion steps due to co-depletion of low abundance proteins including cytokines (Granger et al., 2005) and due to introduction of a bias in protein composition (Schröder et al., 2010). As a result, the practice of pre-fractionation and depletion prior to analysis remains a matter of technical discussion, although protocols allowing to by-pass such pre-analytical steps are certainly preferable in regard to throughput and practicality.

Signal amplification by nucleic acid amplification strategies offers a direct approach to address the analytical sensitivity of protein profiling assays and enhance the lower LOD.

ImmunoPCR (iPCR), introduced by Sano et al. (1992), represents the first example of such nucleic acid-assisted protein detection methods, benefiting from the intrinsic power of polymerase chain reaction (PCR) for specific amplification of DNA by several orders of magnitude (Saiki et al., 1988). To summarize the general concept of iPCR, antibodies conjugated to DNA probes are utilized for detection of bound proteins and the DNA probes are subsequently amplified by PCR. Although currently the amount of such DNA products can be quantified in quantitative (or real-time) PCR (Heid et al., 1996), this was not the case by the time when iPCR was introduced. In its original format, iPCR involved a labor-intensive step for estimation of the amount of PCR product by gel electrophoresis. Thus, despite its potential, iPCR has been so far underutilized, however there are recent studies demonstrating its utility for sensitive detection of proteins (McDermed et al., 2012).

A similar signal amplification method is *rolling circle amplification* (RCA). Here, upon capture of a target protein by a DNA probe-conjugated antibody, a circular DNA template or a padlock probe (Nilsson et al., 1994) is added, which hybridizes to the DNA probe on the antibody. Amplification by DNA polymerase results in a long, single-stranded DNA product, which can be detected by complementary, fluorescently-labeled detection probes. Several proof-of-concept studies utilizing RCA method reported three orders of magnitude improvement in LODs (Schweitzer et al., 2002, 2000, Shao et al., 2003, Zhou et al., 2004), and the concept of RCA has been applied within several biomarker discovery-oriented studies as well (Kaukola et al., 2004, Mor et al., 2005, Orchekowski et al., 2005, Sanchez-Carbayo et al., 2006). More extensive discussions regarding iPCR and RCA can be found by Adler et al. (2008), Malou and Raoult (2011), Niemeyer et al. (2005) and by Kingsmore and Patel (2003), Partyka et al. (2014), respectively.

At this point, I would like to open a parenthesis and touch upon the "ultra-sensitive" immunoassay technologies, which have been emerging lately. One such technology is the so-called digital ELISA on single-molecule arrays (SiMoA) (Rissin et al., 2010, Rissin and Walt, 2006), which allows for detection of single antibody-protein complexes captured on microspheres by confining them to femtoliter-well arrays. Simply, isolation of single complexes captured on single microspheres into single reaction wells allows for a subsequent digitalization of the detection of captured proteins by counting. In contrast to the traditional method of signal generation where the total signal is an ensemble averaging of fluorophore molecules, in this concept a single fluorophore molecule is often enough to generate a signal over background, thus allowing for sensitive measurement of single molecules, with a reported lower LOD of 6 fg/ml for PSA, for instance. This technology, with a current multiplexing capacity of up to ten proteins, was made commercially available by Quanterix Corporation recently. A similar single-molecule counting platform, with similar levels of reported LODs, is the Erenna® immunoassay system offered by Singulex Corporation. Here, upon capture of proteins by capture antibodies on microspheres, fluorescently labeled detection antibodies are added, which are subsequently eluted for

a single-molecule counting process in a capillary flow system (Todd et al., 2007, Wu et al., 2006). There are indeed other attempts to implement single-molecule detection concept on affinity arrays, such as demonstrated by Tessler and Mitra (2011), describing a planar sandwich assay format using total internal reflection fluorescence (TIRF) imaging for single-molecule counting.

Although these emerging technologies offer great analytical sensitivity, they are limited in terms of multiplexing capacity, as any other technology based on sandwich-assay format. Multiplexing is indeed an equally important aspect as sensitivity, especially when considering discovery-oriented applications. By increasing the degree of multiplexing, single-binder assays are mainly susceptible to an increased possibility of off-target interactions between capture antibodies and proteins present in a given sample. Namely, the cross-reactive interactions which might take place in single-binder assays are mainly sample-driven⋆. In addition to such sample-driven off-target interactions, multiplex sandwich assays are also susceptible to reagent-driven cross-reactive interactions, namely the cross-reactive interactions between detection antibodies with other detection antibodies, as well as with capture antibodies. As highlighted by Juncker et al. (2014), in a sandwich assay for n target proteins, theoretically there are in total $4n(n-1)$ cross-reactive interactions which might take place. Recently, various concepts have been suggested to minimize such reagent-driven cross-reactive interactions in sandwich assays. In the so-called bead-based sequential multiplex analyte capturing assay, Poetz et al. (2010) demonstrated the possibility of analyzing the very same sample with different capture and detection antibody panels related both to abundance and phosphorylation status of different tyrosine kinases, thus reducing the reagent-driven cross-reactivity by a "temporal" separation of the reagents. Similarly, in the so-called antibody colocalization microarrary, Li et al. (2012), Pla-Roca et al. (2012) demonstrated the possibility of "spatially" separating the reagents by spotting detection antibodies on corresponding capture antibody spots upon sample incubation.

Proximity ligation assay (PLA), originally described by Fredriksson et al. (2002), is another elegant approach essentially addressing the reagent-driven cross-reactivity, thus enhancing the assay selectivity. In PLA, similar to iPCR, detection antibodies of a mostly polyclonal source are conjugated to DNA probes. Yet, for each target, antibodies are conjugated to not a single but two (or up to four) different DNA probes, which consist of both a unique identifier and a common ligation sequence. Upon coincident binding of correct pairs of detection antibodies, the probes come into close proximity to be joined by DNA ligation. The ligation product serves as an amplifiable reporter molecule, which can be

⋆As a side note, in the investigation listed as the fifth in the related articles, a sequential capture assay concept was developed to minimize the off-target interactions in single-binder bead-based assays. Briefly, this assay concept relies on a first capture with a bead-based array, labeling of captured targets on beads, elution of labeled targets and read-out using a secondary capture array.

detected by qPCR. PLA was originally described as a homogeneous assay and was adopted by Darmanis et al. (2010) for use on microspheres as solid support, where a multiplexing capacity of up to 36 targets was demonstrated using this latter format (Darmanis et al., 2011). Recent and more extensive discussions regarding PLA have been provided by Nong et al. (2012) and Blokzijl et al. (2014).

Multiplexing is not only a challenging factor for the selectivity of affinity reagent-based assays, but it potentially also restricts the linear dynamic range which can be achieved for certain proteins present in a complex sample. Since each protein is present at a different concentration and since each antibody-protein interaction has its own affinity characteristics, ideally each protein might require for distinct assay conditions for a measurement within a linear dynamic range. Thus, when utilizing arrays with several affinity binders targeting different proteins simultaneously, a linear range might not be achieved for all. This can be even more restricted in case of sandwich assays, depending not only on one but potentially two different affinity characteristics.

As for any affinity-based technology, the quality and nature of affinity reagents utilized on affinity arrays is of uttermost importance and the awareness about this is frequently expressed in recent years (Marx, 2013, Perkel, 2014). Confirmation of the binding specificity of utilized affinity reagents against their target proteins is a key issue, for which various approaches can be taken including a Western blot analysis, as well as utilizing generic binding assays such as the strategy employed within the Human Protein Atlas Project, as I have discussed in section 2.3.1. Epitope mapping, for instance on high-density peptide arrays (Buus et al., 2012, Forsström et al., 2014), can further contribute for a confirmation of antibody specificity and a better understanding of the antigen binding sites. Certainly, a mass spectrometry-based quality-control approach can also be taken to verify that the utilized antibodies in a certain assay environment and sample context are capturing their intended targets, as demonstrated e.g. by Neiman et al. (2013) for assays on bead-based antibody arrays. The type and nature of immunogens utilized to generate the applied antibodies also need to be considered when developing assay protocols. Several commercially available antibodies, including the mono-specific antibodies generated within the Human Protein Atlas project, are for instance raised towards peptides or protein fragments. Utilizing such affinity reagents might require antigen denaturation e.g. by heat in order to enhance the epitope retrieval, as it has been demonstrated by Schwenk et al. (2008, 2010b) and applied within **Article IV** and **Article V**.

4.2.2 Applications of Antibody Arrays for Biomarker Discovery

Since their emergence in the beginning of 2000′s, antibody and other affinity reagent arrays have been increasingly more utilized within discovery-oriented, disease-specific proteomics studies. The vast majority of the early studies have been performed within

the context of cancer (Gao et al., 2005, Hudelist et al., 2004, Miller et al., 2003, Mor et al., 2005). Although cancer field continued to frequently employ array-based strategies for discovery of disease-associated protein profiles (Kim et al., 2009, Orchekowski et al., 2005, Sanchez-Carbayo et al., 2006, Shafer et al., 2007), the spectrum of diseases where affinity reagent arrays are being applied has remarkably extended in recent years, including, but not limited to metabolic diseases (Schwenk et al., 2010a); hyperinflammatory diseases (Hsu et al., 2008); autoimmune diseases such as Sjöogre's syndrome (Szodoray et al., 2004), SLE (Bauer et al., 2006, Carlsson et al., 2011b) or autoimmune pancreatitis (Sandström et al., 2012); neurodegenerative diseases such as Alzheimer's disease (Britschgi et al., 2011, Hye et al., 2014) or Parkinson's disease, as well as infectious diseases such as malaria (Bachmann et al., 2014) or tuberculosis (Nahid et al., 2014). The work presented in **Article V** demonstrates a first application of antibody arrays for protein profiling in the rare disease field, where plasma and serum samples were profiled on bead-based antibody arrays in the Duchenne muscular dystrophy context. Similarly, the work presented in **Article IV** is a first, large-scale application of antibody arrays for protein profiling in plasma within a multiple sclerosis context.

The single-binder assay format constitutes the main strategy applied for discovery-oriented protein profiling studies. There are indeed examples where a multiple sandwich assay format has been utilized e.g. by Kaukola et al. (2004), Szodoray et al. (2004) and Bauer et al. (2006), but these studies have mostly focused on profiling cytokines, growth factors and soluble receptors, or they were investigating a previously identified candidate protein biomarker in an extended sample collection for verification purposes (Qundos et al., 2014). Studies adopting a single-binder assay strategy could investigate more versatile hypothesis-driven sets of targets, including e.g. targets known to be expressed in the affected tissue type or organ or known to be involved in cellular pathways associated with the disease, as demonstrated in **Article IV** in muscular dystrophy context. Furthermore, single-binder assay formats have also been used to profile entirely hypothesis-free assembled sets of targets, as demonstrated e.g. by Bachmann et al. (2014), Nahid et al. (2014), Sattlecker et al. (2014), as well as in **Article V**. There are also several studies which have combined the single-binder assay format with a read-out based on rolling circle amplification (Kader et al., 2005, Mor et al., 2005, Orchekowski et al., 2005, Sanchez-Carbayo et al., 2006, Shafer et al., 2007).

Serum and plasma samples constitute the sample type which have so far been mostly applied in discovery approaches using affinity reagent arrays, however there are example studies utilizing tumor tissue extracts in cancer context (Mehan et al., 2012, Srinivasan et al., 2014) and cerebrospinal fluid in the context of neurological diseases such as multiple sclerosis (Häggmark et al., 2013). Current examples of affinity array-based studies utilize both antibodies and recombinant antibody fragments. The single-binder, bead-based protein profiling approach described by Schwenk et al. (2008) has been implemented

within the context of several diseases, such as kidney disorders (Neiman et al., 2011), small intestine neuroendocrine tumors (Darmanis et al., 2013), malaria (Bachmann et al., 2014), muscular dystrophies (**Article V**) and multiple sclerosis (**Article IV**), where mainly affinity-purified polyclonal antibodies generated within the Human Protein Atlas have been utilized. The single-binder, planar-array based approach described by Wingren et al. (2005) utilized the recombinant single-chain variable fragments (scFv) introduced by Söderlind et al. (2000). This platform has been implemented in various protein profiling studies within e.g. pancreatic cancer (Ingvarsson et al., 2008, Wingren et al., 2012), breast cancer (Carlsson et al., 2008, 2011a) and B-cell lymphoma (Pauly et al., 2014). Similarly, Ramirez et al. (2010) utilized scFv arrays in the context of ovarian cancer. Protocols developed for single-binder, planar-array based analysis of tumor tissue extracts (Alhamdani et al., 2010, Schröder et al., 2010) have also been utilized within cancer context (Srinivasan et al., 2014), as reviewed by Hoheisel et al. (2013). In addition to antibodies and recombinant antibody fragments, an assay based on aptamers has been described by Gold et al. (2010) and Kraemer et al. (2011), which has recently been implemented for protein profiling studies within tuberculosis (Nahid et al., 2014), Alzheimer's disease (Sattlecker et al., 2014) and lung cancer (Mehan et al., 2012).

4.3 Antigen Arrays for Profiling Autoantibody Repertoires of Body Fluids

> I'd like to see a future where people can just think up an experiment, buy an array, and do it.
>
> *Michael Snyder (Eisenstein (2006))*

Antigen arrays, either in planar or bead-based format, are built of "bait" reagents, such as proteins or peptides. Using antigen arrays, both analytical and functional assays can be performed. In a functional assay format, the arrayed antigens are utilized to decipher various binding activities such as protein-protein, protein-drug, protein-peptide or protein-nucleic acid interactions, as reviewed e.g. by Zhu and Qian (2012) and Sutandy et al. (2013). In an analytical assay format, arrays of antigens are used to detect and characterize antibodies, either against self- or non-self antigens. In this section, I will frame the discussion mainly to the analytical assays for antibody profiling on antigen arrays. Although for this purpose other types of bait reagents, such as lipids or carbohydrates, can be utilized as antigens, I will mainly discuss arrays of full-length proteins, or fragments thereof, and arrays of peptides.

Protein Arrays

There are two strategies to fabricate protein arrays: Proteins can be either immobilized on planar or bead-based surfaces upon recombinant expression and purification, or they can be expressed *in situ*, namely *on spot* (or on bead) using cell-free expression systems. I would like to provide an overview of first the *in situ* protein array fabrication strategies, followed by the examples of more conventional approaches.

The *in situ* array production strategy eliminates the need to separately express and immobilize the proteins and relies on the synthesis of proteins directly on the array in a cell-free manner. Here, DNA templates, either in the form of a plasmid or a PCR product are utilized. The DNA template allows for protein synthesis in the presence of a crude cell lysate, which contains all macromolecular components needed for the transcription and translation machineries, such as ribosomes and initiation, elongation and termination factors. These components are also externally supplemented with e.g. amino acids, salts and cofactors. Cell-free protein expression is in fact not a recent concept; the feasibility of such cell-free expression mechanisms was revealed by the work of Nirenberg and Matthaei (1961) in the quest towards deciphering the genetic code. Cell-free protein expression systems were traditionally made from *E. coli*, followed later by cells from eukaryotic species, such as wheat germ cells, insect cells or rabbit reticulocyte cells (immature red blood cells), as reviewed by Endo and Sawasaki (2006), Spirin and Swartz (2008), He et al. (2011) and Carlson et al. (2012). Within the last decade, several *in situ* protein array production strategies emerged upon the marriage between such commercially available cell-free protein synthesis methods and the protein array technology.

A prototype strategy, the so-called *protein in situ* array (PISA), later also known as DiscernArray™ (He and Taussig, 2003), was described by He and Taussig (2001). In PISA, PCR-generated DNA constructs are designed to encode a protein or a fragment thereof, while also including a tag sequence. The array surface is pre-coated with a tag capture agent and after translation using an *E. coli* or rabbit reticulocyte lysate, the synthesized proteins with the tag are directly captured on the array surface and the unbound lysate material is washed off. The originally described method of PISA was however a "macro" method based on 24 wells per array and required 25 µl of DNA-lysate mixture per well. Angenendt et al. (2006) later miniaturized PISA further by utilizing a multiple spotting technique, where first the DNA template and then the lysate were spotted, reducing the required volumes to sub-nanoliter volumes and increasing the theoretical spot density up to 13,000.

An alternative to PISA, the so-called *nucleic acid programmable protein array* (NAPPA), was proposed by Ramachandran et al. (2004). As opposed to the original version of PISA where protein synthesis took place in solution, the concept of NAPPA is based on spotting the DNA template, in this case a biotinylated plasmid, encoding the protein of interest

as a fusion with a GST tag (glutathione *S*-transferase). The array surface is pre-coated with avidin, as well as an antibody targeting the GST tag is immobilized on the array. Upon incubation of the spotted plasmid DNA array with the rabbit reticulocyte lysate, the expressed protein with GST tag is immediately captured within each spot by the antibody against GST tag. A next version of NAPPA was described by Ramachandran et al. (2008), where planar arrays consisting of up to 1,000 proteins were generated and later Wong et al. (2009) adapted the NAPPA strategy on a bead-based array platform. One of the important concepts offered by the NAPPA strategy is the feasibility of storing an array with the DNA template for a long term until usage. In turn, the protein spots (or beads) are *per se* not "pure" since they remain co-localized with the DNA template and the antibody against the tag.

A third *in situ* protein array strategy was described by Tao and Zhu (2006), relying on the concept of fabricating protein arrays by capturing nascent polypeptides during translation by puromycin. Here, cell-free expression takes place on streptavidin-coated arrays with immobilized *in vitro* transcribed mRNA, which is hybridized to an oligonucleotide modified with biotin and puromycin. Upon *in situ* translation, puromycin captures the nascent protein facilitating its immobilization on array surface and removal of mRNA by RNase leaves pure protein spots. Yet, as later discussed by Sutandy et al. (2013), the strategy has not flourished due to difficulties to express proteins with a MW over 60 kDa, as well as generally low protein yields. A fourth *in situ* protein array production strategy, the so-called DNA array to protein array (DAPA) was described by He et al. (2008a) and recently optimized further (Schmidt et al., 2013). The concept behind DAPA strategy involves two slides and a permeable membrane in between. One slide is immobilized with PCR-generated DNA constructs encoding tagged proteins, the other slide is immobilized with tag-capturing reagents, and between the two slides the cell-free protein synthesis is carried out within a filter membrane. The newly synthesized proteins on the first slide diffuse through the membrane and become captured on the second slide. The main advantage offered by the DAPA strategy is the possibility to reuse the same DNA array up to 20 times to generate multiple copies of a protein array (He et al., 2008a). In comparison to NAPPA, DAPA also allows to generate "pure" protein spots on a separate surface than the DNA template. Yet, diffusion has been regarded as a potential limitation to synthesize multimeric proteins. Similarly, the protein spots occupy a large area due to diffusion effects; the so-far reported spot densities for DAPA strategy have been in the range of 100 (Stoevesandt et al., 2011).

Taken together, the *in situ* array production strategies in principle avoid the need to express, purify, and store individual proteins, thus also reduce the concerns about the storage stability of protein arrays. When mammalian cell-free protein synthesis systems are utilized, they also offer for an analysis of proteins with post-translational modifications.

More detailed overview of the *in situ* protein array strategies have been provided by He et al. (2008b), Chandra and Srivastava (2010) and Stoevesandt et al. (2011).

The more conventional approach of fabricating protein arrays is based on expression and purification of proteins individually, followed by immobilization on a solid support. While this is a challenging task in several aspects, the feasibility of fabricating arrays with thousands of proteins have been demonstrated by the earliest pioneering studies in the affinity array field. Zhu et al. (2001) demonstrated the practical feasibility of fabricating an array containing approximately 80% of the yeast proteome by cloning, expressing and purifying almost 6,000 yeast open reading frames as GST-His_6 fusion proteins. Using the same strategy, Chen et al. (2008) demonstrated the feasibility of fabricating a whole-proteome array with over 4,200 *E. coli* proteins, representing 99.3% of the *E. coli* genome. This approach originally described by Zhu et al. (2001) was commercialized by Proteometrix Inc., subsequently acquired by Invitrogen™ (now Life Technologies), which is using GST fusion proteins in its ProtoArray® product line. The current 5.0 version of the ProtoArray® Human Protein Microarray hosts over 9,000 full-length human proteins expressed in insect cells. Similarly, in one of the early protein array studies, Lueking et al. (2003) fabricated arrays with 2,413 full-length human proteins expressed in *E. coli* utilizing a human fetal brain cDNA expression library. This approach was later commercialized as the UNIarray® platform by Protagen Diagnostics. More recently, Jeong et al. (2012) reported an array with 16,368 human open reading frames as GST-His_6 fusion proteins expressed in yeast, representing 12,586 unique genes, thus 60% of the human protein-coding genes. The latter framework has been recently commercialized by CDI Laboratories as HuProt™ Human Proteome Microarray, containing 19,832 full-length human proteins in its version 2.0, corresponding to over 75% of the human protein-coding genes.

The recombinant human protein fragments generated within the Human Protein Atlas represent yet another approach, demonstrating the feasibility of large-scale expression and purification of human antigens. Currently, Human Protein Atlas hosts approximately 46,000 sequence-verified human protein fragments, representing approximately 80% of the human protein-coding genes. As I described in section 2.3.1, these 50-150 amino acid long fragments are expressed in *E. coli* as recombinant fusion human protein fragments, which are subsequently utilized within the Human Protein Atlas pipeline to generate planar antigen arrays. These arrays consist of random sets of 384 human protein fragments and are routinely utilized to characterize the binding specificity of newly generated Human Protein Atlas antibodies (Nilsson et al., 2005, Sjöberg et al., 2012). As demonstrated in **Article I**, the collection of these protein fragment arrays offers a valuable protein array resource. More recently, protein fragment arrays with a larger content have been fabricated as well. These arrays hosting 21,120 human protein fragments and representing 12,412 protein encoding genes were for instance utilized within **Article II**.

Peptide Arrays

In addition to proteins or protein fragments, antigen arrays can be generated also by utilizing *synthetic peptides*. The main disadvantage of synthetic peptides compared to full-length proteins or protein fragments is that they are restricted to mimic discontinuous epitopes and allow to assess continuous epitopes. Yet, peptides offer certain advantages such as they might be relatively inexpensive to synthesize; they are physically and chemically more stable than proteins thus offer more stability e.g. during storage and they offer incorporation of certain post-translational modifications such as citrullination or phosphorylation. Similarly, they allow for incorporation of non-natural amino acids and other modifications for detection or immobilization, such as biotin or a histidine tag, which can be introduced at any required position.

The emergence of peptide arrays dates back earlier than the emergence of protein arrays. In fact, peptide arrays might even be considered to designate the emergence of the array technology. The foundations for peptide arrays were laid when Robert Bruce Merrifield (1921–2006) introduced the concept of solid phase peptide synthesis (Merrifield, 1963) and automated peptide synthesizers (Merrifield et al., 1966). Adopting the solid phase peptide synthesis concept, Geysen et al. (1985, 1984) introduced the early concept of peptide arrays by parallel synthesis of peptides on plastic pins. Shortly thereafter, the so-called SPOT™ synthesis approach was introduced by Frank (1992), Frank and Döring (1988), which demonstrated the possibility of *in situ* synthesis of peptides on cellulose membranes. The principle of SPOT™, applied in a further miniaturized way, is still utilized in currently available commercial peptide array platforms such as PepSpot™ or PepStar™ provided by JPT Peptide Technologies.

Around the same time as the emergence of the SPOT™ approach, Fodor et al. (1991) reported another breakthrough *in situ* approach by combining the solid phase peptide synthesis concept with semiconductor fabrication concept. Here, peptide arrays were generated on glass slides using combinatorial synthesis based on photolithography. This method relied on synthesis of peptides by repeated cycles of coupling of activated amino acids and photodeprotection with a photomask. The strategy described by Fodor et al. (1991) was in fact adopted for photolithographic synthesis of high-density oligonucleotide arrays and this technology laid the foundation for the company Affymetrix Inc. (Lipshutz et al., 1999). Regarding the synthesis of peptides, requirements such as for individual synthesis of amino acid monomers with photolabile protection groups led to further modifications of the photolithographic peptide synthesis approach. For instance, Singh-Gasson et al. (1999) replaced the photomasks with digital micromirrors, which has been utilized also in recent studies, such as by Shin et al. (2010) and by Buus et al. (2012). A similar approach to the one described by Buus et al. (2012) was recently utilized by Roche NimbleGen Inc. to fabricate an array with 2.1 million overlapping peptides representing all

human protein coding-genes (Forsström et al., 2014). Mask-free photolithographic methods have been also developed, such as by Gao et al. (2003), Pellois et al. (2000, 2002), who fabricated peptide arrays on glass slides by photogenerated acid-based photolithography. Using mask-free photolithography, Price et al. (2012) demonstrated the feasibility and utility of peptide arrays using microprocessor-grade silicon wafers as solid support. In addition to these strategies, particle-based peptide array fabrication methods have been introduced as well. Relying on the basic principle of SPOT™, Beyer et al. (2007) developed a combinatorial synthesis method utilizing electrically charged amino acid particles positioned on a solid support by electrical field which is generated either by a computer chip (Beyer et al., 2007) or a laser printer (Stadler et al., 2008). This technology is currently commercialized by PEPperPRINT, providing PEPperCHIP® Peptide Microarrays.

The methods outlined so far exemplify the *in situ* peptide array synthesis approaches and a more detailed overview has been provided e.g. by Gao et al. (2004), Henderson and Bradley (2007), Breitling et al. (2009), Andresen and Grotzinger (2009) and Katz et al. (2011). These approaches currently allow to generate high or even ultra-high density peptide arrays. Certainly, it is also possible to fabricate peptide arrays by immobilizing pre-synthesized peptides. This approach is more suited to generate arrays with a more limited content geared towards more focused investigations, such as in **Article III** which aimed to analyze antibody binding to citrulline- and arginine-containing peptide pairs. Creating arrays with pre-synthesized peptides also allows to utilize peptides upon assessment of their identity and purity, which is not feasible when generating peptide arrays *in situ*.

4.3.1 Technical Challenges and Considerations

> Our circulating antibodies tell the story of what our immune system has seen and how it responded. To extract some of this information using antigen microarrays, we need to decide which epitopes we use for probing and how to detect the binders.
>
> *Prechl et al. (2010)*

Some of the technical and analytical aspects I have discussed with regard to antibody arrays, such as analytical sensitivity, are also relevant for application of antigen arrays for antibody profiling purposes. Indeed, the earliest applications of signal amplification strategies such as RCA were applied on antigen arrays for detection of allergen-specific IgE binding (Wiltshire et al., 2000). Similarly, prior to autoantibody profiling applications

purification of immunoglobulins by means of protein A/G can be considered as a strategy to enrich for immunoglobulins in body fluid samples. Yet, systematic evaluations addressing the effect of such pre-analytical steps on array-based downstream analyses have so far not been reported in detail.

Certainly, the most important consideration for antigen array applications is the type and source of the antigens. Bacterial expression systems, such as *E. coli* are extensively used and optimized for production of human proteins or fragments thereof, as also demonstrated within the Human Protein Atlas. However, eukaryotic expression systems are needed when aspects such as post-translational modifications, native folding and activity need to be addressed. Certainly, the consideration regarding the type of expression system applies also when *in situ* array production strategies are adopted. As a side note, folding and activity of proteins expressed in eukaryotic systems might still be compromised by adsorption or covalent attachment of the proteins to the surface of utilized solid support, which highlights the importance of testing different array surfaces and related parameters for downstream applications such as autoantibody profiling (Balboni et al., 2008). Such changes in antigen structure might affect discontinuous epitopes that might be of relevance, but it can also expose epitopes that otherwise are not accessible to antibodies as recently underlined by Wang et al. (2013). In line with this and as I will discuss in the next section, there is no established ultimate strategy in terms of the type and source of the employed antigens, which can be utilized to explore the diversity of the autoantibody repertoire. There are studies demonstrating not only the value of employing full-length proteins expressed in eukaryotic systems, but also the value of recombinant protein fragments (as demonstrated in **Article I** and **Article II**), synthetic overlapping peptides (Price et al., 2012) or even random sequence peptides (Stafford et al., 2014) or peptidomimetics (Reddy et al., 2011), each with their inherent biases and advantages. Thus, initial discoveries made with a given type of antigen can be re-evaluated and verified using antigens of alternative types and sources.

As highlighted by Prechl et al. (2010), for antigen array applications aiming to profile the autoantibody repertoire in body fluids, the consideration regarding what to detect is as important as what to immobilize as antigens. In a typical autoantibody profiling assay, antigen arrays consisting of proteins, protein fragments, peptides or similar are incubated with body fluid samples and autoantibody binding is detected using fluorescent dye labeled secondary antibodies specific for human immunoglobulins. A majority of autoantibody profiling investigations focus to determine antigens recognized by autoreactive IgG antibodies only. Yet, the feasibility and potential value of detecting subclasses of IgG (Graham et al., 2006, Papp et al., 2008a, Robinson et al., 2002) or different immunoglubulin classes have been demonstrated (Hagedorn et al., 2010, Papp et al., 2008b). While these and similar investigations utilized planar arrays, **Article III** aimed to develop a bead-based assay allowing for a parallel analysis of antibodies of IgG, IgM and IgA classes.

Especially in such applications, but also when analyzing only a single autoantibody class, the specificity of the secondary detection antibody might be a factor to consider. For instance, regarding the choice for secondary reagents to detect IgG autoantibodies, reagents with specificity against both heavy and light chain or with specificity against only the Fc portion of the IgG heavy chain can be utilized. Reagents specific against both heavy and light chain might offer a broader epitope recognition than Fc fragment-specific reagents, but they might react with other immunoglobulin classes since they share the same light chain. Besides, depending on the quality of the reagent, Fc fragment-specific secondary antibodies might still cross-react with other immunoglobulin classes. Therefore, secondary reagents from different vendors with different specificities should preferably be evaluated before a large-scale analysis of clinical samples. Besides, as demonstrated in **Article III**, when serum samples at a low dilution rate are utilized for autoantibody profiling applications, captured serum immunoglobulins building an antigen-antibody complex on the array surface might activate the complement system, which might in turn mask the detectability of captured immunoglobulins by secondary reagents. Thus, prior to a large-scale analysis of serum samples, it might be beneficial evaluate different sample dilution rates and the need to block the complement system by means of suitable assay buffer additives such as EDTA.

Antigen array-based autoantibody profiling approaches using planar arrays mostly adopt a single-color approach, where autoantibodies are detected using fluorescent dye labeled secondary antibodies. Dual-color approaches have also been described for array-based autoantibody profiling approaches. Hartmann et al. (2008) utilized a dual-color approach for detection of both total immunoglobulin and antigen-specific autoantibodies. Kattah et al. (2006) described the co-incubation of two different samples labeled with fluorescent dye-labeled Fab fragments. Approaches like the latter might allow to incorporate a reference sample for internal normalization, which can be a preferable strategy when analyzing for instance longitudinal sample collections. As exercised in **Article I**, when the utilized antigens have a tag, a dual-color approach can also be used to involve a secondary antibody to detect the tag and this information can be utilized both for spot alignment and for normalization purposes.

Upon identification of targets on autoantibodies, competition assays can add valuable information regarding the degree of autoantibody specificity. Here, an assessment of the reduction in antibody binding upon pre-blocking of samples with the identified target would provide support for autoantibody specificity. This can be an especially important point when proteins or protein fragments are expressed as fusion proteins in bacterial expression systems, since body fluid samples from a subset of individuals might contain antibodies reacting e.g. with the expression tag. In line with this, depending on the antigen type and source, a pre-blocking of samples with the expression tag or a lysate originating from the expression system should ideally be evaluated prior to analysis of

large sample collections, as applied in **Article I** and **Article II**. A similar consideration is relevant when utilizing peptide arrays for autoantibody profiling purposes. In several peptide array applications, peptides are generated with a C- or N-terminal biotin and a linker molecule, which allow to immobilize peptides on e.g. neutravidin-coated planar or bead-based surfaces. Yet, as experienced in the investigation in **Article III**, body fluid samples from a subset of individuals might contain antibodies reacting with neutravidin, which might require a pre-blocking of samples.

4.3.2 Applications of Antigen Arrays for Biomarker Discovery

Antigen arrays, in the form of protein or protein fragment arrays, constitute an important tool to characterize the binding specificity of antibodies or other affinity reagents, as demonstrated e.g. by Jeong et al. (2012) using recombinant full-length proteins and by Nilsson et al. (2005), Schwenk et al. (2007) and Sjöberg et al. (2012) using recombinant protein fragments. Here, peptide arrays allow to further identify the epitopes either on planar (Chiari et al., 2005, Forsström et al., 2014, Poetz et al., 2005) or bead-based arrays (Hjelm et al., 2010, Larsson et al., 2009) utilizing overlapping peptides. As I will discuss here, antigen arrays also offer a multiplex and high-throughput technology for the investigation of the immune response in the context of several diseases and conditions. As reviewed e.g. by Harwanegg and Hiller (2005) and Shreffler (2011), antigen arrays of allergens can be utilized to characterize the immune response in allergy. Similarly, as reviewed by Vigil et al. (2010) and Natesan and Ulrich (2010) and exemplified in works by Zhu et al. (2006), Crompton et al. (2010), Vigil et al. (2011) and Price et al. (2013b) among several others, antigen arrays of pathogens can be utilized to characterize the immune response within various infectious diseases, which might help to develop vaccines or monitor vaccine immunogenicity. By utilizing antigens representing human proteins or peptides, autoimmune reactions against self-antigens can be studied in the context of several diseases, including autoimmune diseases, as well as cancer. In the following discussion I will mainly focus on this application area of antigen arrays for autoantibody profiling.

Following the emergence of protein microarray technology, the first demonstrations of the utility of antigen arrays for autoantibody profiling approaches were reported by Joos et al. (2000) and by Robinson et al. (2002). Joos et al. (2000) generated arrays with 18 recombinant proteins well-known to be targets of autoantibodies in autoimmune diseases such as Sjögren's syndrome or rheumatoid arthritis and demonstrated autoantibody binding in serum samples. Similarly, Robinson et al. (2002) created arrays with 1,152 features representing 196 different well-known autoimmune targets including peptides, proteins and other molecules which were analyzed with serum samples from patients with autoimmune rheumatic diseases. Joos et al. (2000) observed detection sensitivities

comparable to ELISA and Robinson et al. (2002) reported four- to eight-fold improved analytical sensitivity than ELISA.

Following these pioneering demonstrations of autoantibody profiling applications and reviews highlighting their potential (Balboni et al., 2006, Fournel and Muller, 2003, Hueber et al., 2002, Robinson et al., 2003b, Utz, 2005), antigen arrays have been used in several hypothesis-driven studies, where selected sets of proteins or both proteins and peptides were utilized to generate antigen arrays. Feng et al. (2004), Hueber et al. (2005) investigated rheumatoid diseases; Fattal et al. (2010), Li et al. (2005), Price et al. (2013a) investigated systemic lupus erythematosus; Balboni et al. (2013) investigated juvenile dermatomyositis and Quintana et al. (2008, 2012) analyzed serum and paired serum-cerebrospinal fluid samples in the context of multiple sclerosis, respectively. Besides these autoimmune conditions, Britschgi et al. (2009) utilized peptide-based antigen arrays within Alzheimer's disease; Gnjatic et al. (2009) applied antigen arrays within non-small cell lung cancer and Hagedorn et al. (2010) investigated serum samples from patients with chronic rejection of lung transplantation. Recently, the utility of high-density peptide arrays for autoantibody profiling applications has been also demonstrated e.g. by Price et al. (2012) or Hecker et al. (2012) in the context of systemic lupus erythematosus and multiple sclerosis, respectively.

There were also non-targeted, namely hypothesis-free, approaches exploring the autoantibody repertoire in various diseases. Adopting the strategy originally described by Zhu et al. (2001), arrays with 1,058 full-length liver proteins were generated by Hu et al. (2007), which were later extended to contain a total of 5,011 human proteins and utilized in the context of autoimmune hepatitis (Song et al., 2010). Similarly, Hudson et al. (2007) utilized arrays with 5,005 full-length human proteins to identify tumor-associated antigens within ovarian cancer. The strategy of utilizing human fetal brain cDNA expression libraries to generate antigen arrays, originally described by Lueking et al. (1999, 2003), was also applied in the context of dilated cardiomyopathy (Horn et al., 2006) and the autoimmune condition alopecia areata (Lueking et al., 2005), where arrays with 37,200 redundant, recombinant human proteins were utilized. This setup, later commercialized by Protagen Diagnostics, and containing 3,101 proteins or protein fragments was utilized within multiple sclerosis for autoantibody profiling in cerebrospinal fluid (Beyer et al., 2012).

The investigations in **Article I** also exemplify the hypothesis-free application of in-house generated antigen arrays, where a total of 11,520 human protein fragments representing 7,644 protein-encoding genes were utilized for autoantibody profiling in a multiple sclerosis-related sample collection. Similarly, in the fourth investigation listed under related articles of this thesis, a hypothesis-free approach was taken. Here, 6,528 protein fragments representing 5,447 protein-encoding genes were utilized for autoantibody profiling in a longitudinal plasma sample collection of pancreatic cancer patients who

survived over five years after hematopoietic stem cell transplantation. All the studies mentioned so far utilized antigen arrays in a planar-array format, implying that bead-based antigen arrays have so far not been widely used for autoantibody profiling applications. In **Article I** and **Article II**, the utility of bead-based antigen arrays for autoantibody profiling in large sample collections was demonstrated.

Within the last few years, a majority of the hypothesis-free autoantibody profiling applications have utilized the commercial ProtoArray®. It has been used in a very broad spectrum of diseases, including chronic renal disease and renal transplantation (Butte et al., 2011, Li et al., 2009); hematological disorders (Mias et al., 2013) and autoimmune conditions such as rheumatoid arthritis (Auger et al., 2009), inflammatory bowel disease(Vermeulen et al., 2011), type-1 diabetes (Koo et al., 2014) and primary Sjögren's syndrome (Hu et al., 2011), where all studies analyzed blood-derived samples except the latter one demonstrating the analysis of saliva samples. Within cancer field, ProtoArray® has been used to identify tumor associated antigens in the context of colorectal cancer (Babel et al., 2009), bladder cancer (Orenes-Piñero et al., 2010) and ovarian and pancreatic cancer (Gnjatic et al., 2010). While these cancer-oriented studies focused on analysis of serum samples, Gunawardana et al. (2009) analyzed ascites fluid in the context of ovarian cancer. ProtoArray® has also been utilized for serum autoantibody profiling in neurological diseases such as Alzheimer's disease (Nagele et al., 2011), Parkinson's disease (Han et al., 2012) and amyotrophic lateral sclerosis (May et al., 2014), whereas Querol et al. (2013) analyzed cerebrospinal fluid in the context of multiple sclerosis. While these studies reported disease-associated autoantibody reactivities, Nagele et al. (2013) profiled serum samples from healthy individuals and highlighted that IgG autoantibodies are ubiquitous in serum.

Antigen arrays generated by utilizing the *in situ* NAPPA strategy, originally described by Ramachandran et al. (2004), have been also frequently implemented for recent autoantibody profiling applications. The first studies utilizing NAPPA had a focus on analysis of serum autoantibodies within breast cancer (Anderson et al., 2008, 2011). More recently, NAPPA has been utilized in the context of autoimmune conditions, including type 1 diabetes (Miersch et al., 2013), anklosing spondylitis (Wright et al., 2012) and juvenile arthritis (Gibson et al., 2012), where the latter study also demonstrated the analysis of synovial fluid samples on NAPPA platform.

Besides protein and peptide arrays, there are also examples of phage arrays which have been applied mostly within the cancer field for the identification of targets of tumor-specific autoantibodies. This approach combines the phage display technology with array technology. As described e.g. by Cekaite et al. (2009), a cDNA library is constructed from mRNA isolated from cancer tissue and the cDNA library is inserted into a phage vector to generate cDNA phage-display library. After several rounds of affinity maturation, the library is enriched for page clones reacting e.g. with serum from patients.

The library of enriched clones is immobilized on solid supports to create phage arrays for further analysis of sample collections. This method has been applied for autoantibody profiling in various cancer types, such as breast cancer (Cekaite et al., 2004); prostate cancer (Wang et al., 2005); lung cancer (Chen et al., 2007, Zhong et al., 2005); ovarian cancer (Chatterjee, 2006) and colorectal cancer (Babel et al., 2011); as well as in some autoimmune conditions such as celiac disease (D'Angelo et al., 2013).

In addition to the use of antigen arrays for identification of autoimmune targets, Papp et al. (2007, 2008b) demonstrated that antigen arrays can also be utilized for a parallel measurement antigen-specific complement activation by detecting the deposited complement C3 fragments on antigen array surfaces. Assessment of complement-activating properties of autoantibodies, as well as their immunoglobulin classes, aims to generate more immune function-related information on antigen-arrays. This strategy, reviewed more by Prechl et al. (2012) was applied e.g. in the context of systemic lupus erythematosus (Papp et al., 2012b). While the originally described strategy by Papp et al. (2007) and later applications utilized planar arrays, the investigation in **Article III** aimed to implement this strategy on bead-based antigen arrays.

4.4 Experimental, Statistical and Methodological Aspects

> Scientific method is what working scientists do, not what other people or even they themselves may say about it. [...] [The working scientist] is not consciously following any prescribed course of action, but feels complete freedom to utilize any method or device which in the particular situation before him seems likely to yield the correct answer. In his attack on his specific problem, he is completely free to adopt any course that his ingenuity is capable of suggesting to him. [...] In short, science is what scientists do, and there are as many scientific methods as there are individual scientists.
>
> *Bridgman (1955)*

In this final section, I would like to reflect on miscellaneous aspects regarding study design for discovery-oriented assays carried out on affinity arrays. I also would like to share some personally collected observations and applied practices within the presented thesis

work regarding experimental design and data and statistical analysis. Certainly, a fruitful outcome of a study depends on the awareness of these, if not several more other aspects.

4.4.1 About Study Design: From Samples to Assays

> Reason has so many forms that we do not know which to choose – experiment has no fewer.
>
> *Michel Eyquem de Montaigne*

For a study aiming to identify differentially abundant proteins or autoantibodies as a potential biomarker candidate for a certain disease, the *choice for control samples* is of utmost importance. While several population-based control samples can be accessed easier from a population register, such controls might not represent the relevant population under the risk of becoming future cases. Hospital controls, especially those with other diseases, might therefore represent a more relevant population, however they might not provide an unbiased representation of a control group (Grimes and Schulz, 2005). Thus, it should be evaluated in close collaboration with experts of the studied disease, whether population controls or hospital controls are more suitable for the intended study outcome. In the initial discovery studies carried out in **Article I** and **Article IV**, hospital controls with other neurological diseases were utilized as controls and compared with multiple sclerosis subtypes, whereas within the follow-up study (**Article II**) with the intention to analyze a large sample collection, population-based controls were analyzed.

Certainly, control samples should ideally be matched in terms of demographic variables such as age and gender, if not for other relevant variables as well (Bland and Altman, 1994). However, this might be difficult for some diseases, e.g. if the sample type is an invasively collected body fluid or if the patients are very young or, which have been the cases in **Article IV** and **Article V**, respectively. Similarly, when several disease subtypes are compared, matching such subtype samples in terms of age can be a challenging task since samples belonging to more advanced stages of the disease are *per se* older, as it has been encountered within **Article IV** and **Article V**. Finally, while one view suggests that one well-selected control group is better than two or more (Wacholder et al., 1992), possibilities to involve well-selected sample collections of a related disease or of another control set might still be worth to consider, especially if this can be allowed by the sample throughput of the utilized array platform.

While there are strategies to statistically determine the theoretically required number of samples per group, such as by the statistical power analysis (Levin, 2011, Skates et al., 2013, Zhou et al., 2012a), analysis of as many samples as the utilized technology allows

for would increase the chances for finding candidates that can be reproduced and verified downstream. However, as recently also highlighted by Wallstrom et al. (2013), statistical expertise should indeed be involved at the early stage of biomarker discovery studies, which is not always practiced or feasible. As wisely expressed by Ronald Aylmer Fisher (1890—1962), "to call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of". If a subset of samples needs to be selected from a larger sample collection to meet the sample throughput capacity of the array platform, then this selection needs to be carried out in close collaboration with experts of the studied disease in order to determine which disease subtypes or age groups might be prioritized in an initial discovery analysis. Certainly, the sample preparation type and condition should meet the requirements of the technical analysis to be carried out. For instance, serum and not plasma samples, which have not been exposed to freeze-thaw cycles prior to analysis are the suitable sample types for measurement of complement activation on antigen bead arrays (**Article III**). More general considerations I have mentioned within section 3.2.1 regarding the choice of body fluid type are also applicable within study design.

Proteomic profiling approaches on affinity arrays very often employ a hypothesis-driven approach for *target selection*. This involves selection of targets upon literature mining, where existing knowledge, such as relation to affected tissue, organ, cellular pathways associated with the disease, as well as mutation screening or expression profiling information on transcript level can be taken into consideration to assemble a target set. This has been the main approach within **Article V**, for instance. However, as also argued for in the review article by Ayoglu et al. (2011), the increasing availability of affinity reagent, as well as antigen collections, can allow for the adoption of hypothesis-free approaches for protein profiling on affinity arrays. Thus, affinity arrays can also be implemented within a true discovery framework, as demonstrated within **Article I** for autoantibody profiling and in **Article IV** for protein profiling in multiple sclerosis, where in the latter study hypothesis-driven and hypothesis-free strategies were combined.

4.4.2 About Experimental and Assay Design: From Assays to Data Points

> The works must be conceived with fire in the soul but executed with clinical coolness.
>
> *Joan Miró*

Basic but effective experimental design elements can help to estimate, isolate and eliminate possible bias in protein profiling experiments. *Randomization* and *replication* are for

instance, two basic but crucial concepts which were proposed as early as 1940′s by Ronald Aylmer Fisher within his work regarding design of experiments (Fisher, 1942).

Randomization of samples protects against introduction of bias, which might be introduced by uncontrollable assay- or instrument-related factors. Especially if assays will be carried out on multiple slides or microtiter plates, samples can be applied in a randomized layout, where even a balanced distribution across age, gender and diagnosis type of samples can be applied by a stratified randomization (Altman and Bland, 1999), as applied within **Article V**, which involved different blood preparation types collected at different clinical sites. Replicating entire experiments can reveal day-to-day, namely inter-assay reproducibility; including replicates of e.g. both sample-free assay buffer and a sample pool, distributed within and across several slides or microtiter plates can reveal intra-assay reproducibility. Utilizing the same sample pool as a replicate in segmented experiments carried out as different batches can reveal the so-called batch effects (Leek et al., 2010), which can be addressed during data processing.

Regardless of the robustness of a given assay protocol, relevant positive and negative control analytes need to be included in each assay. Similarly, depending on the type of antibodies, antigens or samples utilized, even well-established buffer compositions should preferentially be fine-tuned: For instance, in protein profiling assays using mouse monoclonal antibodies, human anti-mouse antibodies (HAMA) might interfere with the assay; similarly, when performing protein profiling assays in samples from patients with certain conditions, such as rheumatoid arthritis, the presence of rheumatoid factors might interfere with the assay unless certain buffer additives are utilized (Kaplan and Levinson, 1999, Kricka, 1999).

Regarding the content of the antibody/antigen arrays, including more than one antibody/antigen generated against/representing different regions of each target might provide a better insight. However, this is certainly restricted by the multiplexing capacity of the array format utilized and the availability of the affinity reagents. Finally, it is beneficial to adapt assay protocols to be compatible with automated liquid handling devices, which are becoming increasingly available and affordable. Randomization, aliquoting and preparation of samples and/or affinity reagents should preferably be carried out on such devices in order to reduce hands-on time and risk of human error. Reducing or eliminating hands-on time as much as possible during assays is an important aspect for a robust outcome especially within large-scale profiling studies on several samples and/or for several targets, where experiments might need to be carried out by different operators at different days.

4.4.3 About Data Analysis and Statistics: From Data Points to Potential Biomarker Candidates

> Conducting data analysis is like drinking a fine wine. It is important to swirl and sniff the wine, to unpack the complex bouquet and to appreciate the experience. Gulping the wine doesn′t work.
>
> *Wright (2003)*

Data processing and analysis usually gets less attention than the aspects regarding study and experimental design. Ironically, this stage of especially large-scale studies can require equal, if not even more, amount of time, effort and planning as generating the data itself. Although the discussion about data analysis and statistical tools applied for affinity array data would be complex enough to fill another thesis, I will briefly summarize some basic aspects.

While the assay read-out for bead-based arrays can be obtained directly from the analytical instrument, there is an intermediate image analysis phase in order to obtain assay read-out for planar array formats. Depending on the number of samples analyzed, the image analysis phase can be very labor-intensive when utilizing planar arrays. It ideally requires a visual inspection of the entire topography of each array surface in order to identify and exclude array features affected e.g. by dust specks or scratches before overlaying a grid and extracting the read-out from an image analysis software.

Once data is obtained, there are several basic inspections which can be made to *assess the quality* of data, such as calculation of coefficient of variation (CV) values for the technical replicates included in the assay or visualization of distribution of data points in histograms. The latter implies whether data points should be transformed (e.g. by logarithmic transformation) if common statistical methods making assumptions on normal distribution would like to be later applied (Bland and Altman, 1996). As discussed e.g. by van den Berg et al. (2006) and Hamelinck et al. (2005), there are no universally applicable strategies for data transformation, scaling and normalization. In order to identify the most suitable one, different strategies and their combinations might need to be evaluated in each study having its unique composition of sample collection and its own research question. Here, the suitability of each strategy can be evaluated e.g. by their impact on technical CV values. Within the large-scale protein profiling works presented in this thesis, namely within **Article IV** and **Article V**, the so-called probabilistic quotient normalization (PQN), originally described by Dieterle et al. (2006), was adopted. The suitability of this normalization method within antibody array-based protein profiling

studies was previously assessed by Schwenk et al. (2010a) and Kato et al. (2011), where it was demonstrated that PQN successfully accounts for possible technical dilution artifacts. On the other hand, normalization by PQN was not considered as an equally indispensable step for the type of data generated within autoantibody profiling studies (**Article I-II**).

Before applying statistical methods on the data for comparative analyses, it is a preferred practice to apply *exploratory data analysis steps* to investigate the presence of any systematic variation in the data set, which might not be disease but sample-related or might have been introduced during the experimental procedure. Here, principal component analysis (Ringnér, 2008) and unsupervised hierarchical clustering (Eisen et al., 1998) are two useful tools to asses e.g. the presence of sample type, sample batch or sample origin effects. These effects were for instance identified within **Article IV** and **Article V**. Such analyses can also identify possible artifacts introduced during experimental procedure, such as position effects within a microtiter plate or on a slide. These analysis tools furthermore allow to decipher outliers in the data set, which might be necessary to be excluded before further statistical analysis (Serneels and Verdonck, 2008, Shieh and Hung, 2009).

Prior to any statistical comparative analysis, it is also important to assess whether there is any systematic variation between case and control groups. Such variation might either be introduced due to differences in the pre-analytical chain of sample collection and storage, or it might also reflect a true biological difference in the overall protein content of case and control samples due to disease pathology, as demonstrated for CSF in the context of multiple sclerosis (Häggmark et al., 2013). Such skewed trends can be reflected by the presence of systematically positive or negative fold changes. This can be traced more directly in experiments where target selection has been made in a hypothesis-free manner but also in highly multiplex assays, where stochastically not every single target is expected to be differentially abundant.

In biomarker discovery-oriented applications using affinity arrays, the ultimate aim is to identify the "features", namely the antibodies or antigens, which reveal differences between the compared sample groups. The identified features should be statistically validated using either the available data set, or preferably using a data set generated for a new sample collection. The *feature selection* methods utilized for this aim can be divided into univariate and multivariate methods (Christin et al., 2013). For a univariate feature selection from data generated on affinity reagent arrays, namely for protein profiling experiments, the non-parametric Wilcoxon rank sum test (or the Kruskal-Wallis test for a multi-group comparison) can be applied as alternatives to the parametric Student's t-test (or ANOVA test for a multi-group comparison). Here, the median (or mean) for a feature is considered to differ significantly when the hypothesis testing statistics is smaller than the set value for α, which is generally set to α=0.01. The data generated using multiplex affinity arrays is though mostly high-dimensional, namely, not one but as many hypothesis tests are performed as the number of targets. This increases the probability of committing a type

I error, namely a false positive error. To address this, multiple testing correction methods (Bland and Altman, 1995, Noble, 2009) can be applied to re-calculate the probabilities.

As demonstrated in **Article I** and **Article II**, for a univariate feature selection from data generated in autoantibody profiling experiments, comparison of autoantibody reactivity frequencies within the sample groups can be more insightful. Here, first thresholds need to be defined for "positive reactivity", which can be set in a sample-specific or a target-specific manner. Once the positive reactivity frequencies are calculated within each sample group, then the non-parametric Chi-square test, or preferably a Fisher's exact test, suitable also for sample group sizes smaller than 20, can be applied upon cross-tabulation of the frequency information (Winters et al., 2010).

It is also possible to utilize a multivariate feature selection strategy, which usually works in concert with a classification method. There are several classification methods, such as partial least squares discriminant analysis (PLS-DA) (Eriksson et al., 2006) applied within **Article I**; support vector machines (Ben-Hur et al., 2008); logistic regression applied within **Article IV** and **Article V**; classification trees and ensemble classifiers such as random forest (Breiman, 2001). Once features have been selected either using a univariate or a multivariate strategy, their performance can be estimated by a cross-validation, ideally using new samples. Alternatively, the dataset can be split into training set and test set parts and methods such as leave-one-out or k-fold cross-validation can be utilized to asses the performance of a classifier. The performance measure of a classifier is given by its diagnostic sensitivity and diagnostic specificity, which are usually plotted together in receiver operating characteristic curve (ROC) and the performance is reported by the area under the curve (AUC). An extended discussion regarding various statistical data analysis tools and strategies relevant for clinical proteomics has been by provided by Smit et al. (2008). It should be highlighted that each dataset generated within the context of a certain disease has different characteristics and there is no universal data analysis strategy applicable within all studies.

The identified features as outcome of a hypothesis-free investigation, can be subjected further for a biological interpretation. Here, functional annotation tools, such as DAVID (Huang et al., 2009), can be utilized to assess whether there are gene-ontology (GO) terms significantly enriched for certain molecular functions, biological processes and cellular compartments, as exercised within **Article II**. Furthermore, tools such as STRING (Franceschini et al., 2013) can be utilized to identify whether the identified set of features contains known protein-protein interaction partners, as exercised within **Article I** and **Article V**. Integration of such sources of information might provide additional insights about the potential and relevance of the identified biomarker candidates.

4.4.4 From Potential Biomarker Candidates To Biomarkers

> One never notices what has been done; one can only see what remains to be done.
>
> *Marie Skłodowska-Curie*

Upon identification of potential targets as outcome of an initial discovery on affinity arrays, extensive *technical verification* steps should be taken. Here, reproducibility of the study outcome should be confirmed within and across multiple experiments and preferably using an orthogonal technical platform. As an initial step, different array platforms, namely bead-based or planar arrays can be utilized, as demonstrated within **Article I** in the context of autoantibody profiling in multiple sclerosis.

For protein profiling studies using single-binder assays, affinity reagents from other sources, preferably raised against different epitopes of the same target protein can be collected. Such focused affinity reagent sets should first be characterized for their specificity by means of a Western blot, immunostaining and/or immunoprecipitation and mass spectrometry-based analysis, as well as by epitope mapping to identify the binding sites for each reagent, as demonstrated within **Article IV**. They can be subsequently utilized for a re-analysis of the original sample collection to confirm the initially discovered differential protein profiles. At this step, a simultaneous effort should be put on developing a sandwich assay format resembling a sandwich ELISA format, which still remains the clinical-grade assay type for in vitro diagnostic applications. This in turn means a thorough evaluation of the reagent-driven cross-reactivity for the focused set of affinity reagents, followed by identification of matching antibody pairs for each target by using the information generated during epitope mapping analysis. Furthermore, in order to establish future clinical adaptability of the assay, the utilized affinity reagents should be renewable, namely monoclonal. This, in turn, might require generation of such reagents *de novo*, adding also a time- and cost-intensive dimension.

For autoantibody profiling studies, a similar initial approach can be taken where either protein fragments or peptides representing different parts of the target antigen can be utilized to gain better understanding of the binding site of autoantibodies. Certainly, for initial discoveries made using protein fragments or peptides, an intuitive first step should be to assess whether such autoantibody reactivity patterns can be revealed from a full-length representation of the target protein. It needs to be considered though that producing the full-length version of certain types of proteins, such as transmembrane proteins, might render an additional challenge.

A technical verification stage with the co-development of an assay with improved analytical performance and utilizing clinically relevant reagent types should be followed by a

"biological" verification stage. The main purpose of the *biological verification* stage is to demonstrate the validity of the initially identified protein or autoantibody profiles in new and independent sample collections, preferably from different geographical locations. At this stage, the diagnostic specificity and sensitivity of the initially identified set of targets should be assessed, as well as the positive predictive value and negative predictive value. The latter are measures combining the prevalence characteristics of the given disease with the diagnostic sensitivity and specificity (Füzery et al., 2013).

If successfully accomplished, the steps I have described so far, starting with a single or a small set of *potential biomarker candidates* would lead to technically and biologically verified *biomarker candidates*. At this stage, the intended use of the biomarker candidates should be clearly defined and the degree to which the use of these new biomarker candidates would lead for an improved diagnosis or monitoring should be clear. However, this stage of the biomarker discovery pipeline is also when commercial partners with resources and diagnostic expertise might need to be involved, because access to thousands of samples for design and logistics of extensive studies calls for a considerable amount of human and financial resource. Certainly, the decision making criteria for commercial partners closely, if not solely, depends on factors such as a foreseeable and reasonable return on their investment (Vitzthum et al., 2005). Thus, as I have highlighted within section 3.1, a biomarker candidate is still several steps far from becoming a biomarker. Indeed, this perspective strongly argues for the strategy I have described above, which integrates the discovery and a full characterization of potential biomarker candidates as part of a thorough pre-clinical discovery process. In the following present investigation section, I will briefly discuss four such affinity-array based discovery-oriented investigations within multiple sclerosis and muscular dystrophies, as well as an assay development study, which have been carried out during this thesis work.

Part II

Present Investigations

5 Present Investigations

As the title chosen for this thesis work implies, the investigations included in this thesis are about developing assays for affinity arrays and utilizing them for profiling proteins and autoantibody repertories in body fluids. The overall ambition was to identify potential biomarker candidates for diseases, where there is an unmet need to support disease diagnosis or monitoring.

In Figure 5.1, the five investigations I have selected to include here as an article are summarized. To this aim, I dissected each investigation into six "building blocks": affinity reagents; studied disease and related sample collections provided by a biobank; the sample type; profiling approach; array platform; and objective and methodology. I consider collections of quality-assessed affinity reagents and well-characterized samples as the main enablers of a successful application of affinity arrays in the quest towards potential biomarker candidates. In Figure 5.1, I accordingly depicted samples as the "foundation" of these investigations and the affinity reagents as the "roof". As demonstrated in this thesis work, upon availability of affinity reagents and samples, several investigations can be carried out, addressing different aspects of a given disease or applying the same technical approach across different diseases.

The affinity reagents utilized in almost all the presented investigations were generated within the framework of the Human Protein Atlas, which I briefly discussed in section 2.3.1. In **Article I** and **Article II**, recombinant human protein fragments and in **Article IV** and **Article V**, affinity purified polyclonal antibodies were used. The assay development study presented in **Article III** made use of mainly short synthetic peptides.

Utilizing these antigens or antibodies, arrays were built, either in a planar or a bead-based format. In **Article I**, several batches of in-house generated planar arrays each with a different antigen content were initially utilized, whereas bead-based antigen or antibody arrays were created to be applied within all investigations including **Article I**. Using antigen arrays, **Article I** and **Article II** aimed to study the plasma autoantibody repertoires, as well as **Article III**, which further aimed to study not one but several immunoglobulin classes for autoimmune reactivity and for their complement activating property. Single-binder assays on antibody arrays were utilized within **Article IV** and **Article V** for protein profiling applications.

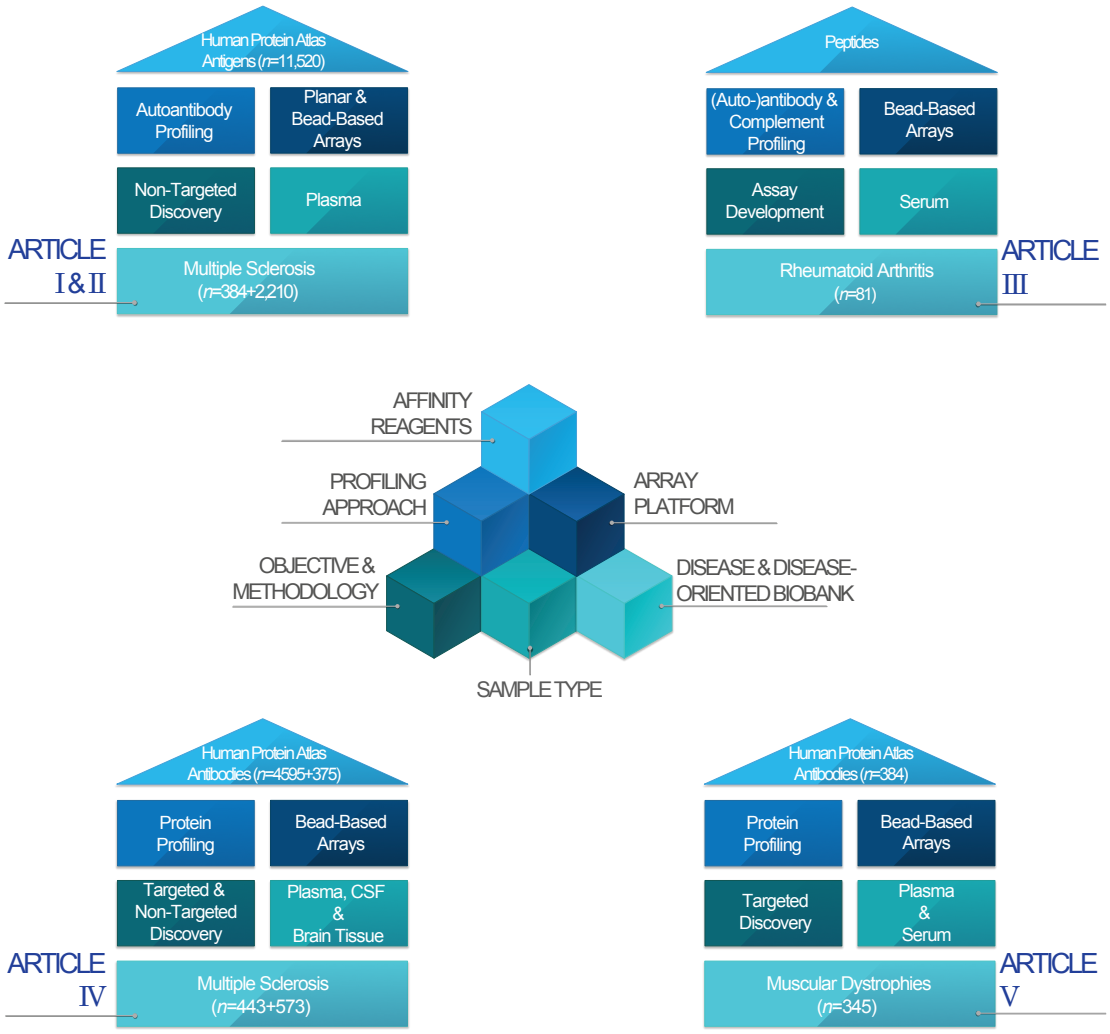

Figure 5.1: Overview of investigations presented in this thesis work.
Six elements can be considered as common "building blocks" for the presented investigations in this thesis. These are i) affinity reagents - antibodies or antigens; ii) studied diseases and access to disease-oriented biobank collections; iii) which provide different sample types such as plasma, serum, cerebrospinal fluid (CSF) or tissue; iv) the profiling approach - protein, autoantibody or complement profiling; v) suitable array platforms for this aim - either planar or bead-based arrays and vi) the methodology. Based on these conceptual building blocks, each investigation has been depicted.

Utilizing antigen arrays, multiple sclerosis-related plasma sample collections were analyzed in **Article I** and **Article II** for their autoantibody reactivity profiles. **Article IV**, on the other hand, investigated protein profiles in a multiple sclerosis-related plasma sample collection by means of antibody arrays. Within **Article IV**, cerebrospinal fluid as well as brain tissue sections were also investigated as additional sample sources. In **Article III**, a rheumatoid arthritis-related serum sample collection was chosen to demonstrate the utility of the developed bead-based antigen array protocol for a co-profiling of autoantibody

reactivity and complement activation. Utilizing antibody arrays, **Article V** investigated both plasma and serum protein profiles within muscular dystrophies.

While the main focus within **Article III** was on assay development, the other articles primarily aimed to identify potential diagnostic biomarker candidates, using a hypothesis-driven (targeted) or a hypothesis-free (non-targeted) approach, as well as both. **Article I** can be considered as a proof-of-concept study, where first assay protocols were developed and fine-tuned for implementation of arrays of human protein fragments for a hypothesis-free autoantibody profiling approach within multiple sclerosis. This entirely hypothesis-free discovery study presented in **Article I** continued as a more focused verification study within **Article II**. **Article IV** also had an entirely hypothesis-free starting point for protein profiling in multiple sclerosis, which was subsequently combined with a hypothesis-driven strategy. **Article V** represents a hypothesis-driven discovery study, where protein profiles for pre-selected targets with potential relevance to muscular dystrophies were generated.

In the following sections, I will provide more insight into each of these investigations, where as a main contributor I performed experimental planning, laboratory work, data analysis, data interpretation and manuscript preparation. The articles describing each of these investigations in detail can be found in the appendix. Here, upon providing a brief background, I will focus on the main aims and most significant findings of each investigation. I will also reflect upon the main possibilities which were exploited, the main challenges which were experienced and the main conclusions which can be drawn.

5.1 Autoantibody Profiling in Multiple Sclerosis using Antigen Arrays – Article I & Article II

Aims of the Investigation

Multiple sclerosis (MS), first described by Jean-Martin Charcot (1825–1893) in 1860′s, is characterized by multifocal demyelination and axonal loss in the central nervous system (CNS) and is the leading cause of chronic neurological disability among young adults (Karussis, 2014, Milo and Miller, 2014). MS has a heterogeneous nature with various subtypes and manifests itself through a wide range of neurological symptoms (Compston and Coles, 2008). As reviewed by Olsson (1992), Steinman (1996), Sospedra and Martin (2005), McFarland and Martin (2007) among several others, there are several indications supporting the autoimmune nature of MS. Based on animal models of MS, autoantibodies against a number of targets, primarily myelin compounds such as myelin basic protein or myelin oligodendrocyte glycoprotein have been shown to be involved in lesion formation, as discussed by Archelos and Hartung (2000), Vincent et al. (1999). In recent years,

several targets other than myelin antigens have been proposed in the context of MS, as recently reviewed by Schirmer et al. (2014). In this regard, antigen arrays have been utilized to investigate the increased diversity of autoantibody responses in MS, presumably caused by epitope spreading mechanisms (McRae, 1995). However, such approaches have mostly utilized dedicated lipid arrays (Brennan et al., 2011, Kanter et al., 2006) or myelin compound arrays (Quintana et al., 2008, 2012, Robinson et al., 2003a), adopting a hypothesis-driven approach.

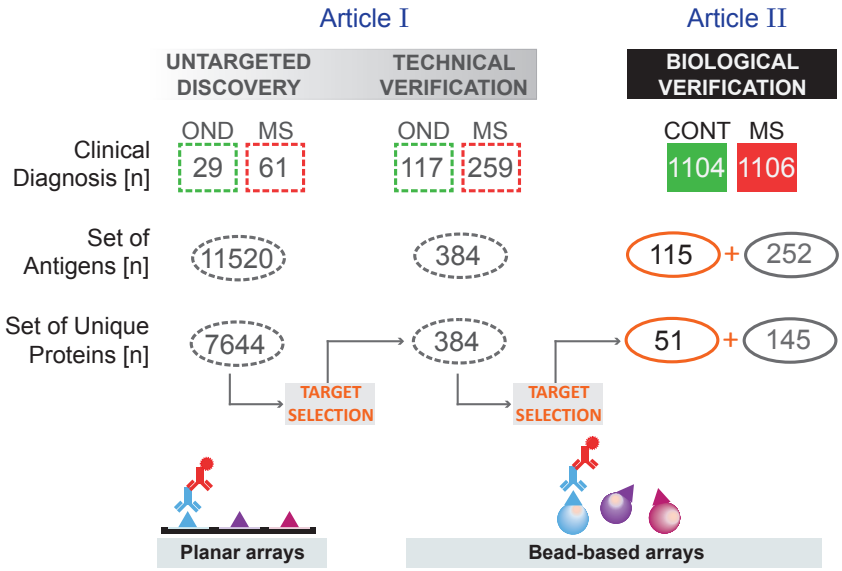

Figure 5.2: Overview of study design within Article I and Article II, where antigen arrays were utilized for autoantibody profiling within multiple sclerosis.
Starting with a hypothesis-free assembled set of 11,520 antigens, **Article I** identified a technically verified set of 51 targets within a representative set of MS-related plasma samples. In **Article II**, these 51 targets were used for autoantibody profiling within an independent plasma sample set of 1,106 MS cases and 1,104 non-diseased controls. In these investigations, recombinant human protein fragments generated within the Human Protein Atlas were utilized as antigens. Figure taken from **Article II**.

Within **Article I**, an entirely hypothesis-free approach was taken to explore the diversity of autoantibody repertoire in MS. As a discovery tool, planar arrays were chosen, which consisted of recombinant human protein fragments generated within the Human Protein Atlas. For this aim, first assay protocols were developed and fine-tuned to utilize the in-house generated antigen arrays for autoantibody profiling purposes. Subsequently, an initial plasma sample collection from 90 individuals was analyzed for IgG reactivity against a total of 11,520 antigens, representing 7,644 unique Ensembl Gene IDs. Upon development and fine-tuning of assay protocols, multiplex bead-based arrays were utilized for an extensive technical verification in an extended collection of 376 samples. As a result, a set of 51 antigens was identified, revealing platform-independent and differential autoantibody reactivity frequencies across various MS subtypes and controls with other

neurological diseases (ONDs). This target set served as the starting point for **Article II**, where using bead-based arrays IgG reactivity against these 51 targets, as well as over 100 additional targets selected upon literature mining, were evaluated in an independent collection of plasma samples from a total of 2,210 individuals (Figure 5.2).

Summary of Findings

Profiling autoantibody repertoires for 11,520 antigens within **Article I** revealed that the number and diversity of plasma autoantibodies varied greatly between individuals and was mainly irrespective of their disease status. As shown in Figure 5.3, there were over 1,500 antigens, which were recognized by plasma autoantibodies of only single individuals. In contrast, there were only a handful of "shared" antigens which were recognized by plasma autoantibodies of a larger proportion of individuals. Thus, one of the main observations of this hypothesis-free investigation was that the autoantibody repertoire in plasma might be under the influence of various individual-driven factors and human plasma presumably hosts autoantibodies toward hundreds, if not thousands, of autoantigens.

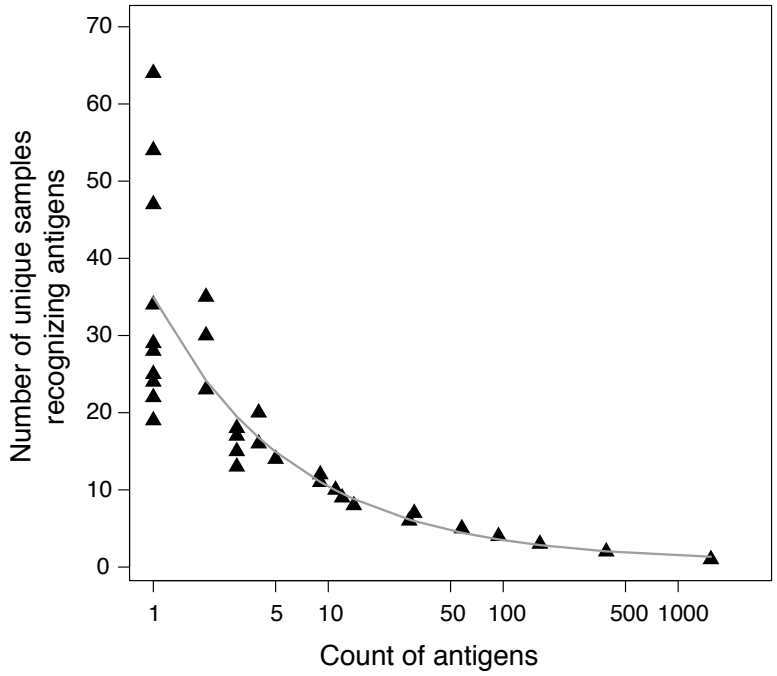

Figure 5.3: Diversity of the autoantibody repertoire as revealed within Article I. In the discovery stage of **Article I**, a representative collection of 90 plasma samples were analyzed for their IgG reactivity profiles against 11,520 antigens. Out of these, 1,539 were recognized in plasma of not more than single individuals, whereas a small number of antigens were recognized in up to 64 individuals. Figure adapted from **Article I** by Ayoglu et al. (2013).

Despite the heterogeneity, a set of 384 antigens were selected for further evaluation within **Article I**, and experimentally verified on a planar array format. This was followed by a technical verification utilizing a bead-based array format, which allowed for an analysis of an extended collection of 376 plasma samples. As outcome of this cross-platform verification approach and competition assays to assess autoantibody specificity, a subset of 51 antigens was revealed with statistically significant differences in reactivity frequency across the MS subtypes and controls with ONDs.

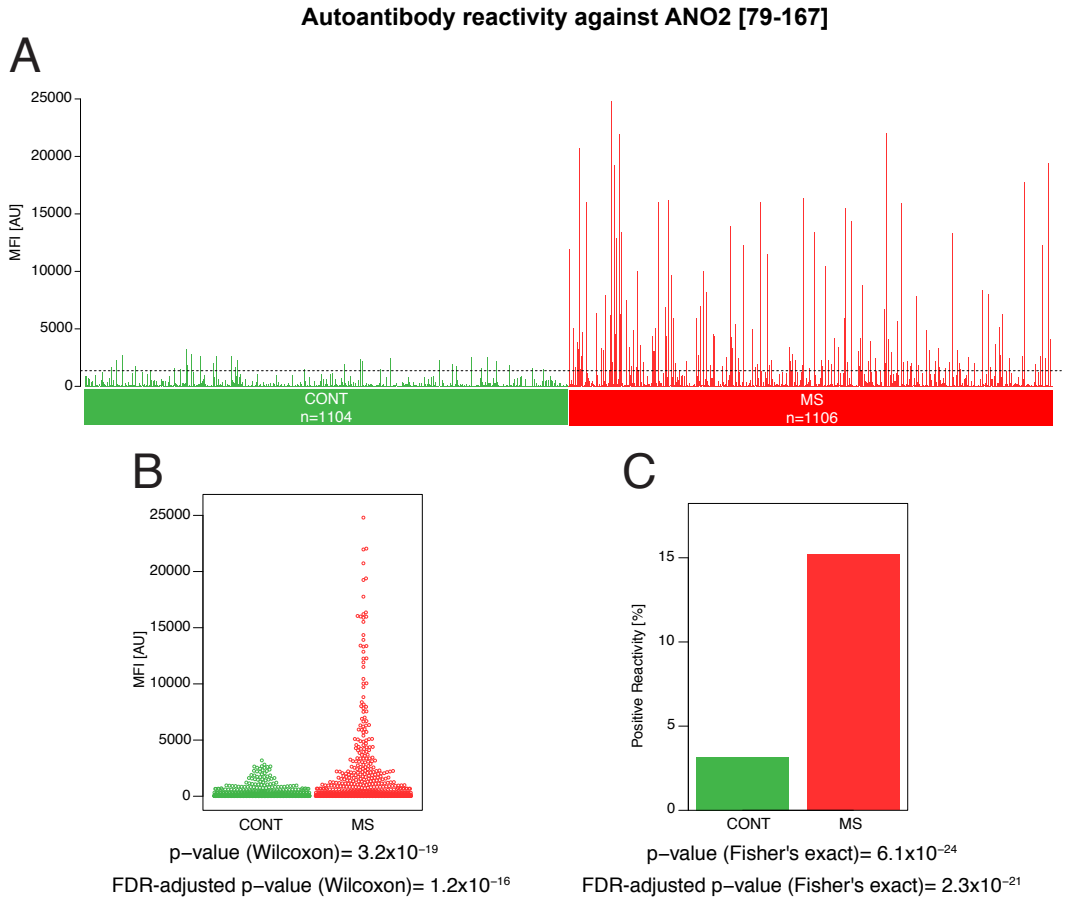

Figure 5.4: Plasma autoantibody reactivity against the protein fragment representing anoctamin 2, identified in Article I and verified in Article II as a potential autoimmune target candidate within MS.
(A) The barplot represents the median fluorescence intensity (MFI) values for plasma reactivity against ANO2 amino acid region [79-167] within 1,104 non-diseased controls and 1,106 MS cases. (B) The dotplot represents the MFI values for plasma reactivity against ANO2 [79-167] within the MS cases and non-diseased controls. The Wilcoxon rank-sum test p-value is reported below the plot. (C) The barplot represents the positive reactivity percentages for ANO2 [79-167] within the MS cases and controls and the respective Fisher′s exact test p-value. Figure adapted from **Article II**.

Article II primarily aimed for a dedicated analysis of autoantibody reactivity against this set of 51 antigens in an independent and much larger collection of plasma samples (Figure 5.2). Comparison of IgG reactivity in plasma from 1,106 MS patients and 1,104 non-diseased controls revealed anoctamin 2 (ANO2), a calcium-activated chloride channel also known as transmembrane protein 16B, as a prominent autoimmune target candidate, which originated from the set of 51 targets identified within **Article I**. A statistically significant difference in positive reactivity between MS cases and non-diseased controls was found in particular for the intracellular, N-terminal region of ANO2 between amino acids [79-167] (Figure 5.4). A preliminary immunofluorescence analysis with antibodies against N-terminal region of ANO2 on human post-mortem MS brain tissue revealed moderate staining in neuronal cell bodies from healthy appearing areas and a clear increase in staining in the form of small cellular aggregates near and inside MS lesions.

The antigen set in **Article II** included also protein fragments representing previously proposed autoimmune targets in the context of MS, such as the potassium channel KIR4.1 (Srivastava et al., 2012). However, the differences revealed for ANO2 outperformed other such targets including KIR4.1. A protein fragment representing interferon beta 1 (IFNB-1) was included as well in the antigen set, which is an immunomodulatory protein and drug used for treatment of MS patients, where however a subset of patients develop antibodies against IFNB-1 and thereby experience reduced therapeutic efficacy (Giovannoni et al., 2002). As shown in Figure 5.5, the reactivity profiles for IFNB-1 protein fragment revealed a significant difference between IFNB-1 treated and untreated cases. This allowed for further identification of IFBN-1 reactivity for a particular drug subtype, namely IFNB-1b type known to be more immunogenic (Hemmer et al., 2005). This finding, in turn, demonstrated the utility of the human protein fragments generated within Human Protein Atlas, being most unique representations of their full-length versions.

Conclusions and Outlook

In **Article I**, a hypothesis-free discovery strategy was adopted to study the plasma autoantibody repertoire in MS on planar arrays consisting of human protein fragments. The potential of this strategy, as well as the utility of human protein fragments as antigens, and the efficiency of using two complementary array platforms for a technical verification has been demonstrated. **Article I** revealed the potentially very diverse nature of the autoantibody repertoire. However, a set of 51 targets could be identified differing in their recognition frequencies among the different MS subtypes. In **Article II**, these 51 targets were evaluated further in over 2,000 samples. Here, reactivity against the protein fragment representing ANO2, a calcium-activated chloride channel, revealed a significant difference between MS cases and non-diseased controls. ANO2 is therefore suggested as a potential autoimmune target candidate within MS.

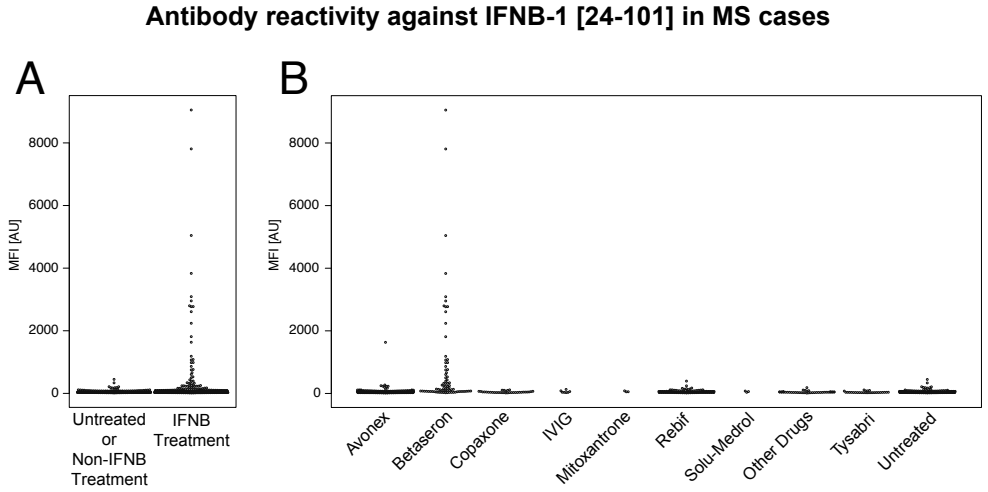

Figure 5.5: Antibody reactivity against the protein fragment representing IFNB-1 in IFNB-1 treated and untreated MS plasma samples analyzed in Article II. (A) The dotplot represents MFI values for plasma reactivity against IFNB1 amino acid region [24-101] within MS cases treated with IFNB1-type drugs (including Betaseron, Avonex and Rebif) and within MS cases which were either untreated or treated with other type of drugs. (B) The dotplot represents the MFI values for plasma reactivity against IFNB1 [24-101] within MS cases dissected further into various treatment categories and drugs. In particular those MS cases treated with IFNB-1b type drug Betaseron revealed presence of binding antibodies against IFNB-1, whereas MS cases with binding antibodies against IFNB-1 were rare among those treated with IFNB-1a type drugs (Avonex and Rebif). Figure adapted from supplementary material of **Article II**.

Ongoing investigations include a continuation of immunostaining analyses on human brain tissue sections; as well as an investigation of the association of the ANO2 reactivity profiles with HLA allele and other risk factor metadata available for the screened sample collection. The protein fragment representing ANO2 has also been utilized for immunization of two different strains of MS animal models, where collected serum sample series will be analyzed utilizing bead-based arrays in addition to histopathological analysis of the brain tissue. In order to determine the binding site of plasma autoantibodies to ANO2, an epitope mapping study utilizing peptide arrays will be carried out. It is a challenging task to express a full-length representation of the transmembrane protein ANO2, yet as a start, the entire intracellular domain will be expressed and a planned analysis of plasma autoantibody reactivity in over 17,000 samples will allow for a further evaluation of ANO2 as a potential autoimmune target in MS. Related future efforts, such as investigation of other related proteins of anoctamin family (Schreiber et al., 2010) or investigation of autoantibody profiles for ANO2 in other related neuroinflammatory conditions, will reveal a more complete picture regarding the relation and contribution of ANO2 to disease pathogenesis in MS, as well as regarding its diagnostic potential.

5.2 Complement Activation on Antigen Arrays: Towards More Functional Antibody Profiles – Article III

Aims of the Investigation

As demonstrated in the previous investigation, antigen arrays represent a powerful tool to identify the targets of autoantibodies or antibodies against infectious agents. Yet, this information can be enriched further by investigating more than one immunoglubulin class -as IgG is mostly the only investigated immunoglobulin class-, as well as by investigating the effector functions of antibodies, including their complement activation properties. Regarding the latter, measurement of antigen-specific C3 fragment deposition offers an efficient approach to monitor the degree of complement activation via all three complement activation pathways described in section 2.1.2. Such assays for a parallel and multiplex analysis of antigen-specific antibody reactivity and complement activation have been previously developed (Papp et al., 2007, 2012a, 2008b) and implemented in the context of autoimmune diseases such as systemic lupus erythematosus (Papp et al., 2012b). However, these previous studies utilized arrays in a planar format, resulting in a lack of established bead-based assay protocols for studying antigen-specific antibody reactivity and complement activation in a single assay. This has been addressed within **Article III**, which primarily aimed to develop an assay for parallel analysis of antibody reactivity and immune-complex induced complement activation on bead-based antigen arrays.

The developed assay workflow outlined in Figure 5.6 was used in the context of *rheumatoid arthritis* (RA). RA is a chronic, inflammatory disease of presumed autoimmune etiology, affecting approximately 0.5-1% of the world population (Silman and Pearson, 2002). Although not exclusive for RA, circulating *rheumatoid factors*, namely antibodies directed against the constant region of immunoglobulins of IgG class, are present in nearly 80% of RA patients (Firestein, 2003). Another group of autoantibodies, targeting proteins such as filaggrin (Simon et al., 1993) or fibrinogen (Masson-Bessière et al., 2001), have been more recently associated with RA. As proposed by Schellekens et al. (1998), these antibodies target epitopes where the amino acid arginine is converted into citrulline as a result of a post-translational modification. Thus, this group of RA-associated autoantibodies are referred to as *anti-citrullinated protein antibodies* (ACPAs). The utility of the developed assay protocol was demonstrated by analyzing a serum sample collection from 41 RA patients and 40 non-diseased controls using a bead-based antigen array. The array consisted of citrulline- and arginine-containing peptide pairs derived from filaggrin and fibrinogen beta chain, as well as other analytes, which allowed for a multiplex profiling of antibody reactivity and complement activation in the context of ACPAs, rheumatoid factors, as well as the viral antigen EBNA-1. Using the developed assay protocol, the levels of IgG, IgM and IgA reactivity, along with their complement activating properties against various antigens could be analyzed in the context of RA.

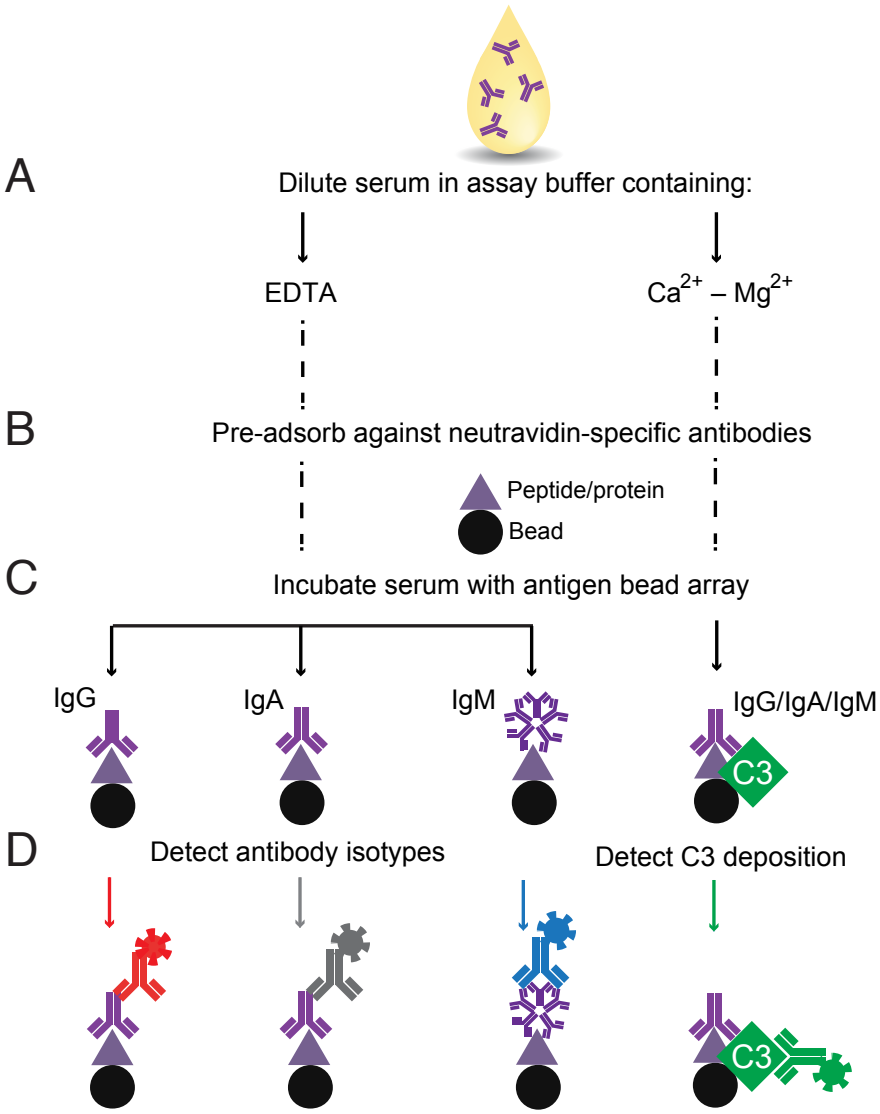

Figure 5.6: Overview of the developed workflow within Article III for a parallel analysis of antigen-specific IgG, IgM and IgA reactivity and their complement activating property.

(A) Serum samples are diluted 1:10, either in a Ca^{2+}-Mg^{2+} containing assay buffer for detection of complement activation, or in an EDTA containing assay buffer for antibody detection. (B) The serum samples are pre-adsorbed against neutravidin-specific antibodies (This step is applicable when the antigen array contains biotinylated peptides, which are immobilized on neutravidin-coated beads.) (C) A mixture of beads coupled to various antigens is distributed into a 384-well plate and the pre-adsorbed samples are added to the bead array. (D) Complement activation driven C3 deposition and the three different antibody classes are detected in parallel with fluorescently labeled secondary antibodies dispensed into individual wells of each quadrant of the 384-well plate. Figure taken from **Article III** by Ayoglu et al. (2014b).

Summary of Findings

Assay optimization efforts within **Article III** initially identified the most suitable serum sample dilution rate allowing for measurement of complement activation. A buffer supplemented with physiologically equivalent concentrations of Ca^{2+} and Mg^{2+} provided the optimal buffer composition for measurement of complement activation. However, this buffer composition did not allow for an equally efficient detection of bound antibodies due to extensive complement fragment deposition, masking in turn the bound antibodies and diminishing their detectability by secondary detection reagents (Figure 5.7). Thus, an assay buffer supplemented with EDTA was identified to suit detection of serum antibodies captured on antigen-coupled beads, as EDTA chelates Ca^{2+} and Mg^{2+} cations needed for complement activation.

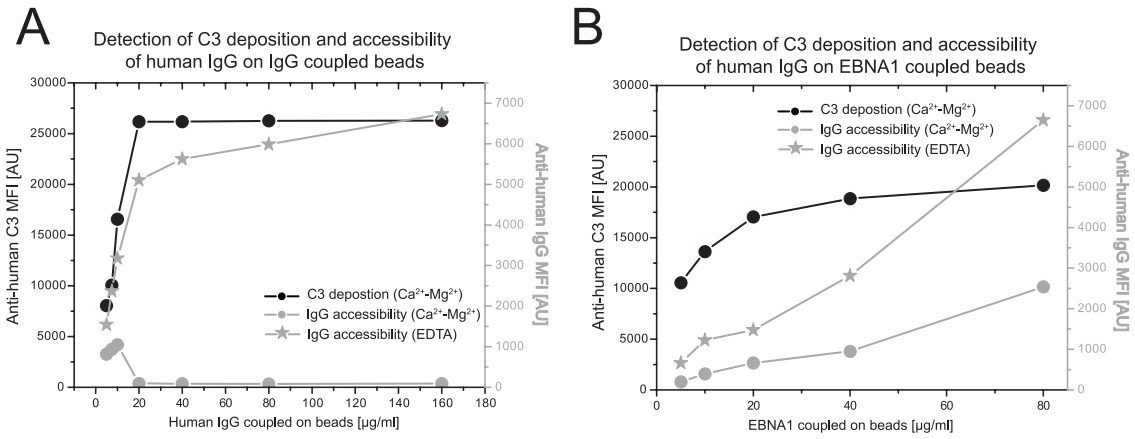

Figure 5.7: Article III revealed the masking effect of complement activation for detection of antibodies.
Beads coupled with human IgG or the viral antigen EBNA-1 were incubated with a serum sample, where the assay buffer was supplemented either with Ca^{2+} and Mg^{2+} (lines with dots) or EDTA (lines with stars). (A) When utilizing the buffer composition promoting complement activation (Ca^{2+}-Mg^{2+} buffer), the detectability of human IgG coupled on beads (grey line with dots) was entirely diminished due to substantial complement activation (black line with dot). (B) Also a remarkable effect on detectability of antigen-IgG complexes was observed when EBNA-1 coupled beads were incubated with a seropositive sample (grey line with dots). These observations indicated the need to utilize an assay buffer supplemented with complement-blocking EDTA for detection of serum antibodies captured on antigen-coupled beads. Figure adapted from **Article III** by Ayoglu et al. (2014b).

Using the developed assay protocol, serum samples from 41 RA patients and 40 non-diseased controls were analyzed. Samples from RA patients revealed the presence of significantly higher levels of IgG-specific IgM and IgA rheumatoid factors, which also contributed for a significantly higher degree of complement activation in serum from RA

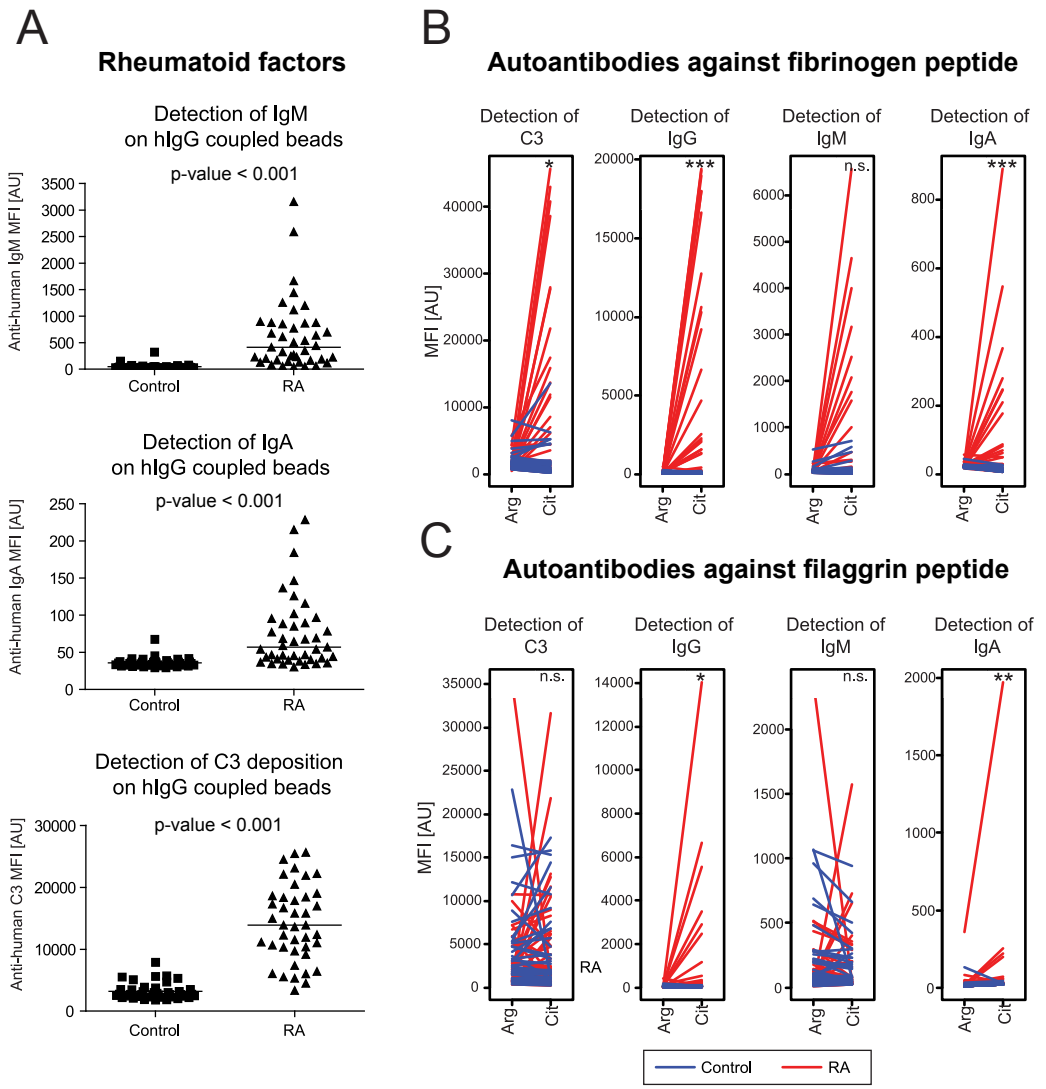

Figure 5.8: Complement activating properties of rheumatoid factors and autoantibodies against citrullinated peptides analyzed using the assay workflow developed in Article IIII.

(A) When incubated with human IgG coupled beads, serum samples from RA patients revealed significantly higher levels of IgG-specific IgM and IgA rheumatoid factors, which in turn resulted in a higher level of C3 fragment deposition in serum of RA patients. (B) Compared to the arginine-containing form, the citrulline containing peptide derived from fibrinogen beta chain revealed higher levels of IgG, IgM and IgA autoantibodies, as well as higher complement activation in serum samples from RA patients (red). The level of IgG and IgA autoantibodies against the citrulline-containing fibrinogen beta peptide and level of complement activation were also significantly higher in RA patients than non-diseased controls. (C) Both arginine- and citrulline-containing peptide derived from filaggrin revealed the presence of IgM autoantibodies, and complement activation by them, in serum from both RA patients and non-diseased controls. However, an IgG and IgA reactivity against the citrulline-containing filaggrin was only detected in serum from RA patients. The significance of difference between RA patients and non-diseased controls for the citrulline-containing forms of peptides is indicated by (***) for p-value<0.001, (**) for p-value<0.01, (*) for p-value<0.05 and (n.s) for non-significant differences. Figure adapted from **Article III** by Ayoglu et al. (2014b).

patients than the non-diseased controls (Figure 5.8-A). The utilized antigen array included citrulline- or arginine-containing 15-mer peptides representing fibrinogen beta chain amino acid region [60-74] and 5-mer peptides representing filaggrin amino acid region [454-458]. These peptide regions were previously demonstrated to be the most dominant epitopes for fibrinogen beta and filaggrin, e.g. by Sebbag et al. (2006) and by Babos et al. (2013), respectively. The parallel analysis of autoantibody classes and their complement activation revealed the presence of IgM autoantibodies against both arginine-containing native form and citrulline-containing form of filaggrin peptide, and a complement activation caused by these autoantibodies in serum samples from both RA patients and non-diseased controls (Figure 5.8-C). Regarding fibrinogen beta, there was a significant difference between RA patients and non-diseased controls in IgA reactivity against the citrulline-containing fibrinogen beta peptide and C3 deposition (Figure 5.8-B). Yet, the IgG reactivity against both the citrulline-containing fibrinogen beta and citrulline-containing filaggrin peptides revealed the most significant differences and they were exclusively observed in serum from RA patients and not in non-diseased controls. These findings illustrated the possibility of measuring antigen-specific antibody reactivity of various classes and antigen-specific complement activation within the same assay and in a multiplex format.

Conclusions and Outlook

This investigation focused on developing an assay protocol for bead-based arrays, where antigen-specific reactivity of different antibody classes against self-antigens or viral antigens, along with their complement activation property can be measured in parallel and in a multiplex format. In the context of RA, availability of tools for a parallel measurement of rheumatoid factor classes (Jónsson and Valdimarsson, 1993), as well as ACPAs (van der Woude et al., 2010) has a clinical relevance. An additional parallel investigation of complement activation might promise an even more complete picture regarding the pathology of diseases such as RA, where autoimmune responses are involved.

The developed assay workflow in **Article III** can certainly be exploited further by analysis of much larger sample collections in the context of several other autoimmune diseases, as well as infectious diseases. Besides, as an alternative to different antibody classes, different subclasses of IgG can be detected, along with an assessment of their complement activating properties. In line with this, investigation of the complement activating properties of the serum autoantibodies against ANO2, the highlighted target within **Article I** and **Article II**, would offer a more functional picture of the ANO2 antibodies within multiple sclerosis. In conclusion, the developed assay protocol and workflow within **Article III** can be regarded as a useful addition to the toolkit of antigen arrays, allowing for potentially more insightful antibody profiles within autoimmune and infectious diseases.

5.3 Protein Profiling in Multiple Sclerosis using Antibody Arrays – Article IV

Aims of the Investigation

As demonstrated in **Article I** and **Article II**, antigen arrays allow for a hypothesis-free exploration of the plasma autoantibody repertoire in multiple sclerosis (MS) and for identification of potential autoimmune target candidates. Antibody arrays offer an equally powerful approach, in this case to investigate plasma protein profiles for identification of targets with a biomarker potential. Yet, antibody or other affinity reagent arrays have so far not been widely utilized in the context of MS. As reviewed by Farias et al. (2014), attempts within MS to identify differentially expressed proteins in plasma or cerebrospinal fluid (CSF) have adopted gel-based and shotgun mass spectrometry-based strategies, relying on an analysis of mostly very limited number of samples. In line with this, **Article IV** aimed to investigate plasma protein profiles in MS generated on antibody arrays, while the sample throughput and multiplexing capacity of bead-based arrays was combined initially with a hypothesis-free screening approach and then with a more targeted analysis.

As summarized in Figure 5.9, using a single-binder assay format first a representative set of plasma samples were profiled by 4,595 antibodies, all generated within the Human Protein Atlas. Next, 56 of these antibodies were combined with antibodies targeting 296 other proteins selected upon literature mining. The resulting array, including antibodies towards proteins both with a hypothesis-free and hypothesis-driven origin, was utilized to analyze plasma samples from 172 individuals. This set included samples from patients with all subtypes of MS, such as relapsing remitting MS (RRMS) and secondary progressive MS (SPMS), as well as the patients exhibiting the early form of MS called clinically isolated syndrome (CIS), patients with other neurological diseases (ONDs) and ONDs with signs of inflammation (iOND). The potentially interesting targets revealing differences mainly between patients with CIS and SPMS were assembled in a new array with 101 antibodies. Using this focused antibody array, an extended collection of 443 plasma samples, as well as a collection of 573 CSF samples from a majority of the plasma sample donors were analyzed. Finally, a refined list of five antibodies were investigated on human brain tissue for their staining patterns.

Summary of Findings

The initial hypothesis-free screening approach, followed by analysis with two smaller and more targeted antibody arrays revealed differences in plasma for five protein profiles (Figure 5.10-A). These were interferon regulatory factor 8 (IRF8), zinc transporter solute

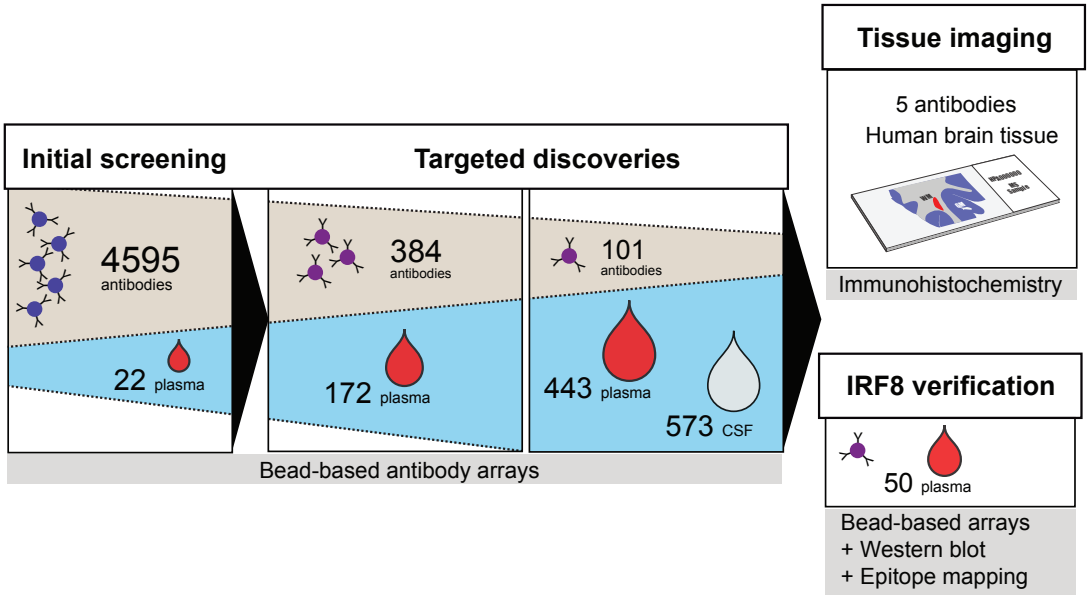

Figure 5.9: Overview of the protein profiling approach within Article IV, where different sample types were analyzed in the context of multiple sclerosis.
Using bead-based arrays and adopting a hypothesis-free approach, 3,450 unique proteins were profiled in a representative set of 16 plasma samples from MS cases and 6 from non-diseased controls. Subsequently, 384 antibodies against 344 targets, including 48 proteins that had been selected from the initial screening, were used for a targeted discovery in plasma from a total of 172 individuals diagnosed with MS, CIS or OND. To confirm the preliminary findings, 46 proteins targeted by 101 antibodies were evaluated in a collection of 443 plasma and 573 CSF samples. These analyses resulted in a list of candidate targets, which were evaluated by immunofluorescence analysis of post-mortem brain tissue sections from MS patients. Profiles for one these candidates, IRF8, was further investigated in an independent set of 50 plasma samples and the utilized antibody targeting IRF8 was characterized by Western blot analysis and epitope mapping. Figure taken from **Article IV**.

carrier family 30 member 7 (SLC30A7), interleukin 7 (IL7), growth associated protein 43 (GAP43) and methyltransferase-like protein 14 (METTL14), where in fact the latter onoriginated from the hypothesis-free set of 3,450 targets. In plasma, all these five targets revealed differential profiles between the SPMS and CIS patients (Figure 5.10-A). For IRF8 and IL7, this was mainly due to higher levels in SPSM patients, whereas for the other three targets, it was due to lower levels in SPMS patients compared to the CIS patients (Figure 5.10-B). Upon analysis of a small and independent plasma sample collection, the differences revealed for IRF8 between SPMS and CIS patients, as well as between SPMS and RRMS patients could be verified (Figure 5.10-C). Subsequently, the specificity of the utilized antibody against IRF8 was confirmed by Western blot analysis and by epitope mapping on high-density peptide arrays.

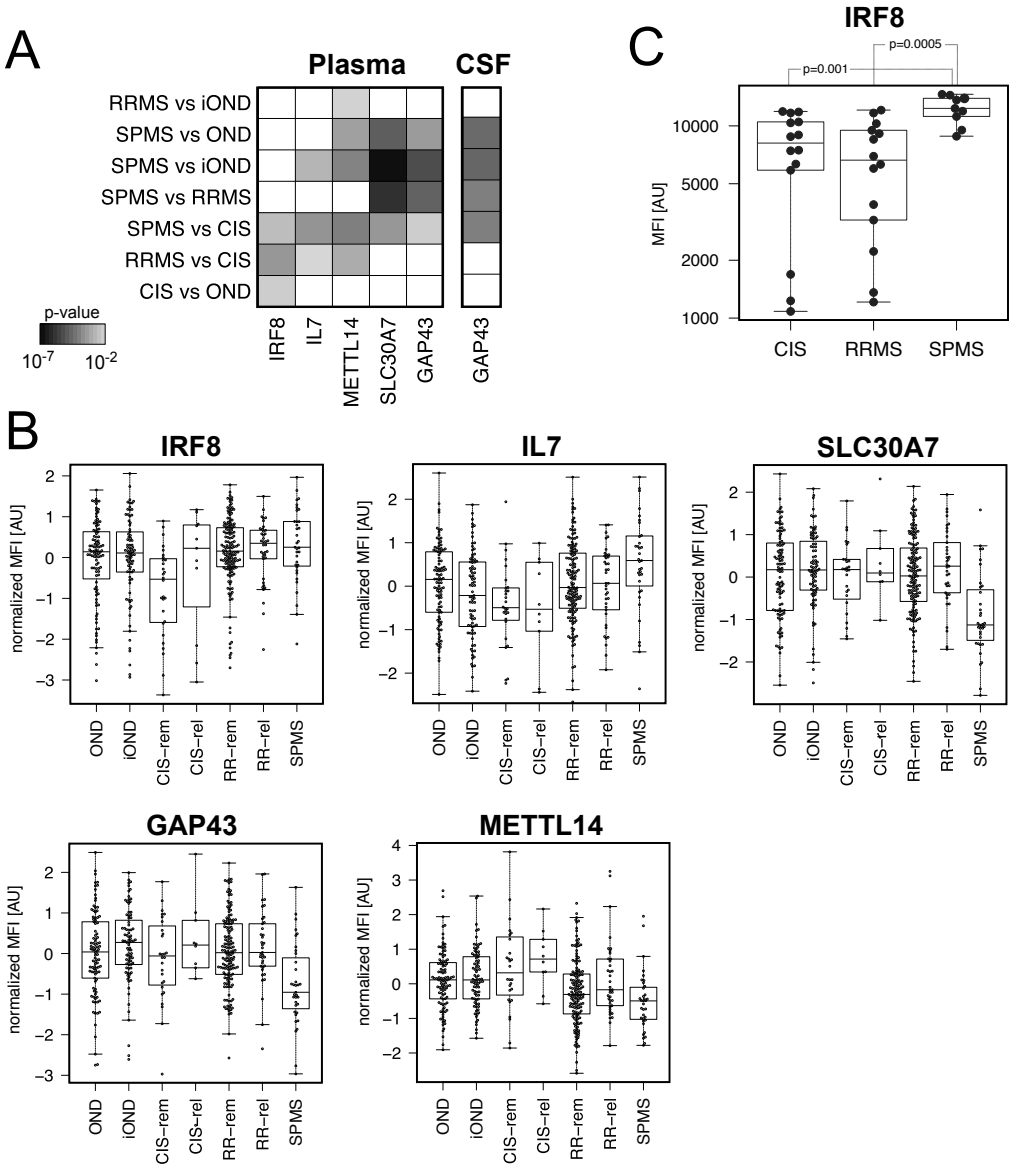

Figure 5.10: Overview of protein profiles for five targets, which revealed differences within an MS-related plasma sample collection in Article IV.
(A) Summary of the univariate analyses performed across CIS, subtypes of MS and controls with ONDs, which revealed significant differences for IRF8, IL7, METTL14, SLC30A7 and GAP43 in plasma. Out of these, only GAP43 revealed significant differences in the analysis of CSF samples. (B) Boxplots representing the normalized MFI values for the five highlighted targets in plasma, where the main differences were observed between the SPMS and CIS patients. (C) Boxplot representing the MFI values for IRF8 in an independent and relatively small plasma sample collection from patients with CIS, RRMS and SPMS, where the elevated levels of IRF8 in SPMS patients compared to RRMS and CIS patients could be verified. Figure adapted from **Article IV**.

In CSF, significant differences were revealed only for GAP43 out of these five highlighted targets (Figure 5.10-A). Interestingly, protein profiles in matched plasma and CSF samples did in general not reveal correlating trends. Besides, when investigating the correlation of the protein profiles between the five identified targets, plasma and CSF revealed different correlation relationships among these five targets. Finally, the five antibodies which revealed differences in plasma were used for immunostaining analyses on human post-mortem brain tissue with MS pathology. The main findings of the immunostaining analyses included staining of neurons (for IRF8, METTL14, IL7 and GAP43) or glial cells (for IL7 and METTL14) in the proximity of MS lesions by the respective antibodies.

Conclusions and Outlook

The investigations within **Article IV** adopted an affinity-based protein profiling approach within multiple sclerosis, where three different types of sample material, namely plasma, CSF and brain tissue were analyzed. This strategy allowed to identify potentially MS-related protein profiles in plasma, followed by assessment of the relation of these profiles with those from CSF and tissue. The initial stage of this investigation employed a hypothesis-free strategy, where over 4,500 antibodies were utilized to generate protein profiles in plasma. Out of these, the protein METTL14 remained in the refined list of five potential biomarker candidates. Considering that currently little is known about this protein and its function, let alone about its potential relation to MS, the supportive immunostaining patterns for this target highlight the potential of large-scale hypothesis-free investigations to suggest novel targets to be followed up in new investigations. This also underlines that an assessment of immunostaining patterns in the related tissue might offer support for the disease relevance of identified targets in body fluids, as demonstrated in this investigation. The observed lack of correlation between protein profiles in plasma and CSF, on the hand, highlights the need to ideally study both blood-derived samples and proximal fluid samples in discovery studies.

As illustrated in Figure 5.10-B for the five highlighted targets, the spread of MFI values across individuals with different MS subtypes implies the heterogeneity of protein profiles within MS, thus the need to analyze large collections of body fluid samples for identification of potential biomarker candidates in MS. In this regard, the bead-based antibody array format is a suitable tool in terms of its sample throughout capacity. While this investigation highlighted IRF8 as the main target of interest, all the five targets merit an analysis of a larger collection of plasma samples on bead-based antibody arrays. Ongoing investigations include such an analysis, as well as an investigation of the protein profiles for these five targets among other neuroinflammatory diseases in order to assess the specificity of these profiles for MS.

5.4 Protein Profiling in Muscular Dystrophies using Antibody Arrays – Article V

Aims of the Investigation

Duchenne muscular dystrophy (DMD), named after Guillaume-Benjamin-Amand Duchenne (1806–1875), is a rare, X-chromosome linked muscular disorder, with an incidence rate between 1:3,600 and 1:6,000 (Bushby et al., 2010). It occurs due to mutations in the dystrophin gene, leading to the absence of or defect in the dystrophin protein (Hoffman et al., 1987), which in turn causes progressive muscle degeneration. Boys affected by DMD present in the first years of life with muscle weakness and delayed motor functions. Progressive muscle weakness leads to loss of ambulation and complete wheelchair dependence before their teens and if untreated, further cardiac and respiratory complications emerge, limiting the life span down to early twenties (Bushby et al., 2010). Becker muscular dystrophy (BMD), named after Peter Emil Becker (1908–2000), is a milder form of DMD, which is caused by alteration in the size or a reduction in the amount of the dystrophin protein. BMD patients usually remain ambulant until or even after age 16. Diagnosis of DMD and BMD relies on genetic testing for mutation detection in the dystrophin gene and muscle biopsy analysis for the expression of dystrophin protein, as well as elevated levels of creatine kinase (CK) protein in serum (Manzur and Muntoni, 2009). Elevated CK levels are however not specific for muscular dystrophies (Brancaccio et al., 2010) and the invasiveness of muscle biopsy renders the currently available tools not ideal for assessments of disease severity and monitoring of disease progression in young patients.

As I briefly reviewed in section 4.2.2, antibody arrays are being applied in a broad spectrum of diseases, aiming to identify potential biomarker candidates in body fluids to support disease diagnosis and monitoring. Despite the prominent need for such protein biomarkers in muscular dystrophies, such attempts, as in other rare diseases, have been entirely lacking, mainly due to the scarcity of sample collections allowing for discovery-driven protein profiling approaches. In **Article V**, access to a unique assembly of plasma and serum collections from four different clinical sites in three different countries allowed for implementation of antibody arrays for protein profiling within muscular dystrophies. Adopting a hypothesis-driven approach, a list of 315 target proteins with a potential relevance to muscular dystrophies was assembled. A total of 384 antibodies generated against these proteins within the Human Protein Atlas were utilized to create bead-based antibody arrays. Using a single-binder assay format, a total of 345 plasma and serum samples of DMD, BMD patients, non-diseased controls and female carriers, all collected within the EU project BIO-NMD, were analyzed for their protein profiles (Figure 5.11).

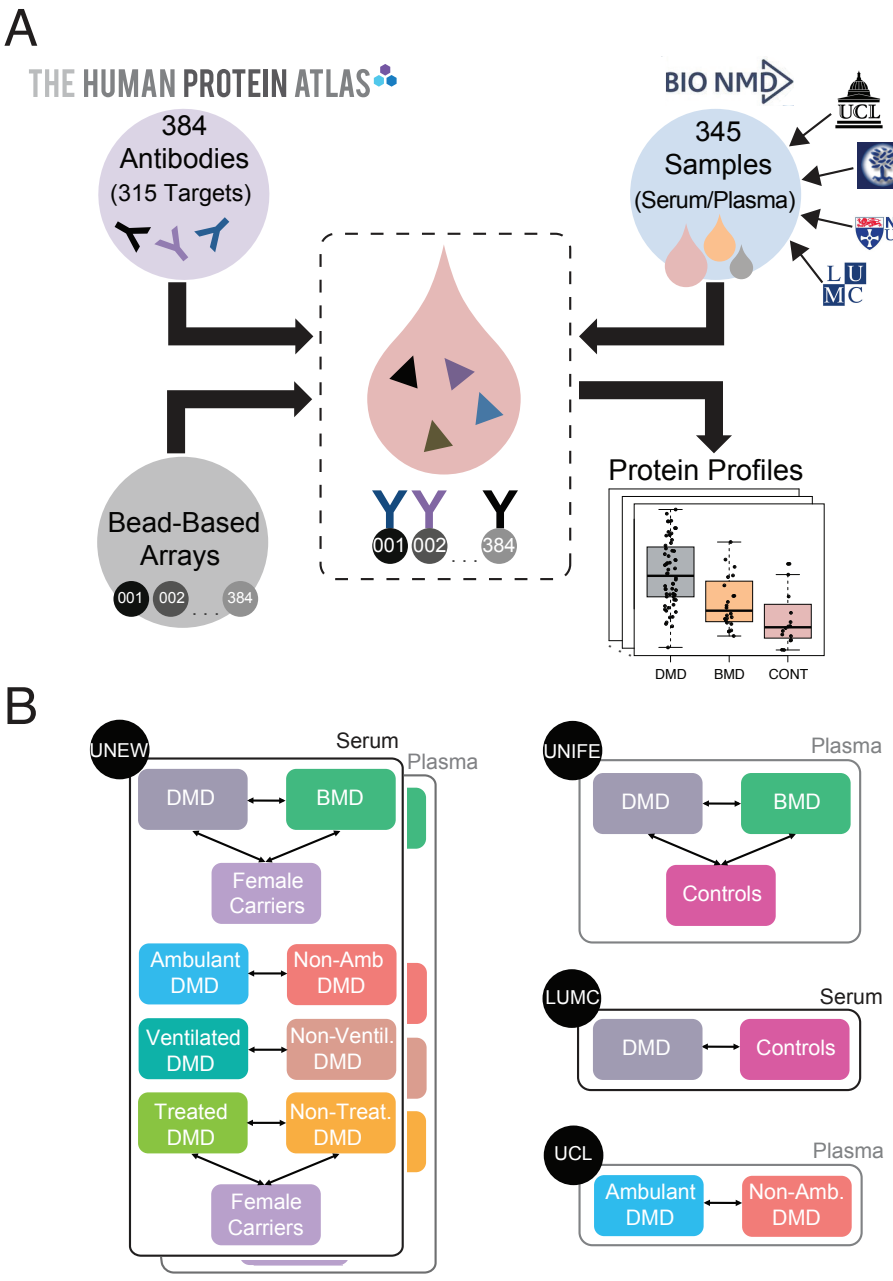

Figure 5.11: Overview of the protein profiling approach within Article V in the context of muscular dystrophies.
(A) Using a multiplex single-binder assay format, bead-based arrays were utilized to analyze 345 serum and plasma samples collected at four different clinical sites in three different countries: UCL (London-UK), UNEW (Newcastle-UK), LUMC (Leiden-Netherlands) and UNIFE (Ferrara-Italy). The array consisted of 384 antibodies, targeting a pre-selected set of 315 proteins. (B) The obtained protein profiles were dissected to identify targets with differential profiles across diagnosis groups, as well as those revealing an association with disease development and clinical parameters, such as ambulation or ventilation status. Figure taken from **Article V** by Ayoglu et al. (2014a).

Summary of Findings

The sample collection analyzed within **Article V** included both plasma and serum samples. As shown in Figure 5.12-A, an exploratory data analysis revealed global differences in serum and plasma protein profiles, which supported previous findings by several other reports as I discussed in section 3.2.1. Furthermore, especially within the plasma collection, a subset of samples derived from one of the clinical sites varied from the other two plasma sample collections (Figure 5.12-B). These observations, in turn, dictated to dissect the data into each sample preparation type and clinical site, followed by a statistical analysis within each sample subset.

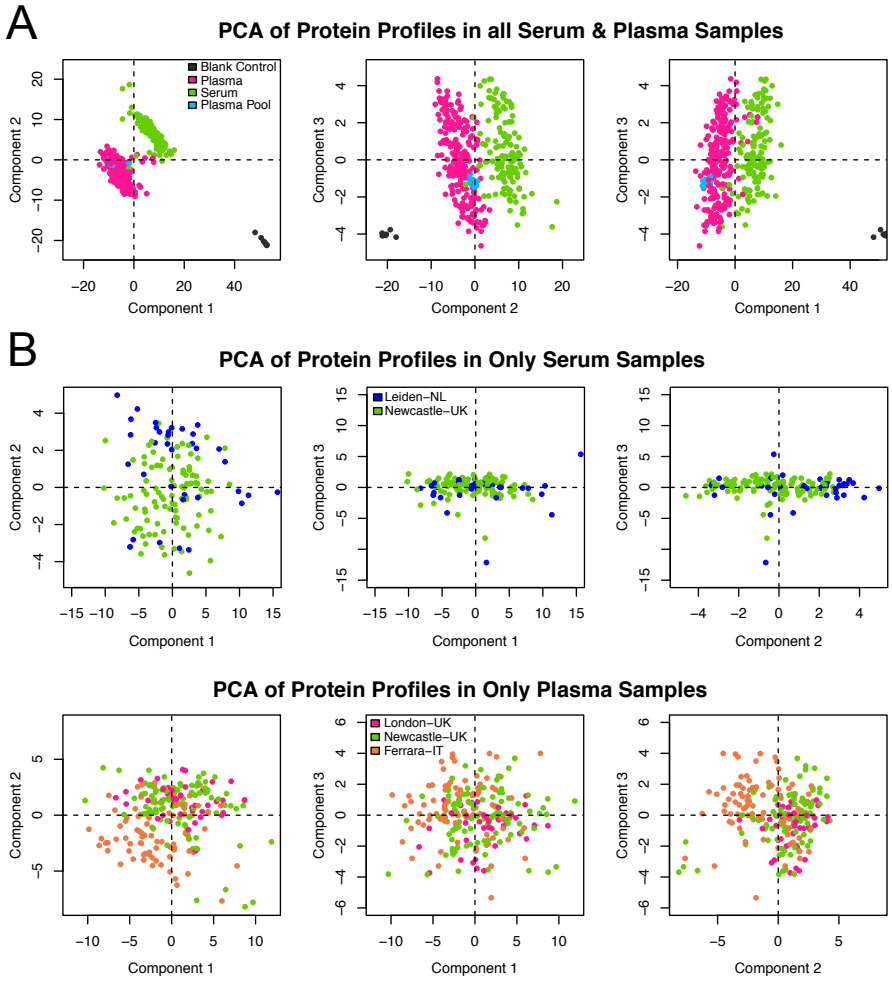

Figure 5.12: Principal component analysis (PCA) revealed sample preparation type and sample origin effects in Article V.
(A) PCA showed that the protein profiles across all plasma and serum samples grouped mainly by the blood preparation type (B) Furthermore, a slight effect of sample origin was observed where the plasma samples originating from Ferrara, Italy varied from the ones collected in London or Newcastle, UK. Figure adapted from supplementary material of **Article V** by Ayoglu et al. (2014a).

A

	DMD versus control groups				BMD versus control groups			DMD versus BMD			AMB versus non-AMB DMDs			VENT versus non-VENT DMDs	
	Plasma		Serum		Plasma		Serum	Plasma		Serum	Plasma		Serum	Plasma	Serum
Antibody name	UNEW	UNIFE	UNEW	LUMC	UNEW	UNIFE	UNEW	UNEW	UNIFE	UNEW	UNEW	UCL	UNEW	UNEW	UNEW
MDH2															
MYL3															
ETFA															
CA3-Ab#1															
CA3-Ab#2															
TNNT3															
ETFB															
CK															
LCP1															
ENO3															
PPM1F															
COL6A1															

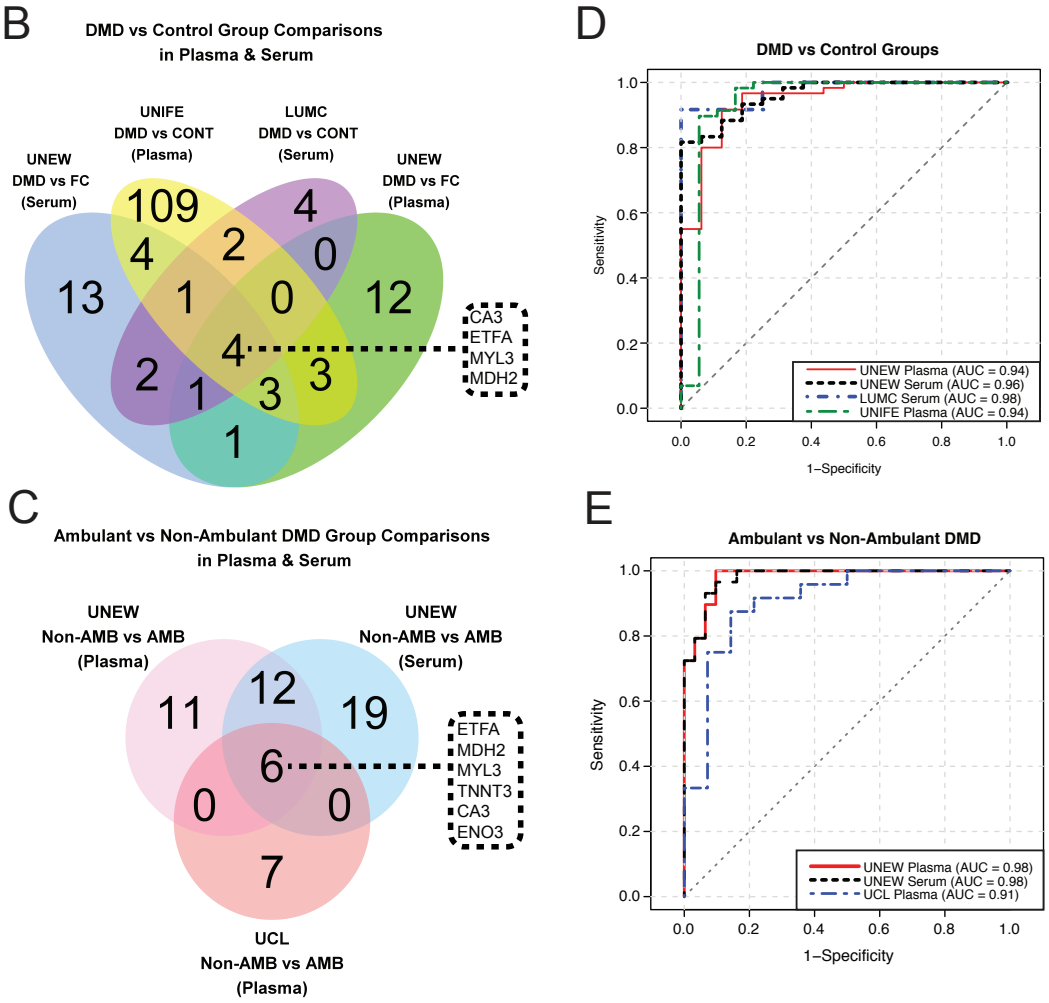

Figure 5.13: Overview of protein profiles identified within Article V, which were associated with disease phenotype and severity in muscular dystrophies.
(A) List of antibodies for which the protein profiles revealed differences with p-value <0.001 (dark blue) or p-value <0.01 (light blue) in various group comparisons. (B-C) Venn diagrams showing the number of proteins revealing significant differences (p-value <0.01) in samples of different preparation type and origin for the comparisons between DMD *vs.* control and ambulant *vs.* non-ambulant DMD, respectively. (D-E) ROC curves demonstrating the classification power of the antibody panels shown in Venn diagrams for the classification between DMD *vs.* control and ambulant *vs.* non-ambulant DMD, respectively. Figure adapted from **Article V** by Ayoglu et al. (2014a).

The outcome of this investigation was a set of eleven potential biomarker candidates (Figure 5.13-A). In the comparisons shown in Figure 5.11-B, profiles for these eleven targets revealed concordant differences for samples originating from more than one clinical site and for both blood preparation types. The identified targets included carbonic anhydrase III (CA3) and myosin light chain 3 (MYL3), both known to be specifically expressed in slow-twitch muscle fibers, and two mitochondrial proteins, malate dehydrogenase 2 (MDH2) and electron transfer flavoprotein A (ETFA). As shown in Figure 5.13-B and D, a panel of these 4 protein profiles could differentiate DMD cases and controls. Similarly, with contribution of enolase 3 (ENO3) and troponin T type 3 (TNNT3), ambulant and non-ambulant DMD patients could be differentiated (Figure 5.13-C and E). In summary, the protein profiling investigations within **Article V** generated a refined list of eleven proteins differing across different phenotypes of muscular dystrophies, as well as between patients presenting different degrees of disease progression.

Conclusions and Outlook

The protein profiling approach demonstrated in **Article V** underlines the potential of multiplex and high-throughput antibody array platforms for the identification of potential plasma biomarker candidates in rare diseases. The eleven protein targets identified in this investigation were independent of sample origin and preparation type and now merit dedicated studies to further verify these protein profiles. This, in turn, emphasizes the need for establishment of multi-national muscular dystrophy biobanks. As experienced in this investigation, despite harmonized sample collection, handling and storage protocols adopted within the BIO-NMD consortium, effects of sample origin, as well as sample preparation type could be found and required a separate assessment of protein profiles within each sample preparation type of different origin. In this particular study, consistent protein profiles could still be identified due to moderate size of sample subsets. However, this might in general render a challenge, thus underlining the the need for adoption of internationally standardized sample collection efforts.

Ongoing investigations include protein profiling for this set of eleven targets in a longitudinal sample collection originating from two of the four clinical sites. This will reveal in more detail the relation between these identified targets and disease progression. Reverse-phase plasma/serum arrays are also being generated to investigate the highlighted targets. CA3, for instance, has an average concentration in normal serum around 10 ng/ml compared to around 250 ng/ml in DMD serum (Mokuno et al., 1985, Ohta et al., 1991), thus analytically allowing for adoption of a reverse-phase assay strategy to analyze hundreds of samples simultaneously. Besides this, efforts will be spent to develop a bead-based sandwich assay upon testing several more antibodies towards the highlighted targets with the aim to identify matching antibody pairs, in line with the strategy I described in more detail in section 4.4.4.

Part III

Conclusions and Perspectives

Conclusions and Perspectives

Knowing is not enough, we must apply.
Willing is not enough, we must do.

Johann Wolfgang von Goethe

While putting down the words for it, I became interested in the numbers behind this thesis work - the number of data points, the number of samples, antigens and antibodies. I made a calculation limiting it to **Article I-II-IV-V**, as well as the fourth article in the related articles list. These investigations represent biomarker discovery-oriented studies where affinity arrays were applied for a multiplex and high-throughput profiling of proteins or autoantibody repertoires. It turned out that over 2.7 million data points were generated within these five investigations. Around 18,500 antigens were used for autoantibody profiling and almost 5,400 antibodies were used for protein profiling. Altogether, affinity arrays were utilized for the analysis of almost 4,300 samples. Calculating these numbers explained the sense I sometimes experienced during these investigations - the sense of searching for a meaning in an ocean of data. However, I also remembered the driving force I frequently experienced during these investigations: The idea that hiding within this ocean might be the small piece of information, which could eventually make a change in the life of a patient, who might be a family member, a close friend or maybe myself.

Goethe's words quoted above summarize my perspective regarding the motive behind the investigations presented in this thesis work. "Willing" to identify protein biomarkers is not enough, so is having a solely critical and discouraging standpoint regarding the possibility of identifying them, as sometimes expressed by experts of the field. We should instead exploit the available tools at hand to generate the maximum information we can generate by means of data. At the end, the price of "doing" and "applying" is probably less than the cost of assuming that it is not possible to identify useful protein biomarkers.

The investigations presented in this thesis suggest that if affinity reagents, sample collections and multiplex and high-throughput array platforms are available and accessible, opportunities for biomarker discovery studies can be created and not only sought after. **Article I** and **Article II** can be considered as a good example to support this point. In

Article I, the availability of a large number of human protein fragments within Human Protein Atlas allowed to identify a set of potential targets of autoantibodies in multiple sclerosis. In **Article II**, these findings could be followed up owing to the availability of a new sample collection with over 2,000 samples and owing to the availability of a bead-based array platform allowing to analyze a sample collection of this size even by a single operator. Three of the targets highlighted as the outcome of **Article II** will now be investigated in over 17,000 samples utilizing bead-based arrays and the outcome of such a large-scale analysis will reveal the biomarker potential of these targets.

This example underlines several other points. First, affinity arrays are not only tools allowing to investigate limited sets of proteins selected by making educated guesses. Affinity arrays can indeed be applied as discovery tools to generate a hypothesis. Second, it highlights the importance of bead-based arrays: As revealed by the overview I have provided in section 4.2.2 and 4.3.2, affinity arrays are in fact applied in the context of several diseases. However, a majority of these discovery studies used planar arrays by analyzing mostly not more than 100 samples and the identified sets of targets were usually not investigated and verified further in extended sample collections. In few cases, this can certainly be due to limited availability of further clinic material, but in several cases it is presumably due to no access to bead-based array platforms, which offer a higher sample throughput than planar arrays for a multiplex analysis of the identified targets as a panel. In this regard, bead-based arrays should be considered as an essential part of the affinity array-based biomarker discovery and verification investigations in the near future. Although frequently expressed, an important detail to highlight is also the low sample consumption offered by affinity arrays. For instance, by consuming only 10 µl of a crude plasma or serum sample, the assays described in **Article IV** and **Article V** can be used to generate protein profiles with 100 different sets of bead-based antibody arrays, each containing 384 different antibodies; thus offering a possibility for protein profiling across 38,400 antibodies. In short, the marriage between comprehensive collections of affinity reagents, such as in the Human Protein Atlas, and multiplex and high-throughput array technologies allows to explore a significant fraction of the protein or autoantibody repertoire of body fluid samples for biomarker discovery applications.

As I emphasized within the thesis, there is a benefit to adopt a more meticulous nomenclature regarding biomarkers. Although there is not much confusion about the definition of a biomarker, the term currently does not have the same meaning when used by researchers involved in biomarker discovery efforts and by clinicians working bedside and needing these biomarkers. The outcome of very initial biomarker discovery efforts is not biomarkers, but potential biomarker candidates. From a clinical point of view, these are in several cases not more than lists with protein names – unless the findings are verified in new and large sample collections, where also clinicians might start to recognize any potential use. The latter point should also help to develop a perspective regarding the

consequences of overstated early claims within biomarker discovery studies. A surplus of prematurely declared discoveries in the biomarker discovery field might potentially lead to a decay in interest, investment and funding within this field. Thus, an increased awareness about this risk should guide for more critical examination of findings, tools, reagents and methods regarding biomarker discovery efforts. In line with this, there is a great benefit to structure proteomic biomarker discovery studies to integrate an initial discovery phase with a verification phase. This perspective might be the key to transform the "lists with some protein names" to "potential biomarker candidates". Certainly, such a refinement is also key to identify those targets having a true potential to become biomarkers, which would merit time- and cost-intensive further studies.

Immunoassays are not perfect and there are several challenges shared by immunoassay- and mass spectrometry-based methods due to the complex nature of the human proteome, especially the blood proteome. Yet, immunoassays are still considered as the clinical-grade assay types. Mass spectrometry-based approaches are certainly very powerful and in near future will become technically even more powerful upon further improvement of aspects such as analytical sensitivity. However, the perception is widely shared that mass spectrometry is not sufficient alone to meet the needs for a routine use in the clinical diagnostics. Discoveries made using mass spectrometry-based approaches therefore usually seek for a verification on immunoassay platforms. In several cases, translation of mass spectrometry-based findings to immunoassays renders a great challenge and results in a gap phase between discovery and verification. This might be considered as one of the reasons for why there are few protein biomarkers in use despite the great number of potential biomarker candidates reported as outcome of mass-spectrometry based profiling approaches. In this regard, affinity array-based discoveries offer an advantage as translation of affinity array-based discoveries is more straightforward to clinical-grade assays. It is therefore worth to underline that affinity arrays are efficient tools suited for both discovery and verification stages of the biomarker discovery pipeline.

Regarding the different technologies for biomarker discovery and verification, the most important perspective might be to focus on understanding and evaluating the strengths and limitations of each technology. This would allow to gain the full benefit from the existing technologies by combining them at relevant stages of the proteomic biomarker discovery efforts. A technology allowing to identify 10,000 proteins in 100 samples might be considered as a good discovery tool. A technology allowing to measure 100 proteins or autoantibody reactivities in 10,000 samples might be an equally good discovery, as well as a verification tool. Taking into consideration the inherent variability across human populations and the heterogeneity of diseases, the latter approach might in fact offer a significantly more efficient route for biomarker discovery and verification. Thus, the current and future technology development efforts within proteomic biomarker discovery should maybe not only focus on improving the sensitivity and selectivity of assays. Since biology

and its degree of complexity will remain irrespective of the existing or future technologies, an equal emphasis should be put on developing efficient assay strategies, workflows and automation procedures allowing for robust analysis of thousands of samples. However, one important aspect to address here is that investigations involving thousands of samples will require more sophisticated statistical analysis strategies, as well as more organized and standardized data documentation, handling and storage frameworks. In order to turn the large amount of data into insight, efforts will be needed to develop efficient strategies for analysis and maintenance of data within such large studies.

The tremendous dynamic range of protein levels in body fluids such as plasma renders protein profiling approaches an analytical challenge. Efforts dedicated to improve the analytical sensitivity of antibody array-based protein profiling assays are therefore very valuable. It might be beneficial for instance to develop new assay procedures by merging the different array platforms with the evolving signal amplification methods. Regarding the unmet needs in terms of technical and analytical developments, sensitivity is however not the only aspect to be addressed. Another important challenge to be recognized is related to antibody cross reactivity and off-target interactions affecting the selectivity of antibody-based protein profiling assays. This aspect by itself receives relatively less attention than development of new assay technologies with improved analytical sensitivity. However, the obstacles for developing better protein profiling assays reside as much in the compromised assay selectivity as in analytical sensitivity. Fortunately, the awareness regarding the quality of affinity reagents is increasing. Findings of protein profiling investigations depend on the assay context and the quality of the utilized antibodies or other types of affinity reagents. In this regard, efforts are crucial to use orthogonal methods to further characterize the utilized antibodies which have led to discoveries. Here, various strategies can be taken including more conventional methods such as a Western blot, as well as epitope mapping utilizing peptide arrays or mass spectrometry-based approaches such as a targeted mass spectrometry analysis. This point highlights the importance of combining the powerful aspects of different proteomics tools and the importance of the interplay between affinity array-focused and mass spectrometry-focused proteomics fields. As recognized within the Human Protein Atlas, there is also a potential benefit in generating more than one affinity reagent per protein target. Availability of such collections encompassing several affinity reagents per target can play a significant role for early development of sandwich assays directly upon initial discoveries.

Regarding the application of antigen arrays for autoantibody profiling purposes, it is foreseeable that the antigen arrays will continue to be exploited in the near future. The need remains to unravel the characteristics of autoantigens and the role of autoantibodies in both health and disease. Therefore, antigen arrays are currently very valuable "fishing expedition" tools, regardless of the type and source of antigens utilized and regardless of whether continuous or discontinuous epitopes are examined. Arrays

composed of full-length proteins, protein fragments, synthetic peptides, peptidomimetics, carbohydrates or lipids are currently all needed and should be utilized more to first increase the understanding about the autoantibody repertoire in body fluid samples. Generating more information will clarify the relevance and clinical value of each type of antigens, as well as the relevance of analysis of autoantibodies directed against continuous or discontinuous epitopes. Studies utilizing arrays with an unbiased content can provide a particularly good starting point to unravel the common features of autoantigens being recognized in body fluid samples both in the context of diseases, as well as in health. In near future, there is also a great benefit in focusing not only on what the autoantibodies are binding to but also what kind of autoantibodies are binding to the self-proteins. More systematic studies shifting the focus to differences in immunoglobulin class frequencies might provide valuable clues about whether certain set of antigens sharing certain features related to e.g. structure or localization co-determine the antibody class, thus the effector functions such as complement activation. The assay workflow described in **Article III** for a parallel analysis of autoantibody classes and antigen-specific complement activation provides a good example for such coming efforts. Regarding the needs in terms of technical developments for antigen arrays, analytical sensitivity can be considered an area to spend further efforts, as for antibody arrays. Utilizing detection methods beyond the conventional fluorescence-based strategies and experimenting with more novel label-free or label-based detection technologies on different antigen array platforms is an area open for development of new assay types on antigen arrays.

An overview of the studies mentioned in section 4.2.2 and 4.3.2 applying affinity arrays reveals that a vast majority of the biomarker discovery studies focus at a single disease and perform a comparative analysis utilizing samples from patients with the related disease and non-diseased controls. Such studies certainly contribute to the understanding of each individual disease on a molecular level. There is still an existing need for such studies in several conditions, which could not yet fully benefit from the possibilities offered by multiplex and high-throughput proteomics technologies. Rare diseases such as Duchenne muscular dystrophy, which was investigated in **Article V**, can be considered as an example. On the other hand, certain disease types such as cancer have been studied extensively, mostly involving a comparison between samples from advanced cancer patients and non-diseased controls and often highlighting e.g. inflammation-related proteins as biomarker candidates. For such study areas, it might be more beneficial to shift the focus from single cancer types and design studies involving several cancer types. Such multi-disease screening approaches might help to identify more disease-specific biomarker candidates. Certainly, the same approach can also be beneficial for other disease groups such as neurodegenerative disorders or systemic autoimmune diseases. Similarly, more studies involving longitudinal sample collections or matched body fluid sample collections might enrich the information generated for potential biomarker candidates. Such studies might help to assess whether the identified potential biomarker candidates are able to

assist disease monitoring and to co-evaluate the biomarker potential of identified proteins in different body fluids, respectively. As shown in **Article II** and **Article IV**, there is also a potential to be realized by extending the analysis of body fluid samples for an expression profiling of identified targets on the affected tissue. Such integrative and collaborative approaches, where immunohistochemistry investigations can be carried out in tandem with affinity-array based discoveries will certainly allow to gain more biological insight at an early stage of discoveries and thus help to assess the relation of identified targets to disease pathogenesis.

As demonstrated in **Article V**, when possible there might also be a great benefit in adopting a multi-center study design approach within biomarker discovery investigations. The use of multi-center sample collections early in the discovery phase significantly helps to identify those targets which are less sensitive to variations in sample collection, handling, storage and transfer. Such discoveries based on "portable" evidence have an increased chance to survive the gradually more challenging verification stages of the biomarker discovery pipeline, which usually involve multi-center sample collections.

Regarding the latter point, one aspect worth to highlight again is the importance of sample quality. Samples are the starting point of the proteomic biomarker discovery pipeline, no matter which tools are utilized. The awareness about the importance of sample quality, form and standardization is fortunately increasing, so does the awareness about the need for more dedicated biobanking efforts. Standardization of sample preparation, handling, storage and transfer procedures might currently not be there yet. But this does not mean end-users like ourselves have no responsibility or possibility to make a change. It is our collective responsibility not to rely on the quality of samples which can be conveniently obtained from local institutions, but to increase awareness and to propose relevant requirements to be adhered. Local institutions or even established biobanks should not be expected to take the significant strides only by themselves. When establishing new collaborations on local or international level, end-users could proactively share experienced or foreseeable challenges or impracticalities for downstream analyses and suggest improvements regarding sample or metadata transfer formats. If biobanks, clinicians and researchers performing proteomic biomarker discovery studies will work more in tandem, it will be possible to gain full benefit from the analyses of good quality and well-characterized sample collections.

In fact, such collaborative and synchronized efforts are required not only to guarantee sample quality but also during the planning, initiation and progress of biomarker discovery projects. Ideally, a multi-disciplinary expertise should be involved in all steps of such projects, including the researchers performing the analysis, statisticians but most importantly the experts of the disease. Any collaborative gaps upon transfer of samples from clinical sites to research laboratories for a proteomic analysis might cause projects to be designed and performed not under optimal clinical expertise, which might cause further

challenges in data analysis and interpretation. In order to deliver results in an efficient manner, projects should be initiated and carried out within multi-disciplinary teams involving the clinical colleagues at all stages of the biomarker discovery process. Efforts addressing the challenges within biomarker discovery field should therefore not only aim for more powerful technologies and should not only focus on technical limitations. Future efforts should also realize that there is room to improve the organization of biomarker discovery projects by involving continuous networking and cross-disciplinary expertise providing medical, technical and biological insight.

The day when this doctoral thesis will be defended, it will be the twentieth birthday of the proteome concept, which was introduced by Marc Wilkins in the first Siena Meeting held between 5–7 September 1994. Similarly, protein and antibody arrays are nowadays celebrating their almost fifteenth birthdays. Predicting the future of proteomics, efforts in generating affinity reagents, affinity arrays and affinity array-based biomarker discovery field is not easy. Yet, things which can not be envisioned can not be reached at all. As I have aimed to argue for in this thesis, affinity arrays are not at their infancy anymore and they can be envisioned as capable tools to address the need for the sought-after biomarkers. There is now enough evidence to appreciate the role of affinity arrays in the quest towards biomarkers and this appreciation is needed to provide a momentum for further technological improvement within the affinity array field. If there were two conclusions to be drawn from this thesis and the investigations performed, the first one would be that affinity arrays can and will contribute for the immense accumulation of valuable information regarding possible molecular mechanisms behind several diseases. The resulting complexity of this information will require an even closer interplay between researchers from various fields including molecular biology, statistics, computer science, engineering and last but not least, medicine. Similarly, a stronger cooperation between different sub-disciplines of proteomics and with other omics fields will be needed. The second conclusion would therefore be that if biomarkers are ever to be found and established, then this will be accomplished in such a multi-disciplinary climate. The road of proteomic biomarker discovery is long, iterative and frequently ends in frustration. It therefore might resemble the myth of Sisyphus, relentlessly rolling a boulder to the top of a mountain and then watching its descent. But by applying the tools at hand and adopting a productive and collaborative perspective, the struggle for biomarkers is worth and it serves for a very good purpose. As Albert Camus (1913–1960) concludes: "The struggle itself toward the heights is enough to fill a man's heart. One must imagine Sisyphus happy."

Part IV

Acknowledgements

Acknowledgements

Scientific research is one of a fun team sport, where progress becomes possible through a good team effort and where achievements become fun when shared with the fellows. I therefore would like to acknowledge the following individuals both from within and outside of research, including mentors, collaborators, colleagues, friends and family.

Peter Nilsson — I am extremely grateful to you for encouraging me to start a PhD. By doing so, you gave me the opportunity to find out what I wanted to do and to eventually discover that research is indeed something for me. Thank you for your continuous support while me generating "massive" data sets, diving in the ocean of data, trying to understand the peculiarities about autoimmunity and eventually developing confidence in what I do, what we do, and how we do things, while at the same time never holding back your constructive opinion and criticism regarding even smallest details. Thank you for your contagious energy and excitement. Thank you for the fun thesis discussions we lately had and for your support, no matter what day and what time. And more importantly, thank you for your trust in me during these years; this has been such a great driving force, such a big help to bring out the determination in me to the surface and to not give up. We have done and do great work together. Det har varit mycket spännande – tusen tack för allt.

Jochen Schwenk — Not everyone like me is lucky enough to find a mentor, a teacher, a life coach and a boss in a single person. Thank you for becoming my all-in-one supervisor – for when I needed your support, humor, advice and encouragement. I could always think out loud around you; I could always count on you to teach me something new, to schwenkinize, to add a different color to things and to provide a helpful mind. Despite the argumentative nature we both seem to share and to enjoy, we never reject each others opinions and suggestions just out of hand; you delicately plant goals and perspectives into my mind, where they always find their way to merge with my own to start growing. We see things sufficiently similar and sufficiently different, which adds a fun and exciting dimension to things we work on. I believe we can always come up with and accomplish cool things together. Du hattest immer eine offene Tür und ein offenes Ohr für meine Gedanken, Fragen und Sorgen. Mit sehr viel Geduld hast Du mir immer Kraft gegeben. Danke für das Verständnis, Danke für die Unterstützung, und Danke für das Vertrauen – Danke für alles.

"Peter & Jochen" — Thank you "Peter & Jochen", which is an imaginary third supervisor of mine, a supervisor where both of my supervisors were fused into a single one. This is actually what I did during these years: No matter how different or similar, I fused suggestions, advices, reactions and perspectives from both of them. There were times when one of them said "black" and the other one said "white", but this soon taught me the possibility and importance of finding the "grey" and all three of us agreeing on the "grey" instead. I learned a lot by working with "Peter & Jochen". Most importantly, I fully enjoyed working with "Peter & Jochen". I am grateful to have had both of them by my side during this exciting voyage.

Mathias Uhlén — Thank you very much for gathering all of us under this very same roof of Human Protein Atlas. Thank you for the great time we had with you and Peter in Istanbul during the ECB15 congress.

The following funding sources and organizations made the work presented here financially possible. I thereby would like to acknowledge **Knut and Allice Wallenberg Foundation**; **ProNova VINN Excellence Centre for Protein Technology** (with special thanks to **Amelie Eriksson Karlström** and **Per-Åke Nygren**); the grants and opportunities provided by the **BIO-NMD** project, as well as the grants provided by the **EuroTEAM** project and by **SciLifeLab**.

All past and present members of the big **Human Protein Atlas** family — Thank you for your great efforts to generate all the precious antigens and antibodies; this work would otherwise not be possible. I also would like to acknowledge **Atlas Antibodies AB** for providing the financial support for all the amazing Human Protein Atlas gatherings.

I would like to thank **"Peter & Jochen"** for the proofreading of this thesis and their constructive feedback. I would like to acknowledge **Fulya Taylan** for the very careful final proofreading. I am grateful to **Peter Nilsson**, **Eva Poijes** and **Ronald Sjöberg** for their help to translate the popular scientific summary to Swedish. I also would like to thank **Nicholas Mitsios** and **Tony Jiménez-Beristain Andersson** for their feedback regarding parts of this thesis. **Jacob Odeberg** is acknowledged for the thesis quality review. I am deeply grateful to **Alper Aydemir** for the LaTeX support he has provided despite the hours of time difference and thousands of kilometers between us. For the lovely artwork on my thesis cover, I owe my gratitude to **Brendan Monroe**.

There are several collaborators, whom I would like to acknowledge for their involvement and contributions in our finished or ongoing projects.

Krisztián Papp (Eötvös Loránd University)— Thank you very much for being such a good role model for me during the C3 project. I learned a lot from you, I got inspired a lot from your dedication. You showed me the possibility of initiating and finishing projects with a clear mindset in short time frames. Thank you for being a great collaborator and a good friend, always keeping in touch and sharing your new experiences with me.
Cristina Al-Khalili Szigyarto (KTH School of Biotechnology) — Thank you for the BIO-NMD journey. I learned a lot and I am very happy about the outcome. We did a a great job together.
Tomas Olsson, Mohsen Khademi (Karolinska Institute) — Thank you for the amazing MS collaboration we have been keeping since I started my master thesis project. Thank you for always feeding our projects with new samples. Thank you for recently involving **André Ortlieb Guerreiro Cacais** and **Ingrid Skelton Kockum** to take the ANO2 findings one step further.
Jan Mulder, Nicholas Mitsios (Karolinska Institute) — Thank you for your great collaboration in our MS-related projects.
Mats Nilsson, Anja Mezger (Stockholm University) — Thank you for the recent and exciting collaboration about RCA. Calculating LODs has never been this much fun.
Matthias Löhr (Karolinska Institute) — Thank you for initiating the pancreatic cancer project with us and thereby giving us a chance to identify very interesting targets to be followed up.
Petter Brodin (Karolinska Institute) — Thank you for initiating our recent collaboration regarding profiling of autoimmune responses against your highly interesting set of antigens.
Manuela Beil-Peter (CyBio) — Thank you for all your support with our lovely SELMA instrument and for all your friendly visits during all these years.

I would like to acknowledge the **SciLifeLab Biobank Profiling-Affinity Proteomics** Group, or the so-called **PAPP**, which grew so much, in several aspects, since the days I started back in 2009 as a master thesis student. I feel lucky to have had a chance to closely witness how this group has been evolving. We lately became a great ensemble, complementing each other in different ways. I hope this positive and productive culture of PAPP will be nurtured further in the future and yield good research, good papers and good friendships. I would like to highlight the companionship of the following colleagues from the **SciLifeLab Biobank Profiling-Affinity Proteomics**, as well as from the **Tissue Profiling** and **Cell Profiling** groups:

Nicholas Mitsios — Thank you for being a caring and fun friend, a supportive workmate and a good collaborator. I wish all the "ethnic enemies" would be like the two of us. The vodka evenings were the rewards for my evening autoimmunity assays; the Sunday visits colored my weekends accompanied by your cute family, and of course by börek and souvlaki. Hanging out with you relieved my homesickness. Thank you for being around whenever needed.

Frauke Henjes — Thank you for all the long chats; for the great company in Nice; for being less German than me according to the "German Quiz"; for the vodka evenings and Sunday afternoons; for all the football games we watched; for encouraging me to use LaTeX and for the support while writing this thesis. Alles Gute in deinem neuen Job; mach′s gut bis wir uns wieder mal sehen.

Elin Birgersson — Thank you for always being a reliable, fun and supportive workmate. Thank you for offering me help and encouragement whenever you felt like it could be needed, especially during the thesis writing journey. Thank you for booking the tables for the fun dinners, which revealed the hidden association in my subconscious mind between football and hamburger. Thank you for tagging along in the DCA journey and allowing me to take a short break while trusting your hands in the lab.

Sanna Byström — Thank you for first being a very good student to supervise and then for becoming a very good colleague and a collaborator to work with in the MS journey. I discovered several similarities between us during all these years and enjoyed more and more our several chats and discussions. Keep up the good work.

Eni Andersson — Thank you for your kindness and for never hesitating to offer me support with things. You proved me over and over that I am not alone in caring about other colleagues around.

Ronald Sjöberg — Thank you for being my faithful office neighbor for so long. I am glad we could comfortably swear under each other's company at the stupid R, stupid Excel, stupid Genepix and student lab reports to be corrected, which were full of stupid standard curves and stupid ten digits after comma. Thank you for always providing your help and opinion whenever needed.

Maja Neiman — Thank you my "big mama" for your support at the final kilometers of this marathon. Here it comes: "No!"

Cecilia Mattsson — Thank you for being my great PrEST registry with Eni. Thank you for all the Mollan lunches and for your support regarding the dissertation party arrangements.

Tony Jiménez-Beristain Andersson — Thank you for being my "xxx snake"; for your compliments about the clothes I wear and the cakes I bake; for approving that I reached a decent "huslig" level. Your Christmas cards made my days every year, your advices and comments about work- and life-related issues opened my eyes every time. I am so glad you are around.

Kimi Drobin — Thank you my "big mama nr. 2" for your support together with Maja during my late-breaking dissertation party crisis. And for all the fun times.

Maria-Jesus Iglesias — Thank you for taking care of the poor FM3D after my several late-night autoimmunity assays. Thank you for always providing a hug when needed at difficult times.

Arash Zandian — Thank you for checking whether all is going fine whenever I would not show up in the lab. Thank you for the nice company in Nice. Continue to be a strict SELMA deputy officer, stay tuned while getting done with papers and keep up the good work until you become the Dr. Zandian for real.
Ulrika Qundos, Anna Häggmark — Thank you for sharing the PhD years.
Mun-Gwan Hong — Thank you for all our soft and fun battles about the way of making ROC curves, calculating p-values and the inevitable differences in age range of different disease subtypes where you represented the statistician faithful to the good manners of statistics and where I represented the realistic life-science researcher who had to make sense of her data in a pragmatic and quick way. We are lucky to have you in PAPP.
Hammou Ait-Blal — Thank you for helping me to finally find a good word to describe myself: The "busy bee". Thank you for all the fun chit-chats.
Björn Forström — Thank you for the fun company in Nice.
Maria Mikus, Claudia Fredolini, Masato Habuka, Fredrik Edfors, Elisa Pin — Thank you for contributing to the nice atmosphere of our group.

There are also some past members of the **SciLifeLab Biobank Profiling-Affinity Proteomics** Group, whom I would like to acknowledge:

Spyros Darmanis — For so many things I learned and for the inspiration I got from him; for helping me to further develop my critical (and maybe sometimes prematurely cynical) attitude on results, projects and papers; for making me acquire the habit of continuously keeping an eye on the hot stuff going on in the field; for all the fun days we had in the lab while trying to make RCA work on reverse phase arrays and for making me realize that assay development projects initiated with great passion and carried out with great effort might still end up in the cemetery of dead projects.
Rebecca Rimini — For her great guidance during my master thesis project; for introducing me to the complex world of multiple sclerosis where I stayed since then.
Anna Gundberg, Mårten Sundberg — For their help and support with PrESTs back in the old days, when there was so much to do with so few hands around.
Ida Hossar, Philippa Reuterswärd — For being my students during their degree and bachelor projects. Supervising them and Sanna taught me a lot.
Josefine Jaxby — For being my Kingfisher and DCA buddy. We laughed, talked and got puzzled about several things; we said "Hey Ho-Let′s Go!" and moved on.

I would like to acknowledge other members of the **Cell Profiling** and **Cell Screening** groups, including **Emma Lundberg**, **Martin Hjelmare**, **Charlotte Stadler**, **Mikaela Wiking**, **Diana Mahdessian**, **Rutger Schutten**, **Frida Danielsson**, **Marie Skogs**, **Bo Lundgren**, **Ulf Martens** and **Maria Häggblad**, for contributing to the lovely atmosphere of our floor.

I want to thank all the responsible organizers of the great courses I have attended during my master studies at **KTH School of Biotechnology**, including the courses given by **Peter Nilsson**, **Amelie Eriksson Karlström**, **Cristina Al-Khalili Szigyarto**, **Afshin Ahmadian**, **Jenny Ottosson Takanen**, **Torbjörb Gräslund**, **Lars Arvestad** and **Joakim Lundeberg**. I also would like to acknowledge the organizers of some of the courses I have taken during my PhD studies, including **Cristina Al-Khalili Szigyarto** (KTH Biotechnology); **Kie Kasuga**, **Svetlana Lagercrantz**, and **Sanela Kjellqvist** (Karolinska Institute).

I would like express my appreciation to the instructors of the amazing **Cold Spring Harbor Laboratory (CSHL) Proteomics Summer School**, including **Ileana Cristea** and **Todd Greco** (Princeton University); **Darryl Pappin** (CSHL); **Michael Bereman** (North Carolina State University) and **Michelle Cilia** (Cornell University), whose limitless energy and teaching effort was very impressive and inspirational. They proved that partying hard until late night is so much fun after spending the evening to generate loads of data. I also would like to thank my CSHL friend **Shaun Garnett** (University of Cape Town) for challenging my perception regarding several research- and life-related subjects just on time. As Kazantzakis said, "since we cannot change reality, let us change the eyes which see reality".

It would be remiss of me not to mention some of the amazing teachers I had in the past. I would like to acknowledge my "first" teacher **Altın Alpaslan**, who always kept in touch and supported me during my entire education journey; **Betül Kırdar**, my bachelor thesis advisor at Bogazici University; and my inspiring and very influential literature teacher **Lars Bäcker** from Deutsche Schule Istanbul.

Last but not least, I would like to thank some of my close friends and family members.

Fulya Taylan — You were there at toughest times. You were caring about whether I take good care of myself during this thesis journey. I am so lucky to have met such a kind friend like you. Herşey için teşekkürler canım arkadaşım.
Oğuzhan Osman Erim — Thank you for being such a supportive friend. Together with **Hakan Akıllıoğlu**, you have been there whenever needed. You have been the source of fun discussions and the company for lunches and dinners at best locations of Stockholm. This city is more fun when knowing you are around.
Mao Yumeng — Thank you for all the fun we had during and after our master studies and thank you for the wonderful Venice trip.
Seda Yıldırım-Babakıray, Neslihan Ballık, Müge Çapan, Ömur Sözer, Doruk Canbilen, Neslihan Arol — Who cares how spread we are on the map of the world when knowing that we will always have things to talk until the daylight would break and that we will always have lots of fun together. Bizi güzel günler bekliyor.

Beyhan Yurdakul, Ismail & Buket Ayoğlu, Fitnat & Süleyman Ayoğlu, Gürcan Yurdakul, Burak Yurdakul, Selahattin & Ayşe Ayoğlu, Halil Ayoğlu, Fitnat Ayoğlu Kunter, Nesrin & Habip Canbilen — Thank you for your incredible love. Sevginiz ve bana verdiğiniz tüm emekler için teşekkür ederim.

Alper Aydemir — The co-pilot of our space shuttle voyaging around the pale blue dot, carrying our dreams. Thank you for letting in rainbows with rare colors through our window; for your amazing sense of humor; for your contagious passion and optimism. Thank you for making sure you are with me while being thousands of kilometers away. Thank you for being who you are and for teaching me that "le cœur n′a pas de rides". Cap ou pas cap, mon bel ami? Ne kadar şanslıyım ki sen hayatımdasın; hayat, seninle daha güzel.

Reyhan Yurdakul Ayoğlu & Şevki Ayoğlu — Thank you for the principles and values you have passed on to me, which are guiding and enlightening my life. Thank you for your undying support, your steady belief that I can achieve anything I would work for and your full respect and constant encouragement about my decisions and choices. Thank you for everything. Size aldığım her nefes, yazdıgım her satır, attıgım her adım için minnettarım; hayatım boyunca gurur duyacagınız bir evlat olmak için çabalamaya devam edeceğim; sizi çok seviyorum.

Part V

References

References

Adler, M., Wacker, R., and Niemeyer, C. M. (2008). Sensitivity by combination: immuno-PCR and related technologies. *The Analyst*, 133(6):702.

Agaton, C., Galli, J., Höidén Guthenberg, I., Janzon, L., Hansson, M., Asplund, A., Brundell, E., Lindberg, S., Ruthberg, I., Wester, K., Wurtz, D., Höög, C., Lundeberg, J., Ståhl, S., Ponten, F., and Uhlén, M. (2003). Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Molecular & Cellular Proteomics: MCP*, 2(6):405–414.

Aguilar-Mahecha, A., Cantin, C., O'Connor-McCourt, M., Nantel, A., and Basik, M. (2009). Development of reverse phase protein microarrays for the validation of clusterin, a mid-abundant blood biomarker. *Proteome Science*, 7:15.

Ahn, S. and Simpson, R. (2007). Body fluid proteomics: Prospects for biomarker discovery. *Proteomics - Clinical Applications*, 1(9):1004–1015.

Älgenäs, C., Agaton, C., Fagerberg, L., Asplund, A., Björling, L., Björling, E., Kampf, C., Lundberg, E., Nilsson, P., Persson, A., Wester, K., Ponten, F., Wernérus, H., Uhlén, M., Ottosson Takanen, J., and Hober, S. (2014). Antibody performance in western blot applications is context-dependent. *Biotechnology Journal*, 9(3):435–445.

Alhamdani, M. S. S., Schröder, C., and Hoheisel, J. D. (2010). Analysis conditions for proteomic profiling of mammalian tissue and cell extracts with antibody microarrays. *Proteomics*, 10(17):3203–3207.

Alper, C. A., Johnson, A. M., Birtch, A. G., and Moore, F. D. (1969). Human C'3: Evidence for the Liver as the Primary Site of Synthesis. *Science*, 163(3864):286–288.

Altman, D. G. and Bland, J. M. (1999). Statistics notes: How to randomise. *BMJ*, 319(7211):703–704.

Amado, F. M. L., Ferreira, R. P., and Vitorino, R. (2013). One decade of salivary proteomics: current approaches and outstanding challenges. *Clinical Biochemistry*, 46(6):506–517.

Anderson, K. S., Ramachandran, N., Wong, J., Raphael, J. V., Hainsworth, E., Demirkan, G., Cramer, D., Aronzon, D., Hodi, F. S., Harris, L., Logvinenko, T., and LaBaer, J. (2008). Application of protein microarrays for multiplexed detection of antibodies to tumor antigens in breast cancer. *Journal of Proteome Research*, 7(4):1490–1499.

Anderson, K. S., Sibani, S., Wallstrom, G., Qiu, J., Mendoza, E. A., Raphael, J., Hainsworth, E., Montor, W. R., Wong, J., Park, J. G., Lokko, N., Logvinenko, T., Ramachandran, N., Godwin, A. K., Marks, J., Engstrom, P., and LaBaer, J. (2011). Protein microarray signature of autoantibody biomarkers for the early detection of breast cancer. *Journal of Proteome Research*, 10(1):85–96.

Anderson, L. (2014). Six decades searching for meaning in the proteome. *Journal of Proteomics*, 107C:24–30.

Anderson, L. and Anderson, N. G. (1977). High resolution two-dimensional electrophoresis of human plasma proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12):5421–5425.

Anderson, N. L. (2010). The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clinical Chemistry*, 56(2):177–185.

Anderson, N. L. and Anderson, N. G. (1991). A two-dimensional gel database of human plasma proteins. *Electrophoresis*, 12(11):883–884.

Anderson, N. L. and Anderson, N. G. (1998). Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis*, 19(11):1853–1861.

Anderson, N. L. and Anderson, N. G. (2002). The human plasma proteome: history, character, and diagnostic prospects. *Molecular & Cellular Proteomics: MCP*, 1(11):845–867.

Anderson, N. L., Anderson, N. G., Haines, L. R., Hardie, D. B., Olafson, R. W., and Pearson, T. W. (2004a). Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *Journal of Proteome Research*, 3(2):235–244.

Anderson, N. L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R. S., Conrads, T. P., Veenstra, T. D., Adkins, J. N., Pounds, J. G., Fagan, R., and Lobley, A. (2004b). The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Molecular & Cellular Proteomics: MCP*, 3(4):311–326.

Andersson, A. C., Stromberg, S., Backvall, H., Kampf, C., Uhlen, M., Wester, K., and Ponten, F. (2006). Analysis of Protein Expression in Cell Microarrays: A Tool for Antibody-based Proteomics. *The Journal of Histochemistry and Cytochemistry*, 54(12):1413–1423.

Andersson, O., Kozlowski, M., Garachtchenko, T., Nikoloff, C., Lew, N., Litman, D. J., and Chaga, G. (2005). Determination of relative protein abundance by internally normalized ratio algorithm with antibody arrays. *Journal of Proteome Research*, 4(3):758–767.

Andresen, H. and Grotzinger, C. (2009). Deciphering the Antibodyome-Peptide Arrays for Serum Antibody Biomarker Diagnostics. *Current Proteomics*, 6(1):1–12.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, 181(4096):223–230.

Angeloni, S., Cordes, R., Dunbar, S. A., Garcia, C., Gibson, G., Martin, C., and Stone, V. (2013). *xMap Cookbook*. Luminex Corp., 1st edition.

Angenendt, P., Kreutzberger, J., Glökler, J., and Hoheisel, J. D. (2006). Generation of high density protein microarrays by cell-free in situ expression of unpurified PCR products. *Molecular & Cellular Proteomics: MCP*, 5(9):1658–1666.

Archelos, J. J. and Hartung, H.-P. (2000). Pathogenetic role of autoantibodies in neurological diseases. *Trends in Neurosciences*, 23(7):317–327.

Aronson, J. K. (2005). Biomarkers and surrogate endpoints. *British Journal of Clinical Pharmacology*, 59(5):491–494.

Atassi, M. Z. and Casali, P. (2008). Molecular mechanisms of autoimmunity. *Autoimmunity*, 41(2):123–132.

Atassi, M. Z. and Smith, J. A. (1978). A proposal for the nomenclature of antigenic sites in peptides and proteins. *Immunochemistry*, 15(8):609–610.

Auger, I., Balandraud, N., Rak, J., Lambert, N., Martin, M., and Roudier, J. (2009). New autoantigens in rheumatoid arthritis (RA): screening 8268 protein arrays with sera from patients with RA. *Annals of the Rheumatic Diseases*, 68(4):591–594.

Avrameas, S. (1991). Natural autoantibodies: from 'horror autotoxicus' to 'gnothi seauton'. *Immunology Today*, 12(5):154–159.

Avrameas, S. and Selmi, C. (2013). Natural autoantibodies in the physiology and pathophysiology of the immune system. *Journal of Autoimmunity*, 41:46–49.

Ayache, S., Panelli, M., Marincola, F. M., and Stroncek, D. F. (2006a). Effects of storage time and exogenous protease inhibitors on plasma protein levels. *American Journal of Clinical Pathology*, 126(2):174–184.

Ayache, S., Panelli, M. C., Byrne, K. M., Slezak, S., Leitman, S. F., Marincola, F. M., and Stroncek, D. F. (2006b). Comparison of proteomic profiles of serum, plasma, and modified media supplements used for cell culture and expansion. *Journal of Translational Medicine*, 4:40.

Ayoglu, B., Chaouch, A., Lochmüller, H., Politano, L., Bertini, E., Spitali, P., Hiller, M., Niks, E. H., Gualandi, F., Ponten, F., Bushby, K., Aartsma-Rus, A., Schwartz, E., Le Priol, Y., Straub, V., Uhlén, M., Cirak, S., t Hoen, P. A. C., Muntoni, F., Ferlini, A., Schwenk, J. M., Nilsson, P., and Al-Khalili Szigyarto, C. (2014a). Affinity proteomics within rare diseases: a BIO-NMD study for blood biomarkers of muscular dystrophies. *EMBO Molecular Medicine*, 6(7):918–936.

Ayoglu, B., Häggmark, A., Khademi, M., Olsson, T., Uhlén, M., Schwenk, J. M., and Nilsson, P. (2013). Autoantibody profiling in multiple sclerosis using arrays of human protein fragments. *Molecular & Cellular Proteomics: MCP*, 12(9):2657–2672.

Ayoglu, B., Häggmark, A., Neiman, M., Igel, U., Uhlén, M., Schwenk, J. M., and Nilsson, P. (2011). Systematic antibody and antigen-based proteomic profiling with microarrays. *Expert Review of Molecular Diagnostics*, 11(2):219–234.

Ayoglu, B., Szarka, E., Huber, K., Orosz, A., Babos, F., Magyar, A., Hudecz, F., Rojkovich, B., Gáti, T., Nagy, G., Schwenk, J. M., Sármay, G., Prechl, J., Nilsson, P., and Papp, K. (2014b). Bead Arrays for Antibody and Complement Profiling Reveal Joint Contribution of Antibody Isotypes to C3 Deposition. *PloS ONE*, 9(5):e96403.

Babel, I., Barderas, R., Díaz-Uriarte, R., Martínez-Torrecuadrada, J. L., Sanchez-Carbayo, M., and Casal, J. I. (2009). Identification of tumor-associated autoantigens for the diagnosis of colorectal cancer in serum using high density protein microarrays. *Molecular & Cellular Proteomics: MCP*, 8(10):2382–2395.

Babel, I., Barderas, R., Diaz-Uriarte, R., Moreno, V., Suarez, A., Fernandez-Acenero, M. J., Salazar, R., Capella, G., and Casal, J. I. (2011). Identification of MST1/STK4 and SULF1 Proteins as Autoantibody Targets for the Diagnosis of Colorectal Cancer by Using Phage Microarrays. *Molecular & Cellular Proteomics: MCP*, 10(3):M110.001784–M110.001784.

Babos, F., Szarka, E., Nagy, G., Majer, Z., Sármay, G., Magyar, A., and Hudecz, F. (2013). Role of N- or C-Terminal Biotinylation in Autoantibody Recognition of Citrullin Containing Filaggrin Epitope Peptides in Rheumatoid

Arthritis. *Bioconjugate Chemistry*, 24(5):817–827.

Bachmann, J., Burté, F., Pramana, S., Conte, I., Brown, B. J., Orimadegun, A. E., Ajetunmobi, W. A., Afolabi, N. K., Akinkunmi, F., Omokhodion, S., Akinbami, F. O., Shokunbi, W. A., Kampf, C., Pawitan, Y., Uhlén, M., Sodeinde, O., Schwenk, J. M., Wahlgren, M., Fernandez-Reyes, D., and Nilsson, P. (2014). Affinity proteomics reveals elevated muscle proteins in plasma of children with cerebral malaria. *PLoS Pathogens*, 10(4):e1004038.

Balboni, I., Chan, S. M., Kattah, M., Tenenbaum, J. D., Butte, A. J., and Utz, P. J. (2006). Multiplexed protein array platforms for analysis of autoimmune diseases. *Annual Review of Immunology*, 24:391–418.

Balboni, I., Limb, C., Tenenbaum, J. D., and Utz, P. J. (2008). Evaluation of microarray surfaces and arraying parameters for autoantibody profiling. *Proteomics*, 8(17):3443–3449.

Balboni, I., Niewold, T. B., Morgan, G., Limb, C., Eloranta, M.-L., Rönnblom, L., Utz, P. J., and Pachman, L. M. (2013). Interferon-α induction and detection of anti-ro, anti-la, anti-sm, and anti-rnp autoantibodies by autoantigen microarray analysis in juvenile dermatomyositis. *Arthritis and Rheumatism*, 65(9):2424–2429.

Bally, M., Syed, S., Binkert, A., Kauffmann, E., Ehrat, M., and Vörös, J. (2011). Fluorescent vesicles for signal amplification in reverse phase protein microarray assays. *Analytical Biochemistry*, 416(2):145–151.

Bandura, D. R., Baranov, V. I., Ornatsky, O. I., Antonov, A., Kinach, R., Lou, X., Pavlov, S., Vorobiev, S., Dick, J. E., and Tanner, S. D. (2009). Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Analytical Chemistry*, 81(16):6813–6822.

Barbe, L., Lundberg, E., Oksvold, P., Stenius, A., Lewin, E., Bjorling, E., Asplund, A., Ponten, F., Brismar, H., Uhlen, M., and Andersson-Svahn, H. (2007). Toward a Confocal Subcellular Atlas of the Human Proteome. *Molecular & Cellular Proteomics: MCP*, 7(3):499–508.

Bartůnková, J., Kayserová, J., and Shoenfeld, Y. (2009). Allergy and autoimmunity: parallels and dissimilarity: the yin and yang of immunopathology. *Autoimmunity Reviews*, 8(4):302–308.

Bauer, J. W., Baechler, E. C., Petri, M., Batliwalla, F. M., Crawford, D., Ortmann, W. A., Espe, K. J., Li, W., Patel, D. D., Gregersen, P. K., and Behrens, T. W. (2006). Elevated serum levels of interferon-regulated chemokines are biomarkers for active human systemic lupus erythematosus. *PLoS Medicine*, 3(12):e491.

Behring, E. and Kitasato, S. (1890). Über das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren. *DMW - Deutsche Medizinische Wochenschrift*, 16:1113–1114.

Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schölkopf, B., and Rätsch, G. (2008). Support vector machines and kernels for computational biology. *PLoS computational biology*, 4(10):e1000173.

Bendall, S. C., Simonds, E. F., Qiu, P., Amir, E.-a. D., Krutzik, P. O., Finck, R., Bruggner, R. V., Melamed, R., Trejo, A., Ornatsky, O. I., Balderas, R. S., Plevritis, S. K., Sachs, K., Pe'er, D., Tanner, S. D., and Nolan, G. P. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*, 332(6030):687–696.

Berglund, L., Andrade, J., Odeberg, J., and Uhlén, M. (2008a). The epitope space of the human proteome. *Protein Science*, 17(4):606–613.

Berglund, L., Björling, E., Jonasson, K., Rockberg, J., Fagerberg, L., Al-Khalili Szigyarto, C., Sivertsson, A., and Uhlén, M. (2008b). A whole-genome bioinformatics approach to selection of antigens for systematic antibody generation. *Proteomics*, 8(14):2832–2839.

Bernard, C. (1974). *Lectures on the Phenomena of Life Common to Animals and Plants. Volume I of Translation by Hoff H E, Guillemin R, Guillemin L* . American Lecture Series.

Beyer, M., Nesterov, A., Block, I., König, K., Felgenhauer, T., Fernandez, S., Leibe, K., Torralba, G., Hausmann, M., Trunk, U., Lindenstruth, V., Bischoff, F. R., Stadler, V., and Breitling, F. (2007). Combinatorial synthesis of peptide arrays onto a microchip. *Science*, 318(5858):1888.

Beyer, N. H., Lueking, A., Kowald, A., Frederiksen, J. L., and Heegaard, N. H. (2012). Investigation of autoantibody profiles for cerebrospinal fluid biomarker discovery in patients with relapsing-remitting multiple sclerosis. *Journal of Neuroimmunology*, 242(1-2):7–7.

Biemann, K., Gapp, G., and Seibl, J. (1959). Application of Mass Spectrometry to Structure Problems. I. Amino Acid Sequence in Peptides. *Journal of the American Chemical Society*, 81(9):2274–2275.

Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P., and Plückthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *Journal of Molecular Biology*, 332(2):489–503.

Björling, E. and Uhlén, M. (2008). Antibodypedia, a portal for sharing antibody and antigen validation data. *Molecular & Cellular Proteomics: MCP*, 7(10):2028–2037.

Bland, J. M. and Altman, D. G. (1994). Matching. *BMJ*, 309(6962):1128.

Bland, J. M. and Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *BMJ*, 310(6973):170.

Bland, J. M. and Altman, D. G. (1996). Transforming data. *BMJ*, 312(7033):770.

Blank, M., Barzilai, O., and Shoenfeld, Y. (2007). Molecular mimicry and auto-immunity. *Clinical Reviews in Allergy & Immunology*, 32(1):111–118.

Blokzijl, A., Nong, R., Darmanis, S., Hertz, E., Landegren, U., and Kamali-Moghaddam, M. (2014). Protein biomarker validation via proximity ligation assays. *Biochimica et Biophysica Acta*, 1844(5):933–939.

Boder, E. T. and Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology*, 15(6):553–557.

Boja, E. S. and Rodriguez, H. (2012). Mass spectrometry-based targeted quantitative proteomics: Achieving sensitive and reproducible detection of proteins. *Proteomics*, 12(8):1093–1110.

Bordet, J. and Gengou, O. (1901). Sur l'existence de substances sensibilisatrices dans la plupart des serum antimicrobiens. *Ann Inst Pasteur*.

Boutell, J. M., Hart, D. J., Godber, B. L. J., Kozlowski, R. Z., and Blackburn, J. M. (2004). Functional protein microarrays for parallel characterisation of p53 mutants. *Proteomics*, 4(7):1950–1958.

Boyd, Z. S., Wu, Q. J., O'Brien, C., Spoerke, J., Savage, H., Fielder, P. J., Amler, L., Yan, Y., and Lackner, M. R. (2008). Proteomic analysis of breast cancer molecular subtypes and biomarkers of response to targeted kinase inhibitors using reverse-phase protein microarrays. *Molecular cancer therapeutics*, 7(12):3695–3706.

Bradbury, A., Velappan, N., Verzillo, V., Ovecka, M., Chasteen, L., Sblattero, D., Marzari, R., Lou, J., Siegel, R., and Pavlik, P. (2003). Antibodies in proteomics I: generating antibodies. *Trends in Biotechnology*, 21(6):275–281.

Bradbury, A. R. M., Sidhu, S., Dübel, S., and McCafferty, J. (2011). Beyond natural antibodies: the power of in vitro display technologies. *Nature Biotechnology*, 29(3):245–254.

Brancaccio, P., Lippi, G., and Maffulli, N. (2010). Biochemical markers of muscular damage. *Clinical Chemistry and Laboratory Medicine*, 48(6).

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Breitling, F., Nesterov, A., Stadler, V., Felgenhauer, T., and Bischoff, F. R. (2009). High-density peptide arrays. *Molecular BioSystems*, 5(3):224–234.

Brennan, K. M., Galban-Horcajo, F., Rinaldi, S., O'Leary, C. P., Goodyear, C. S., Kalna, G., Arthur, A., Elliot, C., Barnett, S., Linington, C., Bennett, J. L., Owens, G. P., and Willison, H. J. (2011). Lipid arrays identify myelin-derived lipids and lipid complexes as prominent targets for oligoclonal band antibodies in multiple sclerosis. *Journal of Neuroimmunology*, 238(1-2):87–95.

Bridgman, P. W. (1955). *Reflections of a Physicist*. Philosophical Library, 1st edition.

Britschgi, M., Olin, C. E., Johns, H. T., Takeda-Uchimura, Y., LeMieux, M. C., Rufibach, K., Rajadas, J., Zhang, H., Tomooka, B., Robinson, W. H., Clark, C. M., Fagan, A. M., Galasko, D. R., Holtzman, D. M., Jutel, M., Kaye, J. A., Lemere, C. A., Leszek, J., Li, G., Peskind, E. R., Quinn, J. F., Yesavage, J. A., Ghiso, J. A., and Wyss-Coray, T. (2009). Neuroprotective natural antibodies to assemblies of amyloidogenic peptides decrease with normal aging and advancing Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America*, 106(29):12145–12150.

Britschgi, M., Rufibach, K., Huang, S. L. B., Clark, C. M., Kaye, J. A., Li, G., Peskind, E. R., Quinn, J. F., Galasko, D. R., and Wyss-Coray, T. (2011). Modeling of pathological traits in Alzheimer's disease based on systemic extracellular signaling proteome. *Molecular & Cellular Proteomics: MCP*, 10(10):M111.008862.

Buchner, H. (1889). Ueber die nähere Natur der bakterientödtenden Substanz im Blutserum. *Centralblatt für Bakteriologie*.

Burnet, F. M. (1957). A Modification of Jerne's Theory of Antibody Production using the Concept of Clonal Selection. *The Australian Journal of Science*, 20(3):67–69.

Bushby, K., Finkel, R., Birnkrant, D. J., Case, L. E., Clemens, P. R., Cripe, L., Kaul, A., Kinnett, K., McDonald, C., and Pandya, S. (2010). Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and pharmacological and psychosocial management. *The Lancet Neurology*, 9(1):77–93.

Butte, A. J., Sigdel, T. K., Wadia, P. P., Miklos, D. B., and Sarwal, M. M. (2011). Protein microarrays discover angiotensinogen and PRKRIP1 as novel targets for autoantibodies in chronic renal disease. *Molecular & Cellular Proteomics: MCP*, 10(3):M110.000497.

Buus, S., Rockberg, J., Forsström, B., Nilsson, P., Uhlén, M., and Schafer-Nielsen, C. (2012). High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. *Molecular & Cellular Proteomics: MCP*, 11(12):1790–1800.

Caiazzo Jr, R., Maher, A., Drummond, M., Lander, C., Tassinari, O., Nelson, B., and Liu, B. (2009). Protein microarrays as an application for disease biomarkers. *Proteomics - Clinical Applications*, 3(2):138–147.

Carlson, E. D., Gan, R., Hodgman, C. E., and Jewett, M. C. (2012). Cell-free protein synthesis: applications come of age. *Biotechnology advances*, 30(5):1185–1194.

Carlsson, A., Wingren, C., Ingvarsson, J., Ellmark, P., Baldertorp, B., Fernö, M., Olsson, H., and Borrebaeck, C. A. K. (2008). Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays. *European Journal of Cancer*, 44(3):472–480.

Carlsson, A., Wingren, C., Kristensson, M., Rose, C., Fernö, M., Olsson, H., Jernström, H., Ek, S., Gustavsson, E., Ingvar, C., Ohlsson, M., Peterson, C., and Borrebaeck, C. A. K. (2011a). Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34):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. *Molecular & Cellular Proteomics: MCP*, 10(5):M110.005033.

Carrascal, M., Gay, M., Ovelleiro, D., Casas, V., Gelpí, E., and Abian, J. (2010). Characterization of the human plasma phosphoproteome using linear ion trap mass spectrometry and multiple search engines. *Journal of Proteome Research*, 9(2):876–884.

Cekaite, L., Haug, O., Myklebost, O., Aldrin, M., Østenstad, B., Holden, M., Frigessi, A., Hovig, E., and Sioud, M. (2004). Analysis of the humoral immune response to immunoselected phage-displayed peptides by a microarray-based method. *Proteomics*, 4(9):2572–2582.

Cekaite, L., Hovig, E., and Sioud, M. (2009). Monitoring B Cell Response to Immunoselected Phage-Displayed Peptides by Microarrays. In *Methods in Molecular Biology*, pages 273–285. Humana Press, Totowa, NJ.

Chan, S. M., Ermann, J., Su, L., Fathman, C. G., and Utz, P. J. (2004). Protein microarrays for multiplex analysis of signal transduction pathways. *Nature Medicine*, 10(12):1390–1396.

Chandra, H., Reddy, P. J., and Srivastava, S. (2011). Protein microarrays and novel detection platforms. *Expert Review of Proteomics*, 8(1):61–79.

Chandra, H. and Srivastava, S. (2010). Cell-free synthesis-based protein microarrays and their applications. *Proteomics*, 10(4):717–730.

Chatterjee, M. (2006). Diagnostic Markers of Ovarian Cancer by High-Throughput Antigen Cloning and Detection on Arrays. *Cancer Research*, 66(2):1181–1190.

Chatterjee, S. K. and Zetter, B. R. (2005). Cancer biomarkers: knowing the present and predicting the future. *Future Oncology*, 1(1):37–50.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., and Fodor, S. P. (1996). Accessing genetic information with high-density DNA arrays. *Science*, 274(5287):610–614.

Chen, C.-S., Korobkova, E., Chen, H., Zhu, J., Jian, X., Tao, S.-C., He, C., and Zhu, H. (2008). A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli. *Nature Methods*, 5(1):69–74.

Chen, G., Wang, X., Yu, J., Varambally, S., Yu, J., Thomas, D. G., Lin, M.-Y., Vishnu, P., Wang, Z., Wang, R., Fielhauer, J., Ghosh, D., Giordano, T. J., Giacherio, D., Chang, A. C., Orringer, M. B., El-Hefnawy, T., Bigbee, W. L., Beer, D. G., and Chinnaiyan, A. M. (2007). Autoantibody profiles reveal ubiquilin 1 as a humoral immune response target in lung adenocarcinoma. *Cancer Research*, 67(7):3461–3467.

Chiari, M., Cretich, M., Corti, A., Damin, F., Pirri, G., and Longhi, R. (2005). Peptide microarrays for the characterization of antigenic regions of human chromogranin A. *Proteomics*, 5(14):3600–3603.

Christen, U. and Herrath, M. G. v. (2004). Initiation of autoimmunity. *Current Opinion In Immunology*, 16(6):759–767.

Christin, C., Hoefsloot, H. C. J., Smilde, A. K., Hoekman, B., Suits, F., Bischoff, R., and Horvatovich, P. (2013). A critical assessment of feature selection methods for biomarker discovery in clinical proteomics. *Molecular & Cellular Proteomics: MCP*, 12(1):263–276.

Chruscinski, A. J., Singh, H., Chan, S. M., and Utz, P. J. (2013). Broad-Scale Phosphoprotein Profiling of Beta Adrenergic Receptor (β-AR) Signaling Reveals Novel Phosphorylation and Dephosphorylation Events. *PloS ONE*, 8(12):e82164.

Clackson, T., Hoogenboom, H. R., Griffiths, A. D., and Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature*, 352(6336):624–628.

Colwill, K., Persson, H., Jarvik, N. E., Wyrzucki, A., Wojcik, J., Koide, A., Kossiakoff, A. A., Koide, S., Sidhu, S., Dyson, M. R., Pershad, K., Pavlovic, J. D., Karatt-Vellatt, A., Schofield, D. J., Kay, B. K., McCafferty, J., Mersmann, M., Meier, D., Mersmann, J., Helmsing, S., Hust, M., Dübel, S., Berkowicz, S., Freemantle, A., Spiegel, M., Sawyer, A., Layton, D., Nice, E., Dai, A., Rocks, O., Williton, K., Fellouse, F. A., Hersi, K., Pawson, T., Nilsson, P., Sundberg, M., Sjöberg, R., Sivertsson, A., Schwenk, J. M., Takanen, J. O., Hober, S., Uhlén, M., Dahlgren, L.-G., Flores, A., Johansson, I., Weigelt, J., Crombet, L., Loppnau, P., Kozieradzki, I., Cossar, D., Arrowsmith, C. H., Edwards, A. M., and Gräslund, S. (2011). A roadmap to generate renewable protein binders to the human proteome. *Nature Methods*, 8(7):551–558.

Compston, A. and Coles, A. (2008). Multiple sclerosis. *The Lancet*, 372(9648):1502–1517.

Coutinho, A., Kazatchkine, M. D., and Avrameas, S. (1995). Natural autoantibodies. *Current Opinion In Immunology*, 7(6):812–818.

Cravatt, B. F., Simon, G. M., and Yates, III, J. R. (2007). The biological impact of mass-spectrometry-based proteomics. *Nature*, 450(7172):991–1000.

Cretu, D., Diamandis, E. P., and Chandran, V. (2013). Delineating the synovial fluid proteome: recent advancements and ongoing challenges in biomarker research. *Critical reviews in clinical laboratory sciences*, 50(2):51–63.

Crick, F. H. (1970). Central dogma of molecular biology. *Nature*, 227(5258):561–563.

Crick, F. H., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature*, 192:1227–1232.

Crompton, P. D., Kayala, M. A., Traore, B., Kayentao, K., Ongoiba, A., Weiss, G. E., Molina, D. M., Burk, C. R., Waisberg, M., Jasinskas, A., Tan, X., Doumbo, S., Doumtabe, D., Kone, Y., Narum, D. L., Liang, X., Doumbo, O. K., Miller, L. H., Doolan, D. L., Baldi, P., Felgner, P. L., and Pierce, S. K. (2010). A prospective analysis of the Ab response to Plasmodium falciparum before and after a malaria season by protein microarray. *Proceedings of the National Academy of Sciences of the United States of America*, 107(15):6958–6963.

D'Alessandro, A., Righetti, P. G., and Zolla, L. (2010). The red blood cell proteome and interactome: an update. *Journal of Proteome Research*, 9(1):144–163.

D'Angelo, S., Mignone, F., Deantonio, C., Di Niro, R., Bordoni, R., Marzari, R., De Bellis, G., Not, T., Ferrara, F., Bradbury, A., Santoro, C., and Sblattero, D. (2013). Profiling celiac disease antibody repertoire. *Clinical Immunology*, 148(1):99–109.

Darmanis, S., Cui, T., Drobin, K., Li, S.-C., Oberg, K., Nilsson, P., Schwenk, J. M., and Giandomenico, V. (2013). Identification of candidate serum proteins for classifying well-differentiated small intestinal neuroendocrine tumors. *PloS ONE*, 8(11):e81712.

Darmanis, S., Nong, R. Y., Hammond, M., Gu, J., Alderborn, A., Vänelid, J., Siegbahn, A., Gustafsdottir, S., Ericsson, O., Landegren, U., and Kamali-Moghaddam, M. (2010). Sensitive plasma protein analysis by microparticle-based proximity ligation assays. *Molecular & Cellular Proteomics: MCP*, 9(2):327–335.

Darmanis, S., Nong, R. Y., Vänelid, J., Siegbahn, A., Ericsson, O., Fredriksson, S., Bäcklin, C., Gut, M., Heath, S., Gut, I. G., Wallentin, L., Gustafsson, M. G., Kamali-Moghaddam, M., and Landegren, U. (2011). ProteinSeq: high-performance proteomic analyses by proximity ligation and next generation sequencing. *PloS ONE*, 6(9):e25583.

Dasso, J., Lee, J., Bach, H., and Mage, R. G. (2002). A comparison of ELISA and flow microsphere-based assays for quantification of immunoglobulins. *Journal of Immunological Methods*, 263(1-2):23–33.

Daugherty, P. S., Chen, G., Olsen, M. J., Iverson, B. L., and Georgiou, G. (1998). Antibody affinity maturation using bacterial surface display. *Protein engineering*, 11(9):825–832.

Davidson, A. and Diamond, B. (2001). Autoimmune diseases. *The New England Journal of Medicine*, 345(5):340–350.

D'Cruz, D. (2000). Autoimmune diseases associated with drugs, chemicals and environmental factors. *Toxicology Letters*, 112-113:421–432.

de Hoffmann and Vincent Stroobant, E. (2007). *Mass Spectrometry*. John Wiley & Sons, Ltd, 3rd edition.

De Jager, W., Prakken, B. J., Bijlsma, J. W. J., Kuis, W., and Rijkers, G. T. (2005). Improved multiplex immunoassay performance in human plasma and synovial fluid following removal of interfering heterophilic antibodies. *Journal of Immunological Methods*, 300(1-2):124–135.

De Jager, W., te Velthuis, H., Prakken, B. J., Kuis, W., and Rijkers, G. T. (2003). Simultaneous detection of 15 human cytokines in a single sample of stimulated peripheral blood mononuclear cells. *Clinical and Vaccine Immunology*, 10(1):133–139.

de Wit, M., Fijneman, R. J. A., Verheul, H. M. W., Meijer, G. A., and Jimenez, C. R. (2013). Proteomics in colorectal cancer translational research: Biomarker discovery for clinical applications. *Clinical Biochemistry*, 46(6):466–479.

Delehanty, J. B. (2004). Printing Functional Protein Microarrays Using Piezoelectric Capillaries. In *Protein Arrays*, pages 135–144. Humana Press, New Jersey.

Delehanty, J. B. and Ligler, F. S. (2003). Method for Printing Functional Protein Microarrays. *BioTechniques*, 34(2):380–385.

Dezfouli, M., Vickovic, S., Iglesias, M. J., Nilsson, P., Schwenk, J. M., and Ahmadian, A. (2014). Magnetic bead assisted labeling of antibodies at nanogram scale. *Proteomics*, 14(1):14–18.

Diamandis, E. P. (2014). Present and future of cancer biomarkers. *Clinical Chemistry and Laboratory Medicine*, 52(6):791–794.

Diamandis, E. P. and Christopoulos, T. K. (1991). The biotin-(strept)avidin system: principles and applications in biotechnology. *Clinical Chemistry*, 37(5):625–636.

Dieterle, F., Ross, A., Schlotterbeck, G., and Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Analytical Chemistry*, 78(13):4281–4290.

Dighiero, G. and Rose, N. R. (1999). Critical self-epitopes are key to the understanding of self-tolerance and autoimmunity. *Immunology Today*, 20(9):423–428.

Domon, B. and Aebersold, R. (2010). Options and considerations when selecting a quantitative proteomics strategy. *Nature Biotechnology*, 28(7):710–721.

Donath, J. and Landsteiner, K. (1904). Ueber paroxysmale Hämoglobinurie. *Münchener medizinische Wochenschrift*, 51:1590–1603.

Doniach, D. and Roitt, I. M. (1957). Auto-immunity in Hashimoto's disease and its implications. *The Journal of Clinical Endocrinology & Metabolism*, 17(11):1293–1304.

Drabovich, A. P., Saraon, P., Jarvi, K., and Diamandis, E. P. (2014). Seminal plasma as a diagnostic fluid for male reproductive system disorders. *Nature reviews. Urology*, 11(5):278–288.

Drobin, K., Nilsson, P., and Schwenk, J. M. (2013). Highly multiplexed antibody suspension bead arrays for plasma protein profiling. *Methods in Molecular Biology*, 1023:137–145.

Dübel, S., Stoevesandt, O., Taussig, M. J., and Hust, M. (2010). Generating recombinant antibodies to the complete human proteome. *Trends in Biotechnology*, 28(7):333–339.

Dunkelberger, J. R. and Song, W.-C. (2010). Complement and its role in innate and adaptive immune responses. *Cell research*, 20(1):34–50.

duPont, N. C., Wang, K., Wadhwa, P. D., Culhane, J. F., and Nelson, E. L. (2005). Validation and comparison of luminex multiplex cytokine analysis kits with ELISA: determinations of a panel of nine cytokines in clinical sample culture supernatants. *Journal of reproductive immunology*, 66(2):175–191.

Eaton, W. W., Rose, N. R., Kalaydjian, A., Pedersen, M. G., and Mortensen, P. B. (2007). Epidemiology of autoimmune diseases in Denmark. *Journal of Autoimmunity*, 29(1):1–9.

Edelman, G. M. (1959). Dissociation of γ-globulin. *Journal of the American Chemical Society*, 81(12):3155–3156.

Edelman, G. M. (1970). The structure and function of antibodies. *Scientific American*, 223(2):34–42.

Ehrlich, P. (1891). Experimentelle Untersuchungen über Immunität. II. Über Abrin. *DMW - Deutsche Medizinische Wochenschrift*, 17(44):1218–1219.

Ehrlich, P. and Morgenroth, J. (1899). Ueber Hämolysine, II. Mitteilung . *Berliner Klinische Wochenschrift*, 2:156–164.

Ehrlich, P. and Morgenroth, J. (1901). Ueber Hämolysine: fünfte Mittheilung. *Berliner Klinische Wochenschrift*, pages 251–257.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25):14863–14868.

Eisenstein, M. (2006). Protein arrays: growing pains. *Nature*, 444(7121):959–962.

Ekins, R. P. (1989). Multi-analyte immunoassay. *Journal of pharmaceutical and biomedical analysis*, 7(2):155–168.

Ekins, R. P. and Chu, F. (1994). Developing multianalyte assays. *Trends in Biotechnology*, 12(3):89–94.

Ekins, R. P. and Chu, F. W. (1991). Multianalyte microspot immunoassay-microanalytical" compact disk" of the future. *Clinical Chemistry*, 37(11):1955–1967.

Ekins, R. P. and Edwards, P. (1997). On the meaning of "sensitivity". *Clinical Chemistry*, 43(10):1824–1831.

Elgert, K. D. (2009). *Immunology: Understanding The Immune System*. John Wiley & Sons, New Jersey, 2nd edition.

Elkon, K. and Casali, P. (2008). Nature and functions of autoantibodies. *Nature Clinical Practice. Rheumatology*, 4(9):491–498.

Ellington, A. D. and Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346(6287):818–822.

Elshal, M. F. and McCoy, J. P. (2006). Multiplex bead array assays: performance evaluation and comparison of sensitivity to ELISA. *Methods*, 38(4):317–323.

Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., Weiss, R. A., and Liotta, L. A. (1996). Laser Capture Microdissection. *Science*, 274(5289):998–1001.

Endo, Y. and Sawasaki, T. (2006). Cell-free expression systems for eukaryotic protein production. *Current Opinion In Biotechnology*, 17(4):373–380.

Engvall, E. and Perlmann, P. (1971). Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. *Immunochemistry*, 8(9):871–874.

Erdei, A., Fust, G., and Gergely, J. (1991). The role of C3 in the immune response. *Immunology Today*, 12(9):332–337.

Eriksson, L., Kettaneh-Wold, N., Trygg, J., Wikström, C., and Wold, S. (2006). *Multi- and Megavariate Data Analysis : Part I: Basic Principles and Applications*. Umetrics Inc, 2nd edition.

Escher, C., Lochmüller, H., Fischer, D., Frank, S., Reimann, J., Walter, M. C., Ehrat, M., Ruegg, M. A., and Gygax, D. (2010). Reverse protein arrays as novel approach for protein quantification in muscular dystrophies. *Neuromuscular Disorders: NMD*, 20(5):302–309.

Espina, V., Mehta, A. I., Winters, M. E., Calvert, V., Wulfkuhle, J., Petricoin, E. F., and Liotta, L. A. (2003). Protein microarrays: Molecular profiling technologies for clinical specimens. *Proteomics*, 3(11):2091–2100.

Etzioni, A. (2003). Immune deficiency and autoimmunity. *Autoimmunity Reviews*, 2(6):364–369.

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. *Nature Reviews Cancer*, 3(4):243–252.

Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjöstedt, E., Lundberg, E., Szigyarto, C. A.-K., Skogs, M., Takanen, J. O., Berling, H., Tegel, H., Mulder, J., Nilsson, P., Schwenk, J. M., Lindskog, C., Danielsson, F., Mardinoglu, A., Sivertsson, A., von Feilitzen, K., Forsberg, M., Zwahlen, M., Olsson, I., Navani, S., Huss, M., Nielsen, J., Ponten, F., and Uhlén, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics: MCP*, 13(2):397–406.

Farias, A. S., Pradella, F., Schmitt, A., Santos, L. M. B., and Martins-de Souza, D. (2014). Ten years of proteomics in multiple sclerosis. *Proteomics*, 14(4-5):467–480.

Farina, A. (2014). Proximal fluid proteomics for the discovery of digestive cancer biomarkers. *Biochimica et Biophysica Acta*, 1844(5):988–1002.

Farrah, T., Deutsch, E. W., Omenn, G. S., Campbell, D. S., Sun, Z., Bletz, J. A., Mallick, P., Katz, J. E., Malmström, J., Ossola, R., Watts, J. D., Lin, B., Zhang, H., Moritz, R. L., and Aebersold, R. (2011). A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Molecular & Cellular Proteomics: MCP*, 10(9):M110–006353.

Farrah, T., Deutsch, E. W., Omenn, G. S., Sun, Z., Watts, J. D., Yamamoto, T., Shteynberg, D., Harris, M. M., and Moritz, R. L. (2014). State of the human proteome in 2013 as viewed through PeptideAtlas: comparing the kidney, urine, and plasma proteomes for the biology- and disease-driven Human Proteome Project. *Journal of Proteome Research*, 13(1):60–75.

Fattal, I., Shental, N., Mevorach, D., Anaya, J.-M., Livneh, A., Langevitz, P., Zandman-Goddard, G., Pauzner, R., Lerner, M., Blank, M., Hincapie, M.-E., Gafter, U., Naparstek, Y., Shoenfeld, Y., Domany, E., and Cohen, I. R. (2010). An antibody profile of systemic lupus erythematosus detected by antigen microarray. *Immunology*, 130(3):337–343.

Fearon, D. T. (1997). Seeking wisdom in innate immunity. *Nature*, 388(6640):323–324.

Fearon, D. T. and Locksley, R. M. (1996). The Instructive Role of Innate Immunity in the Acquired Immune Response. *Science*, 272(5258):50–54.

Feinberg, J. G. (1961). A 'Microspot' Test for Antigens and Antibodies. *Nature*, 192(4806):985–986.

Feinberg, J. G. and Wheeler, A. W. (1963). Detection of auto-immune antibody and tissue antigens by the 'microspot' technique. *Journal of Clinical Pathology*, 16:282–284.

Feng, Y., Ke, X., Ma, R., Chen, Y., Hu, G., and Liu, F. (2004). Parallel detection of autoantibodies with microarrays in rheumatoid diseases. *Clinical Chemistry*, 50(2):416–422.

Fenn, J., Mann, M., Meng, C., Wong, S., and Whitehouse, C. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71.

Ferenčík, M. (1993). The complement system. *Handbook of Immunochemistry*.

Ferguson, R. E., Hochstrasser, D. F., and Banks, R. E. (2007). Impact of preanalytical variables on the analysis of biological fluids in proteomic studies. *Proteomics - Clinical Applications*, 1(8):739–746.

Ferrata, A. (1907). Die Unwirksamkeit der Komplexen Hämolysine in salzfreien Lösungen und ihre Ursache. *Berliner Klinische Wochenschrift*.

Firestein, G. S. (2003). Evolving concepts of rheumatoid arthritis. *Nature*, 423(6937):356–361.

Fischer, E. (1906). Untersuchungen über Aminosäuren, Polypeptide und Proteïne. *Berichte der deutschen chemischen Gesellschaft*, 39(1):530–610.

Fisher, R. A. (1942). *The Design of Experiments*. Hafner Publishing Company New York, 3rd edition.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S., Johnson, N., Juettemann, T., Kähäri, A. K., Keenan, S., Kulesha, E., Martin, F. J., Maurel, T., McLaren, W. M., Murphy, D. N., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H. S., Ruffier, M., Sheppard, D., Taylor, K., Thormann, A., Trevanion, S. J., Vullo, A., Wilder, S. P., Wilson, M., Zadissa, A., Aken, B. L., Birney, E., Cunningham, F., Harrow, J., Herrero, J., Hubbard, T. J. P., Kinsella, R., Muffato, M., Parker, A., Spudich, G., Yates, A., Zerbino, D. R., and Searle, S. M. J. (2014). Ensembl 2014. *Nucleic Acids Research*, 42(Database issue):D749–55.

Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., and Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science*, 251(4995):767–773.

Forsström, B., Axnäs, B. B., Stengele, K.-P., Bühler, J., Albert, T. J., Richmond, T. A., Hu, F. J., Nilsson, P., Hudson, E. P., Rockberg, J., and Uhlén, M. (2014). Proteome-wide Epitope Mapping of Antibodies Using Ultra-dense Peptide Arrays. *Molecular & Cellular Proteomics: MCP*, 13(6):1585–1597.

Fournel, S. and Muller, S. (2003). Synthetic peptides in the diagnosis of systemic autoimmune diseases. *Current protein & peptide science*, 4(4):261–274.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., and Jensen, L. J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*, 41(Database issue):D808–15.

Frank, R. (1992). Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron*, 48(42):9217–9232.

Frank, R. and Döring, R. (1988). Simultaneous multiple peptide synthesis under continuous flow conditions on cellulose paper discs as segmental solid supports. *Tetrahedron*, 44(19):6031–6040.

Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gústafsdóttir, S. M., Östman, A., and Landegren, U. (2002). Protein detection using proximity-dependent DNA ligation assays. *Nature Biotechnology*, 20(5):473–477.

Fujita, T., Matsushita, M., and Endo, Y. (2004). The lectin complement pathway-its role in innate immunity and evolution. *Immunological Reviews*, 198(1):185–202.

Fulton, R. J., McDade, R. L., Smith, P. L., Kienker, L. J., and Kettman, J. R. (1997). Advanced multiplexed analysis with the FlowMetrix system. *Clinical Chemistry*, 43(9):1749–1756.

Füzery, A. K., Levin, J., Chan, M. M., and Chan, D. W. (2013). Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. *Clinical Proteomics*, 10(1):1–1.

Gallien, S., Duriez, E., and Domon, B. (2011). Selected reaction monitoring applied to proteomics. *Journal of Mass Spectrometry*, 46(3):298–312.

Gao, W.-M., Kuick, R., Orchekowski, R. P., Misek, D. E., Qiu, J., Greenberg, A. K., Rom, W. N., Brenner, D. E., Omenn, G. S., Haab, B. B., and Hanash, S. M. (2005). Distinctive serum protein profiles involving abundant proteins in lung cancer patients based upon antibody microarray analysis. *BMC Cancer*, 5:110.

Gao, X., Pellois, J. P., Na, Y., Kim, Y., Gulari, E., and Zhou, X. (2004). High density peptide microarrays. In situ synthesis and applications. *Molecular Diversity*, 8(3):177–187.

Gao, X., Zhou, X., and Gulari, E. (2003). Light directed massively parallel on-chip synthesis of peptide arrays with t-Boc chemistry. *Proteomics*, 3(11):2135–2141.

George, J., Gilburd, B., and Shoenfeld, Y. (1997). The emerging concept of pathogenic natural autoantibodies. *Human Antibodies*, 8(2):70–75.

Georgiou, G., Ippolito, G. C., Beausang, J., Busse, C. E., Wardemann, H., and Quake, S. R. (2014). The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature Biotechnology*, 32(2):158–168.

Geysen, H. M., Barteling, S. J., and Meloen, R. H. (1985). Small peptides induce antibodies with a sequence and structural requirement for binding antigen comparable to antibodies raised against the native protein. *Proceedings of the National Academy of Sciences of the United States of America*, 82(1):178–182.

Geysen, H. M., Meloen, R. H., and Barteling, S. J. (1984). Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proceedings of the National Academy of Sciences of the United States of America*, 81(13):3998–4002.

Gibson, D. S., Qiu, J., Mendoza, E. A., Barker, K., Rooney, M. E., and LaBaer, J. (2012). Circulating and synovial antibody profiling of juvenile arthritis patients by nucleic acid programmable protein arrays. *Arthritis research & therapy*, 14(2):R77.

Gibson, D. S. and Rooney, M. E. (2007). The human synovial fluid proteome: A key factor in the pathology of joint disease - Gibson - 2007 - PROTEOMICS - Clinical Applications - Wiley Online Library. *Proteomics - Clinical Applications*.

Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics: MCP*, 11(6):O111.016717.

Giovannoni, G., Munschauer, F. E., and Deisenhammer, F. (2002). Neutralising antibodies to interferon beta during the treatment of multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 73(5):465–469.

Gleicher, N. and Barad, D. H. (2007). Gender as risk factor for autoimmune diseases. *Journal of Autoimmunity*, 28(1):1–6.

Glick, S. (2011). Rosalyn Sussman Yalow (1921-2011). *Nature*, 474(7353):580.

Gnjatic, S., Ritter, E., Büchler, M. W., Giese, N. A., Brors, B., Frei, C., Murray, A., Halama, N., Zörnig, I., Chen, Y.-T., Andrews, C., Ritter, G., Old, L. J., Odunsi, K., and Jäger, D. (2010). Seromic profiling of ovarian and pancreatic cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11):5088–5093.

Gnjatic, S., Wheeler, C., Ebner, M., Ritter, E., Murray, A., Altorki, N. K., Ferrara, C. A., Hepburne-Scott, H., Joyce, S., Koopman, J., McAndrew, M. B., Workman, N., Ritter, G., Fallon, R., and Old, L. J. (2009). Seromic analysis of antibody responses in non-small cell lung cancer patients and healthy donors using conformational protein arrays. *Journal of Immunological Methods*, 341(1-2):50–58.

Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E. N., Carter, J., Dalby, A. B., Eaton, B. E., Fitzwater, T., Flather, D., Forbes, A., Foreman, T., Fowler, C., Gawande, B., Goss, M., Gunn, M., Gupta, S., Halladay, D., Heil, J., Heilig, J., Hicke, B., Husar, G., Janjic, N., Jarvis, T., Jennings, S., Katilius, E., Keeney, T. R., Kim, N., Koch, T. H., Kraemer, S., Kroiss, L., Le, N., Levine, D., Lindsey, W., Lollo, B., Mayfield, W., Mehan, M., Mehler, R., Nelson, S. K., Nelson, M., Nieuwlandt, D., Nikrad, M., Ochsner, U., Ostroff, R. M., Otis, M., Parker, T., Pietrasiewicz, S., Resnicow, D. I., Rohloff, J., Sanders, G., Sattin, S., Schneider, D., Singer, B., Stanton, M., Sterkel, A., Stewart, A., Stratford, S., Vaught, J. D., Vrkljan, M., Walker, J. J., Watrobka, M., Waugh, S., Weiss, A., Wilcox, S. K., Wolfson, A., Wolk, S. K., Zhang, C., and Zichi, D. (2010). Aptamer-based multiplexed proteomic technology for biomarker discovery. *PloS ONE*, 5(12):e15004.

Goodnow, C. C., Sprent, J., Fazekas de St Groth, B., and Vinuesa, C. G. (2005). Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature*, 435(7042):590–597.

Graham, K. L., Vaysberg, M., Kuo, A., and Utz, P. J. (2006). Autoantigen arrays for multiplex analysis of antibody isotypes. *Proteomics*, 6(21):5720–5724.

Granger, J., Siddiqui, J., Copeland, S., and Remick, D. (2005). Albumin depletion of human plasma also removes low abundance proteins including the cytokines. *Proteomics*, 5(18):4713–4718.

Grimes, D. A. and Schulz, K. F. (2005). Compared to what? Finding controls for case-control studies. *The Lancet*, 365(9468):1429–1433.

Grönwall, C., Vas, J., and Silverman, G. J. (2012). Protective Roles of Natural IgM Antibodies. *Frontiers in immunology*, 3:66.

Grote, T., Siwak, D. R., Fritsche, H. A., Joy, C., Mills, G. B., Simeone, D., Whitcomb, D. C., and Logsdon, C. D. (2008). Validation of reverse phase protein array for practical screening of potential biomarkers in serum and plasma: accurate detection of CA19-9 levels in pancreatic cancer. *Proteomics*, 8(15):3051–3060.

Grubb, R. L., Calvert, V. S., Wulkuhle, J. D., Paweletz, C. P., Linehan, W. M., Phillips, J. L., Chuaqui, R., Valasco, A., Gillespie, J., Emmert-Buck, M., Liotta, L. A., and Petricoin, E. F. (2003). Signal pathway profiling of prostate cancer using reverse phase protein arrays. *Proteomics*, 3(11):2142–2146.

Guilleaume, B., Buness, A., Schmidt, C., Klimek, F., Moldenhauer, G., Huber, W., Arlt, D., Korf, U., Wiemann, S., and Poustka, A. (2005). Systematic comparison of surface coatings for protein microarrays. *Proteomics*, 5(18):4705–4712.

Gunawardana, C. G., Memari, N., and Diamandis, E. P. (2009). Identifying novel autoantibody signatures in ovarian cancer using high-density protein microarrays. *Clinical Biochemistry*, 42(4-5):426–429.

Gutman, S. and Kessler, L. G. (2006). The US Food and Drug Administration perspective on cancer biomarker development. *Nature Reviews Cancer*, 6(7):565–571.

Guyton, A. C. and Hall, J. E. (2006). *Textbook of Medical Physiology*. Elsevier, 11th edition.

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, 17(10):994–999.

Gyorgy, A. B., Walker, J., Wingo, D., Eidelman, O., Pollard, H. B., Molnar, A., and Agoston, D. V. (2010). Reverse phase protein microarray technology in traumatic brain injury. *Journal of neuroscience methods*, 192(1):96–101.

Haab, B. B. (2005). Antibody arrays in cancer research. *Molecular & Cellular Proteomics: MCP*, 4(4):377–383.

Haab, B. B. (2006). Applications of antibody array platforms. *Current Opinion In Biotechnology*, 17(4):415–421.

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 Biology*, 2(2):RESEARCH0004.

Haab, B. B., Geierstanger, B. H., Michailidis, G., Vitzthum, F., Forrester, S., Okon, R., Saviranta, P., Brinker, A., Sorette, M., Perlee, L., Suresh, S., Drwal, G., Adkins, J. N., and Omenn, G. S. (2005). Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: Systematic variation between sample types and calibration of mass spectrometry data. *Proteomics*, 5(13):3278–3291.

Haab, B. B., Paulovich, A. G., Anderson, N. L., Clark, A. M., Downing, G. J., Hermjakob, H., LaBaer, J., and Uhlen, M. (2006). A Reagent Resource to Identify Proteins and Peptides of Interest for the Cancer Community: A WORKSHOP REPORT. *Molecular & Cellular Proteomics: MCP*, 5(10):1996–2007.

Hagedorn, P. H., Burton, C. M., Carlsen, J., Steinbruchel, D., Andersen, C. B., Sahar, E., Domany, E., Cohen, I. R., Flyvbjerg, H., and Iversen, M. (2010). Chronic rejection of a lung transplant is characterized by a profile of specific autoantibodies. *Immunology*, 130(3):427–435.

Häggmark, A., Byström, S., Ayoglu, B., Qundos, U., Uhlén, M., Khademi, M., Olsson, T., Schwenk, J. M., and Nilsson, P. (2013). Antibody-based profiling of cerebrospinal fluid within multiple sclerosis. *Proteomics*, 13(15):2256–2267.

Hall, D. A., Ptacek, J., and Snyder, M. (2007). Protein microarray technology. *Mechanisms of ageing and development*, 128(1):161–167.

Hamelinck, D., Zhou, H., Li, L., Verweij, C., Dillon, D., Feng, Z., Costa, J., and Haab, B. B. (2005). Optimized normalization for antibody microarrays and application to serum-protein profiling. *Molecular & Cellular Proteomics: MCP*, 4(6):773–784.

Han, M., Nagele, E., Demarshall, C., Acharya, N., and Nagele, R. (2012). Diagnosis of Parkinson's disease based on disease-specific autoantibody profiles in human sera. *PloS ONE*, 7(2):e32383.

Hanash, S. (2003). Disease proteomics. *Nature*, 422(6928):226–232.

Hanes, J. and Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences of the United States of America*, 94(10):4937–4942.

Hanly, W. C., Artwohl, J. E., and Bennett, B. T. (1995). Review of Polyclonal Antibody Production Procedures in Mammals and Poultry. *ILAR journal / National Research Council, Institute of Laboratory Animal Resources*, 37(3):93–118.

Harrison, P. M., Kumar, A., Lang, N., Snyder, M., and Gerstein, M. (2002). A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Research*, 30(5):1083–1090.

Hartley, H. (1951). Origin of the word 'protein'. *Nature*, 168(4267):244.

Hartmann, M., Roeraade, J., Stoll, D., Templin, M. F., and Joos, T. O. (2009). Protein microarrays for diagnostic assays. *Analytical and bioanalytical chemistry*, 393(5):1407–1416.

Hartmann, M., Schrenk, M., Döttinger, A., Nagel, S., Roeraade, J., Joos, T. O., and Templin, M. F. (2008). Expanding assay dynamics: a combined competitive and direct assay system for the quantification of proteins in multiplexed immunoassays. *Clinical Chemistry*, 54(6):956–963.

Harwanegg, C. and Hiller, R. (2005). Protein microarrays for the diagnosis of allergic diseases: state-of-the-art and future development. *Clinical Chemistry and Laboratory Medicine*, 43(12):1321–1326.

Hayter, S. M. and Cook, M. C. (2012). Updated assessment of the prevalence, spectrum and case definition of autoimmune disease. *Autoimmunity Reviews*, 11(10):754–765.

He, M., He, Y., Luo, Q., and Wang, M. (2011). From DNA to protein: No living cells required. *Process Biochemistry*, 46(3):615–620.

He, M., Stoevesandt, O., Palmer, E. A., Khan, F., Ericsson, O., and Taussig, M. J. (2008a). Printing protein arrays from DNA arrays. *Nature Methods*, 5(2):175–177.

He, M., Stoevesandt, O., and Taussig, M. J. (2008b). In situ synthesis of protein arrays. *Current Opinion In Biotechnology*, 19(1):4–9.

He, M. and Taussig, M. J. (2001). Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). *Nucleic Acids Research*, 29(15):E73–3.

He, M. and Taussig, M. J. (2003). DiscernArray technology: a cell-free method for the generation of protein arrays from PCR DNA. *Journal of Immunological Methods*, 274(1-2):265–270.

Hecker, M., Lorenz, P., Steinbeck, F., Hong, L., Riemekasten, G., Li, Y., Zettl, U. K., and Thiesen, H.-J. (2012). Computational analysis of high-density peptide microarray data with application from systemic sclerosis to multiple sclerosis. *Autoimmunity Reviews*, 11(3):180–190.

Heid, C. A., Stevens, J., Livak, K. J., and Williams, P. M. (1996). Real time quantitative PCR. *Genome Research*, 6(10):986–994.

Hemmer, B., Stüve, O., Kieseier, B., Schellekens, H., and Hartung, H.-P. (2005). Immune response to immunotherapy: the role of neutralising antibodies to interferon beta in the treatment of multiple sclerosis. *Lancet Neurology*, 4(7):403–412.

Henderson, G. and Bradley, M. (2007). Functional peptide arrays for high-throughput chemical biology based applications. *Current Opinion In Biotechnology*, 18(4):326–330.

Hjelm, B., Fernández, C. D., Löfblom, J., Ståhl, S., Johannesson, H., Rockberg, J., and Uhlén, M. (2010). Exploring epitopes of antibodies toward the human tryptophanyl-tRNA synthetase. *New Biotechnology*, 27(2):129–137.

Hjelm, B., Forsström, B., Igel, U., Johannesson, H., Stadler, C., Lundberg, E., Ponten, F., Sjöberg, A., Rockberg, J., Schwenk, J. M., Nilsson, P., Johansson, C., and Uhlén, M. (2011). Generation of monospecific antibodies based on affinity capture of polyclonal antibodies. *Protein Science*, 20(11):1824–1835.

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. *Molecular & Cellular Proteomics: MCP*, 10(2):M110.002857.

Hoffman, E. P., Brown, R. H., and Kunkel, L. M. (1987). Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell*, 51(6):919–928.

Hoheisel, J. D., Alhamdani, M. S. S., and Schröder, C. (2013). Affinity-based microarrays for proteomic analysis of cancer tissues. *Proteomics - Clinical Applications*, 7(1-2):8–15.

Holliger, P. and Hudson, P. J. (2005). Engineered antibody fragments and the rise of single domains. *Nature Biotechnology*, 23(9):1126–1136.

Hoogenboom, H. R. (2005). Selecting and screening recombinant antibody libraries. *Nature Biotechnology*, 23(9):1105–1116.

Horan, P. and Wheeless, L. (1977). Quantitative single cell analysis and sorting. *Science*, 198(4313):149–157.

Horn, S., Lueking, A., Murphy, D., Staudt, A., Gutjahr, C., Schulte, K., König, A., Landsberger, M., Lehrach, H., Felix, S. B., and Cahill, D. J. (2006). Profiling humoral autoimmune repertoire of dilated cardiomyopathy (DCM) patients and development of a disease-associated protein chip. *Proteomics*, 6(2):605–613.

Hsieh, S.-Y., Chen, R.-K., Pan, Y.-H., and Lee, H.-L. (2006). Systematical evaluation of the effects of sample collection procedures on low-molecular-weight serum/plasma proteome profiling. *Proteomics*, 6(10):3189–3198.

Hsu, H.-Y., Joos, T. O., and Koga, H. (2009). Multiplex microsphere-based flow cytometric platforms for protein analysis and their application in clinical proteomics - from assays to results. *Electrophoresis*, 30(23):4008–4019.

Hsu, H.-Y., Wittemann, S., Schneider, E. M., Weiss, M., and Joos, T. O. (2008). Suspension microarrays for the identification of the response patterns in hyperinflammatory diseases. *Medical engineering & physics*, 30(8):976–983.

Hu, S., Li, Y., Liu, G., Song, Q., Wang, L., Han, Y., Zhang, Y., Song, Y., Yao, X., Tao, Y., Zeng, H., Yang, H., Wang, J., Zhu, H., Chen, Z.-N., and Wu, L. (2007). A protein chip approach for high-throughput antigen identification and characterization. *Proteomics*, 7(13):2151–2161.

Hu, S., Loo, J. A., and Wong, D. T. (2006). Human body fluid proteome analysis. *Proteomics*, 6(23):6326–6353.

Hu, S., Vissink, A., Arellano, M., Roozendaal, C., Zhou, H., Kallenberg, C. G. M., and Wong, D. T. (2011). Identification of autoantibody biomarkers for primary Sjögren's syndrome using protein microarrays. *Proteomics*, 11(8):1499–1507.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1):44–57.

Huang, R. P. (2001). Detection of multiple proteins in an antibody-based protein microarray system. *Journal of Immunological Methods*, 255(1-2):1–13.

Hudelist, G., Pacher-zavisin, M., Singer, C., Holper, T., Kubista, E., Schreiber, M., Manavi, M., Bilban, M., and Czerwenka, K. (2004). Use of High-Throughput Protein Array for Profiling of Differentially Expressed Proteins in Normal and Malignant Breast Tissue. *Breast Cancer Research and Treatment*, 86(3):283–293.

Hudson, M. E., Pozdnyakova, I., Haines, K., Mor, G., and Snyder, M. (2007). Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. *Proceedings of the National Academy of Sciences of the United States of America*, 104(44):17494–17499.

Hueber, W., Kidd, B. A., Tomooka, B. H., Lee, B. J., Bruce, B., Fries, J. F., Sønderstrup, G., Monach, P., Drijfhout, J. W., van Venrooij, W. J., Utz, P. J., Genovese, M. C., and Robinson, W. H. (2005). Antigen microarray profiling of autoantibodies in rheumatoid arthritis. *Arthritis and Rheumatism*, 52(9):2645–2655.

Hueber, W., Utz, P. J., Steinman, L., and Robinson, W. H. (2002). Autoantibody profiling for the study and treatment of autoimmune disease. *Arthritis Research*, 4(5):290–295.

Human Genome Sequencing Consortium, I. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945.

Hunt, D. F., Buko, A. M., Ballard, J. M., Shabanowitz, J., and Giordani, A. B. (1981). Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. *Biomedical mass spectrometry*, 8(9):397–408.

HUPO-The Human Proteome Organization (2010). A gene-centric human proteome project. *Molecular & Cellular Proteomics: MCP*, 9(2):427–429.

Hurrell, J. G., Smith, J. A., Todd, P. E., and Leach, S. J. (1977). Cross-reactivity between mammalian myoglobins: linear vs spatial antigenic determinants. *Immunochemistry*, 14(4):283–288.

Hye, A., Riddoch-Contreras, J., Baird, A. L., Ashton, N. J., Bazenet, C., Leung, R., Westman, E., Simmons, A., Dobson, R., Sattlecker, M., Lupton, M., Lunnon, K., Keohane, A., Ward, M., Pike, I., Zucht, H. D., Pepin, D., Zheng, W., Tunnicliffe, A., Richardson, J., Gauthier, S., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., and Lovestone, S. (2014). Plasma proteins predict conversion to dementia from prodromal disease. *Alzheimer's & Dementia*.

Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T., and Yamashina, I. (1987). Serum lectin with known structure activates complement through the classical pathway. *Journal of Biological Chemistry*, 262(16):7451–7454.

Ingvarsson, J., Lindstedt, M., Borrebaeck, C. A. K., and Wingren, C. (2006). One-Step Fractionation of Complex Proteomes Enables Detection of Low Abundant Analytes Using Antibody-Based Microarrays. *Journal of Proteome Research*, 5(1):170–176.

Ingvarsson, J., Wingren, C., Carlsson, A., Ellmark, P., Wahren, B., Engström, G., Harmenberg, U., Krogh, M., Peterson, C., and Borrebaeck, C. A. K. (2008). Detection of pancreatic cancer using antibody microarray-based serum protein profiling. *Proteomics*, 8(11):2211–2219.

Ioannidis, J. P. A. (2013). Biomarker failures. *Clinical Chemistry*, 59(1):202–204.

Jackson, D. H. and Banks, R. E. (2010). Banking of clinical samples for proteomic biomarker studies: A consideration of logistical issues with a focus on pre-analytical variation. *Proteomics - Clinical Applications*, 4(3):250–270.

Jacobson, J. W., Oliver, K. G., Weiss, C., and Kettman, J. (2006). Analysis of individual data from bead-based assays ("bead arrays"). *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, 69(5):384–390.

Jambunathan, K. and Galande, A. K. (2014). Sample collection in clinical proteomics-Proteolytic activity profile of serum and plasma. *Proteomics - Clinical Applications*, 8(5-6):299–307.

Janeway, C. A. J., Travers, P., Walport, M., and Shlomchik, M. J. (2001). *Immunobiology: The Immune System in Health and Disease*. New York: Garland Science, 5th edition.

Janzi, M., Kull, I., Sjöberg, R., Wan, J., Melén, E., Bayat, N., Ostblom, E., Pan-Hammarström, Q., Nilsson, P., and Hammarström, L. (2009a). Selective IgA deficiency in early life: association to infections and allergic diseases during childhood. *Clinical Immunology*, 133(1):78–85.

Janzi, M., Odling, J., Pan-Hammarström, Q., Sundberg, M., Lundeberg, J., Uhlén, M., Hammarström, L., and Nilsson, P. (2005). Serum microarrays for large scale screening of protein levels. *Molecular & Cellular Proteomics: MCP*, 4(12):1942–1947.

Janzi, M., Sjöberg, R., Wan, J., Fischler, B., von Döbeln, U., Isaac, L., Nilsson, P., and Hammarström, L. (2009b). Screening for C3 deficiency in newborns using microarrays. *PloS ONE*, 4(4):e5321.

Jarius, S. and Wildemann, B. (2011). Effect of storage conditions and freeze/thaw cycles on aquaporin-4 antibody (NMO-IgG) serum levels. *Clinical Chemistry and Laboratory Medicine*, 49(12):2121–2122.

Jennette, J. C. and Falk, R. J. (2010). The rise and fall of horror autotoxicus and forbidden clones. *Kidney International*, 78(6):533–535.

Jeong, J. S., Jiang, L., Albino, E., Marrero, J., Rho, H. S., Hu, J., Hu, S., Vera, C., Bayron-Poueymiroy, D., Rivera-Pacheco, Z. A., Ramos, L., Torres-Castro, C., Qian, J., Bonaventura, J., Boeke, J. D., Yap, W. Y., Pino, I., Eichinger, D. J., Zhu, H., and Blackshaw, S. (2012). Rapid identification of monospecific monoclonal antibodies using a human proteome microarray. *Molecular & Cellular Proteomics: MCP*, 11(6):O111.016253.

Jones, H. B. (1848). On a new substance occurring in the urine of a patient with mollities ossium. *Philosophical Transactions of the Royal Society of London*, 138:55–62.

Jónsson, T. and Valdimarsson, H. (1993). Is measurement of rheumatoid factor isotypes clinically useful? *Annals of the Rheumatic Diseases*, 52(2):161–164.

Joos, T. O., Schrenk, M., Höpfl, P., Kröger, K., Chowdhury, U., Stoll, D., Schörner, D., Dürr, M., Herick, K., Rupp, S., Sohn, K., and Hämmerle, H. (2000). A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis*, 21(13):2641–2650.

Juncker, D., Bergeron, S., Laforte, V., and Li, H. (2014). Cross-reactivity in antibody microarrays and multiplexed sandwich assays: shedding light on the dark side of multiplexing. *Current Opinion in Chemical Biology*, 18:29–37.

Kader, H. A., Tchernev, V. T., Satyaraj, E., Lejnine, S., Kotler, G., Kingsmore, S. F., and Patel, D. D. (2005). Protein microarray analysis of disease activity in pediatric inflammatory bowel disease demonstrates elevated serum PLGF, IL-7, TGF-beta1, and IL-12p40 levels in Crohn's disease and ulcerative colitis patients in remission versus active disease. *The American Journal of Gastroenterology*, 100(2):414–423.

Kampf, C., Andersson, A.-C., Wester, K., Björling, E., Uhlén, M., and Ponten, F. (2005). Antibody-Based Tissue Profiling As a Tool for Clinical Proteomics. *Clinical Proteomics*, 1(3-4):285–300.

Kanter, J. L., Narayana, S., Ho, P. P., Catz, I., Warren, K. G., Sobel, R. A., Steinman, L., and Robinson, W. H. (2006). Lipid microarrays identify key mediators of autoimmune brain inflammation. *Nature Medicine*, 12(1):138–143.

Kaplan, I. V. and Levinson, S. S. (1999). When is a heterophile antibody not a heterophile antibody? When it is an antibody against a specific immunogen. *Clinical Chemistry*, 45(5):616–618.

Karas, M. and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, 60(20):2299–2301.

Karu, A. E., Bell, C. W., and Chin, T. E. (1995). Recombinant Antibody Technology. *ILAR journal / National Research Council, Institute of Laboratory Animal Resources*, 37(3):132–141.

Karush, F. (1978). The affinity of antibody: range, variability, and the role of multivalence. In *Immunoglobulins*, pages 85–116. Plenum Publishing Corporation.

Karussis, D. (2014). The diagnosis of multiple sclerosis and the various related demyelinating syndromes: A critical review. *Journal of Autoimmunity*, 48-49C:134–142.

Kato, B. S., Nicholson, G., Neiman, M., Rantalainen, M., Holmes, C. C., Barrett, A., Uhlén, M., Nilsson, P., Spector, T. D., and Schwenk, J. M. (2011). Variance decomposition of protein profiles from antibody arrays using a longitudinal twin model. *Proteome Science*, 9:73.

Kattah, M. G., Alemi, G. R., Thibault, D. L., Balboni, I., and Utz, P. J. (2006). A new two-color Fab labeling method for autoantigen protein microarrays. *Nature Methods*, 3(9):745–751.

Katz, C., Levy-Beladev, L., Rotem-Bamberger, S., Rito, T., Rüdiger, S. G. D., and Friedler, A. (2011). Studying protein-protein interactions using peptide arrays. *Chemical Society reviews*, 40(5):2131–2145.

Kaufmann, S. H. E. (2008). Immunology's foundation: the 100-year anniversary of the Nobel Prize to Paul Ehrlich and Elie Metchnikoff. *Nature Immunology*, 9(7):705–712.

Kaukola, T., Satyaraj, E., Patel, D. D., Tchernev, V. T., Grimwade, B. G., Kingsmore, S. F., Koskela, P., Tammela, O., Vainionp, L., Pihko, H., rimaa, T., and Hallman, M. (2004). Cerebral palsy is characterized by protein mediators in cord serum. *Annals of Neurology*, 55(2):186–194.

Kellar, K. (2002). Multiplexed microsphere-based flow cytometric assays. *Experimental Hematology*, 30(11):1227–1237.

Kellar, K. L. and Oliver, K. G. (2004). Multiplexed microsphere assays for protein and DNA binding reactions. *Methods in Cell Biology*, 75:409–429.

Kern, S. E. (2012). Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. *Cancer Research*, 72(23):6097–6101.

Kettman, J. R., Davies, T., Chandler, D., Oliver, K. G., and Fulton, R. J. (1998). Classification and properties of 64 multiplexed microsphere sets. *Cytometry*, 33(2):234–243.

Kim, B. K., Lee, J. W., Park, P. J., Shin, Y. S., Lee, W. Y., Lee, K. A., Ye, S., Hyun, H., Kang, K. N., Yeo, D., Kim, Y., Ohn, S. Y., Noh, D. Y., and Kim, C. W. (2009). The multiplex bead array approach to identifying serum biomarkers associated with breast cancer. *Breast Cancer Research: BCR*, 11(2):R22.

Kim, D. D. and Song, W.-C. (2006). Membrane complement regulatory proteins. *Clinical Immunology*, 118(2-3):127–136.

Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabuddhe, N. A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L. D. N., Patil, A. H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S. K., Marimuthu, A., Sathe, G. J., Chavan, S., Datta, K. K., Subbannayya, Y., Sahu, A., Yelamanchi, S. D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K. R., Syed, N., Goel, R., Khan, A. A., Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T.-C., Zhong, J., Wu, X., Shaw, P. G., Freed, D., Zahari, M. S., Mukherjee, K. K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C. J., Shankar, S. K., Satishchandra, P., Schroeder, J. T., Sirdeshmukh, R., Maitra, A., Leach, S. D., Drake, C. G., Halushka, M. K., Prasad, T. S. K., Hruban, R. H., Kerr, C. L., Bader, G. D., Iacobuzio-Donahue, C. A., Gowda, H., and Pandey, A. (2014). A draft map of the human proteome. *Nature*, 509(7502):575–581.

Kindt, T. J., Goldsby, R. A., Osborne, B. A., and Kuby, J. (2007). *Kuby Immunology*. W. H. Freeman, New York, 6th edition.

Kingsmore, S. F. and Patel, D. D. (2003). Multiplexed protein profiling on antibody-based microarrays by rolling circle amplification. *Current Opinion In Biotechnology*, 14(1):74–81.

Kivity, S., Agmon-Levin, N., Blank, M., and Shoenfeld, Y. (2009). Infections and autoimmunity – friends or foes? *Trends in Immunology*, 30(8):409–414.

Klein, J. and Sato, A. (2000). The HLA system. Second of two parts. *The New England Journal of Medicine*, 343(20):782–786.

Klein, P., Glaser, E., Grogan, L., Keane, M., Lipkowitz, S., Soballe, P., Brooks, L., Jenkins, J., Steinberg, S. M., DeMarini, D. M., and Kirsch, I. (2001). Biomarker assays in nipple aspirate fluid. *Breast Journal*, 7(6):378–387.

Klinman, D. M. and Steinberg, A. D. (1987). Systemic autoimmune disease arises from polyclonal B cell activation. *The Journal of Experimental Medicine*, 165(6):1755–1760.

Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, 26(3):231–243.

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., and Krizman, D. B. (2001). Proteomic profiling of the cancer microenvironment by antibody arrays. *Proteomics*, 1(10):1271–1278.

Kodadek, T. (2001). Protein microarrays: prospects and problems. *Chemistry & biology*, 8(2):105–115.

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

Koide, A., Bailey, C. W., Huang, X., and Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins. *Journal of Molecular Biology*, 284(4):1141–1151.

Koo, B. K., Chae, S., Kim, K. M., Kang, M. J., Kim, E. G., Kwak, S. H., Jung, H. S., Cho, Y. M., Choi, S. H., Park, Y. J., Shin, C. H., Jang, H. C., Shin, C. S., Hwang, D., Yi, E. C., and Park, K. S. (2014). Identification of Novel Autoantibodies in Type 1 Diabetic Patients Using a High-Density Protein Microarray. *Diabetes*.

Korf, U., Löbke, C., Sahin, Ö., Haller, F., Sültmann, H., Arlt, D., and Poustka, A. (2009). Reverse-phase protein arrays for application-orientated cancer research. *Proteomics - Clinical Applications*, 3(10):1140–1150.

Korngold, L. and Lipari, R. (1956). Multiple-myeloma proteins.III. The antigenic relationship of Bence Jones proteins to normal gamma-globulin and multiple-myeloma serum proteins. *Cancer*, 9(2):262–272.

Kraemer, S., Vaught, J. D., Bock, C., Gold, L., Katilius, E., Keeney, T. R., Kim, N., Saccomano, N. A., Wilcox, S. K., Zichi, D., and Sanders, G. M. (2011). From SOMAmer-based biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. *PloS ONE*, 6(10):e26332.

Kricka, L. J. (1999). Human anti-animal antibody interferences in immunological assays. *Clinical Chemistry*, 45(7):942–956.

Kulasingam, V. and Diamandis, E. P. (2008). Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature clinical practice Oncology*, 5(10):588–599.

Kuriyama, M., Wang, M. C., Papsidero, L. D., Killian, C. S., Shimano, T., Valenzuela, L., Nishiura, T., Murphy, G. P., and Chu, T. M. (1980). Quantitation of prostate-specific antigen in serum by a sensitive enzyme immunoassay. *Cancer Research*, 40(12):4658–4662.

Kusnezow, W., Banzon, V., Schröder, C., Schaal, R., Hoheisel, J. D., Rüffer, S., Luft, P., Duschl, A., and Syagailo, Y. V. (2007). Antibody microarray-based profiling of complex specimens: systematic evaluation of labeling strategies. *Proteomics*, 7(11):1786–1799.

Kusnezow, W. and Hoheisel, J. D. (2003). Solid supports for microarray immunoassays. *Journal of Molecular Recognition: JMR*, 16(4):165–176.

LaBaer, J. (2012). Improving international research with clinical specimens: 5 achievable objectives. *Journal of Proteome Research*, 11(12):5592–5601.

Lakkis, F. G. (2012). *Immunotherapy in Transplantation: Principles and Practice*. Transplantation: Principles and Practice, 1st edition.

Landegren, U., Vänelid, J., Hammond, M., Nong, R. Y., Wu, D., Ullerås, E., and Kamali-Moghaddam, M. (2012). Opportunities for sensitive plasma proteome analysis. *Analytical Chemistry*, 84(4):1824–1830.

Larsson, K., Eriksson, C., Schwenk, J. M., Berglund, L., Wester, K., Uhlen, M., Hober, S., and Wernerus, H. (2009). Characterization of PrEST-based antibodies towards human Cytokeratin-17. *Journal of Immunological Methods*, 342(1-2):20–32.

Lederberg, J. (1959). Genes and Antibodies: Do antigens bear instructions for antibody specificity or do they select cell lines that arise by mutation? *Science*, 129(3364):1649–1653.

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., and Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11(10):733–739.

Leenaars, M. and Hendriksen, C. F. M. (2005). Critical steps in the production of polyclonal and monoclonal antibodies: evaluation and recommendations. *ILAR journal / National Research Council, Institute of Laboratory Animal Resources*, 46(3):269–279.

Leff, S. E., Rosenfeld, M. G., and Evans, R. M. (1986). Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Annual Review of Biochemistry*, 55:1091–1117.

Legrain, P., Aebersold, R., Archakov, A., Bairoch, A., Bala, K., Beretta, L., Bergeron, J., Borchers, C. H., Corthals, G. L., Costello, C. E., Deutsch, E. W., Domon, B., Hancock, W., He, F., Hochstrasser, D., Marko-Varga, G., Salekdeh, G. H., Sechi, S., Snyder, M., Srivastava, S., Uhlén, M., Wu, C. H., Yamamoto, T., Paik, Y.-K., and Omenn, G. S. (2011). The human proteome project: current state and future direction. *Molecular & Cellular Proteomics: MCP*, 10(7):M111.009993–M111.009993.

Lequin, R. M. (2005). Enzyme Immunoassay (EIA)/Enzyme-Linked Immunosorbent Assay (ELISA). *Clinical Chemistry*, 51(12):2415–2418.

Lernmark, Å. (2001). Autoimmune diseases: are markers ready for prediction? *The Journal of Clinical Investigation*, 108(8):1091–1096.

Levin, Y. (2011). The role of statistical power analysis in quantitative proteomics. *Proteomics*, 11(12):2565–2567.

Li, H., Bergeron, S., and Juncker, D. (2012). Microarray-to-microarray transfer of reagents by snapping of two chips for cross-reactivity-free multiplex immunoassays. *Analytical Chemistry*, 84(11):4776–4783.

Li, L., Wadia, P., Chen, R., Kambham, N., Naesens, M., Sigdel, T. K., Miklos, D. B., Sarwal, M. M., and Butte, A. J. (2009). Identifying compartment-specific non-HLA targets after renal transplantation by integrating transcriptome and "antibodyome" measures. *Proceedings of the National Academy of Sciences of the United States of America*, 106(11):4148–4153.

Li, Q.-Z., Zhen, Q. L., Xie, C., Wu, T., Mackay, M., Aranow, C., Putterman, C., and Mohan, C. (2005). Identification of autoantibody clusters that best predict lupus disease activity using glomerular proteome arrays. *The Journal of Clinical Investigation*, 115(12):3428–3439.

Linderstrøm-Lang, K. U. (1952). Lane Medical Lectures: Proteins and Enzymes. *Stanford Univ Publ Med Series*, 6:1–115.

Lindskog, M., Rockberg, J., Uhlén, M., and Sterky, F. (2005). Selection of protein epitopes for antibody production. *BioTechniques*, 38(5):723–727.

Lipman, N. S., Jackson, L. R., Trudel, L. J., and Weis-Garcia, F. (2005). Monoclonal versus polyclonal antibodies: distinguishing characteristics, applications, and information resources. *ILAR journal / National Research Council, Institute of Laboratory Animal Resources*, 46(3):258–268.

Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nature Genetics*, 21(1 Suppl):20–24.

Liu, N., Liang, W., Ma, X., Li, X., Ning, B., Cheng, C., Ou, G., Wang, B., Zhang, J., and Gao, Z. (2013). Simultaneous and combined detection of multiple tumor biomarkers for prostate cancer in human serum by suspension array technology. *Biosensors & bioelectronics*, 47:92–98.

Liumbruno, G., D'Alessandro, A., Grazzini, G., and Zolla, L. (2010). Blood-related proteomics. *Journal of Proteomics*, 73(3):483–507.

Löfblom, J., Frejd, F. Y., and Ståhl, S. (2011). Non-immunoglobulin based protein scaffolds. *Current Opinion In Biotechnology*, 22(6):843–848.

Lollo, B., Steele, F., and Gold, L. (2014). Beyond antibodies: new affinity reagents to unlock the proteome. *Proteomics*, 14(6):638–644.

Luckert, K., Gujral, T. S., Chan, M., Sevecka, M., Joos, T. O., Sorger, P. K., MacBeath, G., and Pötz, O. (2012). A dual array-based approach to assess the abundance and posttranslational modification state of signaling proteins. *Science signaling*, 5(206):pl1.

Lueking, A., Horn, M., Eickhoff, H., Büssow, K., Lehrach, H., and Walter, G. (1999). Protein Microarrays for Gene Expression and Antibody Screening. *Analytical Biochemistry*, 270(1):103–111.

Lueking, A., Huber, O., Wirths, C., Schulte, K., Stieler, K. M., Blume-Peytavi, U., Kowald, A., Hensel-Wiegel, K., Tauber, R., Lehrach, H., Meyer, H. E., and Cahill, D. J. (2005). Profiling of alopecia areata autoantigens based on protein microarray technology. *Molecular & Cellular Proteomics: MCP*, 4(9):1382–1390.

Lueking, A., Possling, A., Huber, O., Beveridge, A., Horn, M., Eickhoff, H., Schuchardt, J., Lehrach, H., and Cahill, D. J. (2003). A nonredundant human protein chip for antibody screening and serum profiling. *Molecular & Cellular Proteomics: MCP*, 2(12):1342–1349.

Lundby, A. and Olsen, J. V. (2011). GeLC-MS for In-Depth Protein Characterization and Advanced Analysis of Proteomes. In *Gel-Free Proteomics*, pages 143–155. Humana Press, Totowa, NJ.

MacBeath, G. and Schreiber, S. L. (2000). Printing proteins as microarrays for high-throughput function determination. *Science*, 289(5485):1760–1763.

Maecker, H. T., Lindstrom, T. M., Robinson, W. H., Utz, P. J., Hale, M., Boyd, S. D., Shen-Orr, S. S., and Fathman, C. G. (2012). New tools for classification and monitoring of autoimmune diseases. *Nature Reviews Rheumatology*, 8(6):317–328.

Magi, B., Bargagli, E., Bini, L., and Rottoli, P. (2006). Proteome analysis of bronchoalveolar lavage in lung diseases. *Proteomics*, 6(23):6354–6369.

Mak, T. W., Saunders, M. E., and Jett, B. D. (2013). *Primer to The Immune Response*. AP Cell - Elsevier, 1 edition.

Malou, N. and Raoult, D. (2011). Immuno-PCR: a promising ultrasensitive diagnostic method to detect antigens and antibodies. *Trends in Microbiology*, 19(6):295–302.

Mann, M. and Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology*, 21(3):255–261.

Mann, M., Kulak, N. A., Nagaraj, N., and Cox, J. (2013). The coming age of complete, accurate, and ubiquitous proteomes. *Molecular cell*, 49(4):583–590.

Mannello, F. (2008). Serum or Plasma Samples?: The "Cinderella" Role of Blood Collection Procedures Preanalytical Methodological Issues Influence the Release and Activity of Circulating Matrix Metalloproteinases and Their Tissue Inhibitors, Hampering Diagnostic Trueness and Leading to Misinterpretation. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 28(4):611–614.

Manzur, A. Y. and Muntoni, F. (2009). Diagnosis and new treatments in muscular dystrophies. *Journal of Neurology, Neurosurgery, and Psychiatry*, 80(7):706–714.

Marko-Varga, G., Végvári, A., Welinder, C., Lindberg, H., Rezeli, M., Edula, G., Svensson, K. J., Belting, M., Laurell, T., and Fehniger, T. E. (2012). Standardization and utilization of biobank resources in clinical protein science with examples of emerging applications. *Journal of Proteome Research*, 11(11):5124–5134.

Marrack, P., Kappler, J., and Kotzin, B. L. (2001). Autoimmune disease: why and where it occurs. *Nature Medicine*, 7(8):899–905.

Martins, T. B., Burlingame, R., von Mühlen, C. A., Jaskowski, T. D., Litwin, C. M., and Hill, H. R. (2004). Evaluation of multiplexed fluorescent microsphere immunoassay for detection of autoantibodies to nuclear antigens. *Clinical and diagnostic laboratory immunology*, 11(6):1054–1059.

Martins, T. B., Litwin, C. M., and Hill, H. R. (2008). Evaluation of a multiplex fluorescent microsphere immunoassay for the determination of epstein-barr virus serologic status. *American Journal of Clinical Pathology*, 129(1):34–41.

Marx, V. (2013). Finding the right antibody for the job. *Nature Methods*, 10(8):703–707.

Masson-Bessière, C., Sebbag, M., Girbal-Neuhauser, E., Nogueira, L., Vincent, C., Senshu, T., and Serre, G. (2001). The major synovial targets of the rheumatoid arthritis-specific antifilaggrin autoantibodies are deiminated forms of the alpha- and beta-chains of fibrin. *Journal of Immunology*, 166(6):4177–4184.

May, C., Nordhoff, E., Casjens, S., Turewicz, M., Eisenacher, M., Gold, R., Brüning, T., Pesch, B., Stephan, C., Woitalla, D., Penke, B., Janáky, T., Virók, D., Siklós, L., Engelhardt, J. I., and Meyer, H. E. (2014). Highly Immunoreactive IgG Antibodies Directed against a Set of Twenty Human Proteins in the Sera of Patients with Amyotrophic Lateral Sclerosis Identified by Protein Array. *PloS ONE*, 9(2):e89596.

McCafferty, J., Griffiths, A. D., Winter, G., and Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*, 348(6301):552–554.

McDermed, J. E., Sanders, R., Fait, S., Klem, R. E., Sarno, M. J., Adams, T. H., and Diamandis, E. P. (2012). Nucleic Acid Detection Immunoassay for Prostate-Specific Antigen Based on Immuno-PCR Methodology. *Clinical Chemistry*, 58(4):732–740.

McFarland, H. F. and Martin, R. (2007). Multiple sclerosis: a complicated picture of autoimmunity. *Nature Immunology*, 8(9):913–919.

McHugh, T. M., Miner, R. C., Logan, L. H., and Stites, D. P. (1988). Simultaneous detection of antibodies to cytomegalovirus and herpes simplex virus by using flow cytometry and a microsphere-based fluorescence immunoassay. *Journal of clinical microbiology*, 26(10):1957–1961.

McRae, B. L. (1995). Functional evidence for epitope spreading in the relapsing pathology of experimental autoimmune encephalomyelitis. *Journal of Experimental Medicine*, 182(1):75–85.

Mehan, M. R., Ayers, D., Thirstrup, D., Xiong, W., Ostroff, R. M., Brody, E. N., Walker, J. J., Gold, L., Jarvis, T. C., Janjic, N., Baird, G. S., and Wilcox, S. K. (2012). Protein signature of lung cancer tissues. *PloS ONE*, 7(4):e35157.

Merrifield, R. B. (1963). Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *Journal of the American Chemical Society*, 85(14):2149–2154.

Merrifield, R. B., Stewart, J. M., and Jernberg, N. (1966). Instrument for automated synthesis of peptides. *Analytical Chemistry*, 38(13):1905–1914.

Mias, G. I., Chen, R., Zhang, Y., Sridhar, K., Sharon, D., Xiao, L., Im, H., Snyder, M. P., and Greenberg, P. L. (2013). Specific plasma autoantibody reactivity in myelodysplastic syndromes. *Scientific reports*, 3:3311.

Miersch, S., Bian, X., Wallstrom, G., Sibani, S., Logvinenko, T., Wasserfall, C. H., Schatz, D., Atkinson, M., Qiu, J., and LaBaer, J. (2013). Serological autoantibody profiling of type 1 diabetes by protein arrays. *Journal of Proteomics*, 94:486–496.

Miles, L. E. M. and Hales, C. N. (1968). Labelled Antibodies and Immunological Assay Systems. *Nature*, 219(5150):186–189.

Miller, J. C., Zhou, H., Kwekel, J., Cavallo, R., Burke, J., Butler, E. B., Teh, B. S., and Haab, B. B. (2003). Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. *Proteomics*, 3(1):56–63.

Milo, R. and Miller, A. (2014). Autoimmunity Reviews. *Autoimmunity Reviews*, 13(4-5):518–524.

Milstein, C. (1981). The Wellcome Foundation Lecture, 1980: monoclonal antibodies from hybrid myelomas. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, 211(1185):393–412.

Mironov, A. A., Fickett, J. W., and Gelfand, M. S. (1999). Frequent alternative splicing of human genes. *Genome Research*, 9(12):1288–1293.

Mischak, H., Apweiler, R., Banks, R. E., Conaway, M., Coon, J., Dominiczak, A., Ehrich, J. H. H., Fliser, D., Girolami, M., Hermjakob, H., Hochstrasser, D., Jankowski, J., Julian, B. A., Kolch, W., Massy, Z. A., Neususess, C., Novak, J., Peter, K., Rossing, K., Schanstra, J., Semmes, O. J., Theodorescu, D., Thongboonkerd, V., Weissinger, E. M., Van Eyk, J. E., and Yamamoto, T. (2007). Clinical proteomics: A need to define the field and to begin to set adequate standards. *Proteomics - Clinical Applications*, 1(2):148–156.

Mitchell, B. L., Yasui, Y., Li, C. I., Fitzpatrick, A. L., and Lampe, P. D. (2005). Impact of freeze-thaw cycles and storage time on plasma samples used in mass spectrometry based biomarker discovery projects. *Cancer informatics*, 1:98–104.

Mitchell, P. (2002). A perspective on protein microarrays. *Nature Biotechnology*, 20(3):225–229.

Miteva, Y. V., Budayeva, H. G., and Cristea, I. M. (2013). Proteomics-based methods for discovery, quantification, and validation of protein-protein interactions. *Analytical Chemistry*, 85(2):749–768.

Mokuno, K., Riku, S., Matsuoka, Y., Sobue, I., and Kato, K. (1985). Serum carbonic anhydrase III in progressive muscular dystrophy. *Journal of the Neurological Sciences*, 67(2):223–228.

Mor, G., Visintin, I., Lai, Y., Zhao, H., Schwartz, P., Rutherford, T., Yue, L., Bray-Ward, P., and Ward, D. C. (2005). Serum protein markers for early detection of ovarian cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7677–7682.

Morente, M. M., Fernández, P. L., and de Alava, E. (2008). Biobanking: old activity or young discipline? *Seminars in Diagnostic Pathology*, 25(4):317–322.

Morgan, B. P. and Gasque, P. (1997). Extrahepatic complement biosynthesis: where, when and why? *Clinical and experimental immunology*, 107(1):1–7.

Mueller, C., Liotta, L. A., and Espina, V. (2010). Reverse phase protein microarrays advance to use in clinical trials. *Molecular Oncology*, 4(6):461–481.

Müller-Eberhard, H. J. (1986). The membrane attack complex of complement. *Annual Review of Immunology*, 4(1):503–528.

Müller-Eberhard, H. J. (1988). Molecular Organization and Function of the Complement System. *Annual Review of Biochemistry*, 57:321–347.

Müller-Eberhard, H. J. and Schreiber, R. D. (1980). Molecular biology and chemistry of the alternative pathway of complement. *Advances in Immunology*, 29:1–53.

Nagele, E., Han, M., Demarshall, C., Belinka, B., and Nagele, R. (2011). Diagnosis of Alzheimer's disease based on disease-specific autoantibody profiles in human sera. *PloS ONE*, 6(8):e23112.

Nagele, E. P., Han, M., Acharya, N. K., Demarshall, C., Kosciuk, M. C., and Nagele, R. G. (2013). Natural IgG Autoantibodies Are Abundant and Ubiquitous in Human Sera, and Their Number Is Influenced By Age, Gender, and Disease. *PloS ONE*, 8(4):e60726.

Nahid, P., Bliven-Sizemore, E., Jarlsberg, L. G., De Groote, M. A., Johnson, J. L., Muzanyi, G., Engle, M., Weiner, M., Janjic, N., Sterling, D. G., and Ochsner, U. A. (2014). Aptamer-based proteomic signature of intensive phase treatment response in pulmonary tuberculosis. *Tuberculosis*, 94(3):187–196.

Naparstek, Y. and Plotz, P. H. (1993). The role of autoantibodies in autoimmune disease. *Annual Review of Immunology*, 11(1):79–104.

Natesan, M. and Ulrich, R. G. (2010). Protein microarrays and biomarkers of infectious disease. *International journal of molecular sciences*, 11(12):5165–5183.

National Institute of Health-Definitions Working Group (2000). Biomarkers and surrogate endpoints in clinical research: definitions and conceptual model. In *Biomarkers and Surrogate Endpoints. Elsevier*, pages 1–9, Amsterdam.

National Institute of Health-The Autoimmune Disease Coordinating Committee (2005). Progress in Autoimmune Diseases Research. Technical report, National Institute of Health.

Neiman, M., Fredolini, C., Johansson, H., Lehtiö, J., Nygren, P.-Å., Uhlén, M., Nilsson, P., and Schwenk, J. M. (2013). Selectivity analysis of single binder assays used in plasma protein profiling. *Proteomics*, 13(23-24):3406–3410.

Neiman, M., Hedberg, J. J., Dönnes, P. R., Schuppe-Koistinen, I., Hanschke, S., Schindler, R., Uhlén, M., Schwenk, J. M., and Nilsson, P. (2011). Plasma profiling reveals human fibulin-1 as candidate marker for renal impairment. *Journal of Proteome Research*, 10(11):4925–4934.

Neumann, E. J. and Bonistalli, K. N. (2009). Effect of blood sample handling post-collection on Erysipelothrix rhusiopathiae antibody titres. *Veterinary Journal*, 180(3):325–329.

Nicol, G. R., Han, M., Kim, J., Birse, C. E., Brand, E., Nguyen, A., Mesri, M., FitzHugh, W., Kaminker, P., Moore, P. A., Ruben, S. M., and He, T. (2008). Use of an immunoaffinity-mass spectrometry-based approach for the quantification of protein biomarkers from serum samples of lung cancer patients. *Molecular & Cellular Proteomics: MCP*, 7(10):1974–1982.

Nielsen, U. B., Cardone, M. H., Sinskey, A. J., MacBeath, G., and Sorger, P. K. (2003). Profiling receptor tyrosine kinase activation by using Ab microarrays. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16):9330–9335.

Niemeyer, C. M., Adler, M., and Wacker, R. (2005). Immuno-PCR: high sensitivity detection of proteins by nucleic acid amplification. *Trends in Biotechnology*, 23(4):208–216.

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B. P., and Landegren, U. (1994). Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science*, 265(5181):2085–2088.

Nilsson, P., Paavilainen, L., Larsson, K., Odling, J., Sundberg, M., Andersson, A.-C., Kampf, C., Persson, A., Al-Khalili Szigyarto, C., Ottosson, J., Björling, E., Hober, S., Wernérus, H., Wester, K., Ponten, F., and Uhlén, M. (2005). Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics*, 5(17):4327–4337.

Nirenberg, M. W. and Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 47:1588–1602.

Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W. C., Waltham, M., Kouros-Mehr, H., Bussey, K. J., Lee, J. K., Espina, V., Munson, P. J., Petricoin, E., Liotta, L. A., and Weinstein, J. N. (2003). Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proceedings of the National Academy of Sciences of the United States of America*, 100(24):14229–14234.

Nisonoff, A., Wissler, F. C., Lipman, L. N., and Woernley, D. L. (1960). Separation of univalent fragments from the bivalent rabbit antibody molecule by reduction of disulfide bonds. *Archives of biochemistry and biophysics*, 89:230–244.

Noble, W. S. (2009). How does multiple testing correction work? *Nature Biotechnology*, 27(12):1135–1137.

Nonaka, M. (2001). Evolution of the complement system. *Current Opinion In Immunology*, 13(1):69–73.

Nong, R. Y., Gu, J., Darmanis, S., Kamali-Moghaddam, M., and Landegren, U. (2012). DNA-assisted protein detection technologies. *Expert Review of Proteomics*, 9(1):21–32.

Nygren, P.-Å. (2008). Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold. *The FEBS Journal*, 275(11):2668–2676.

O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, 250(10):4007–4021.

Ohta, M., Itagaki, Y., Itoh, N., Hayashi, K., Nishitani, H., and Ohta, K. (1991). Carbonic anhydrase III in serum in muscular dystrophy and other neurological disorders: relationship with creatine kinase. *Clinical Chemistry*, 37(1):36–39.

Oliveira, B. M., Coorssen, J. R., and Martins-de Souza, D. (2014). 2DE: The Phoenix of Proteomics. *Journal of Proteomics*, 104:140–150.

Olsson, N., Wingren, C., Mattsson, M., James, P., O'Connell, D., Nilsson, F., Cahill, D. J., and Borrebaeck, C. A. K. (2011). Proteomic analysis and discovery using affinity proteomics and mass spectrometry. *Molecular & Cellular Proteomics: MCP*, 10(10):M110.003962.

Olsson, T. (1992). Immunology of multiple sclerosis. *Current Opinion in Neurology*, 5(2):195.

Omenn, G. S. (2004). The Human Proteome Organization Plasma Proteome Project pilot phase: reference specimens, technology platform comparisons, and standardized data submissions and analyses. *Proteomics*, 4(5):1235–1240.

Omenn, G. S. (2007). The HUPO Human Plasma Proteome Project. *Proteomics - Clinical Applications*, 1(8):769–779.

Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H., Paik, Y.-K., Yoo, J.-S., Ping, P., Pounds, J., Adkins, J., Qian, X., Wang, R., Wasinger, V., Wu, C. Y., Zhao, X., Zeng, R., Archakov, A., Tsugita, A., Beer, I., Pandey, A., Pisano, M., Andrews, P., Tammen, H., Speicher, D. W., and Hanash, S. M. (2005). Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics*, 5(13):3226–3245.

O'Neal, W. K., Anderson, W., Basta, P. V., Carretta, E. E., Doerschuk, C. M., Barr, R. G., Bleecker, E. R., Christenson, S. A., Curtis, J. L., Han, M. K., Hansel, N. N., Kanner, R. E., Kleerup, E. C., Martinez, F. J., Miller, B. E., Peters, S. P., Rennard, S. I., Scholand, M. B., Tal-Singer, R., Woodruff, P. G., Couper, D. J., Davis, S. M., and SPIROMICS Investigators (2014). Comparison of serum, EDTA plasma and P100 plasma for luminex-based biomarker multiplex assays in patients with chronic obstructive pulmonary disease in the SPIROMICS study. *Journal of Translational Medicine*, 12:9.

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. *Molecular & Cellular Proteomics: MCP*, 1(5):376–386.

Orchekowski, R., Hamelinck, D., Li, L., Gliwa, E., vanBrocklin, M., Marrero, J. A., Vande Woude, G. F., Feng, Z., Brand, R., and Haab, B. B. (2005). Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer. *Cancer Research*, 65(23):11193–11202.

Orenes-Piñero, E., Barderas, R., Rico, D., Casal, J. I., Gonzalez-Pisano, D., Navajo, J., Algaba, F., Piulats, J. M., and Sanchez-Carbayo, M. (2010). Serum and tissue profiling in bladder cancer combining protein and tissue arrays. *Journal of Proteome Research*, 9(1):164–173.

Ostroff, R., Foreman, T., Keeney, T. R., Stratford, S., Walker, J. J., and Zichi, D. (2010). The stability of the circulating human proteome to variations in sample collection and handling procedures measured with an aptamer-based proteomics array. *Journal of Proteomics*, 73(3):649–666.

Paone, J. F., Waalkes, T. P., Baker, R. R., and Shaper, J. H. (1980). Serum UDP-galactosyl transferase as a potential biomarker for breast carcinoma. *Journal of surgical oncology*, 15(1):59–66.

Papp, K., Szekeres, Z., Erdei, A., and Prechl, J. (2008a). Two-dimensional immune profiles improve antigen microarray-based characterization of humoral immunity. *Proteomics*, 8(14):2840–2848.

Papp, K., Szekeres, Z., Terényi, N., Isaák, A., Erdei, A., and Prechl, J. (2007). On-chip complement activation adds an extra dimension to antigen microarrays. *Molecular & Cellular Proteomics: MCP*, 6(1):133–140.

Papp, K., Végh, P., Hóbor, R., Erdei, A., and Prechl, J. (2012a). Characterization of factors influencing on-chip complement activation to optimize parallel measurement of antibody and complement proteins on antigen microarrays. *Journal of Immunological Methods*, 375(1-2):75–83.

Papp, K., Végh, P., Hóbor, R., Szittner, Z., Vokó, Z., Podani, J., Czirják, L., and Prechl, J. (2012b). Immune complex signatures of patients with active and inactive SLE revealed by multiplex protein binding analysis on antigen microarrays. *PloS ONE*, 7(9):e44824.

Papp, K., Végh, P., Miklós, K., Németh, J., Rásky, K., Péterfy, F., Erdei, A., and Prechl, J. (2008b). Detection of complement activation on antigen microarrays generates functional antibody profiles and helps characterization of disease-associated changes of the antibody repertoire. *Journal of Immunology*, 181(11):8162–8169.

Papsidero, L. D., Wang, M. C., Valenzuela, L. A., Murphy, G. P., and Chu, T. M. (1980). A prostate antigen in sera of prostatic cancer patients. *Cancer Research*, 40(7):2428–2432.

Parkin, J. and Cohen, B. (2001). An overview of the immune system. *The Lancet*, 357(9270):1777–1789.

Partyka, K., Wang, S., Zhao, P., Cao, B., and Haab, B. (2014). Array-based immunoassays with rolling-circle amplification detection. *Methods in Molecular Biology*, 1105:3–15.

Paulovich, A. G., Whiteaker, J. R., Hoofnagle, A. N., and Wang, P. (2008). The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. *Proteomics - Clinical Applications*, 2(10-11):1386–1402.

Pauly, F., Smedby, K. E., Jerkeman, M., Hjalgrim, H., Ohlsson, M., Rosenquist, R., Borrebaeck, C. A. K., and Wingren, C. (2014). Identification of B-cell lymphoma subsets by plasma protein profiling using recombinant antibody microarrays. *Leukemia research*, 38(6):682–690.

Pavlou, M. P., Diamandis, E. P., and Blasutig, I. M. (2013). The long journey of cancer biomarkers from the bench to the clinic. *Clinical Chemistry*, 59(1):147–157.

Paweletz, C. P., Charboneau, L., Bichsel, V. E., Simone, N. L., Chen, T., Gillespie, J. W., Emmert-Buck, M. R., Roth, M. J., Petricoin, III, E. F., and Liotta, L. A. (2001). Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene*, 20(16):1981–1989.

Pawlak, M., Schick, E., Bopp, M. A., Schneider, M. J., Oroszlan, P., and Ehrat, M. (2002). Zeptosens' protein microarrays: a novel high performance microarray platform for low abundance protein analysis. *Proteomics*, 2(4):383–393.

Pellois, J. P., Wang, W., and Gao, X. (2000). Peptide synthesis based on t-Boc chemistry and solution photogenerated acids. *Journal of combinatorial chemistry*, 2(4):355–360.

Pellois, J. P., Zhou, X., Srivannavit, O., Zhou, T., Gulari, E., and Gao, X. (2002). Individually addressable parallel peptide synthesis on microchips. *Nature Biotechnology*, 20(9):922–926.

Pender, M. P. (2004). *Epstein-Barr Virus and Autoimmunity*. Infection and Autoimmunity. Elsevier.

Perkel, J. M. (2014). Antibody challenge. *BioTechniques*, pages 1–3.

Pernemalm, M., De Petris, L., Eriksson, H., Brandén, E., Koyi, H., Kanter, L., Lewensohn, R., and Lehtiö, J. (2009). Use of narrow-range peptide IEF to improve detection of lung adenocarcinoma markers in plasma and pleural effusion. *Proteomics*, 9(13):3414–3424.

Pernemalm, M. and Lehtiö, J. (2014). Mass spectrometry-based plasma proteomics: state of the art and future outlook. *Expert Review of Proteomics*, 11(4):431–448.

Phizicky, E., Bastiaens, P. I. H., Zhu, H., Snyder, M., and Fields, S. (2003). Protein analysis on a proteomic scale. *Nature*, 422(6928):208–215.

Pla-Roca, M., Leulmi, R. F., Tourekhanova, S., Bergeron, S., Laforte, V., Moreau, E., Gosline, S. J. C., Bertos, N., Hallett, M., Park, M., and Juncker, D. (2012). Antibody colocalization microarray: a scalable technology for multiplex protein analysis in complex samples. *Molecular & Cellular Proteomics: MCP*, 11(4):M111.011460.

Plotz, P. H. (2003). The autoantibody repertoire: searching for order. *Nature Reviews Immunology*, 3(1):73–78.

Poetz, O., Henzler, T., Hartmann, M., Kazmaier, C., Templin, M. F., Herget, T., and Joos, T. O. (2010). Sequential Multiplex Analyte Capturing for Phosphoprotein Profiling. *Molecular & Cellular Proteomics: MCP*, 9(11):2474–2481.

Poetz, O., Hoeppe, S., Templin, M. F., Stoll, D., and Joos, T. O. (2009). Proteome wide screening using peptide affinity capture. *Proteomics*, 9(6):1518–1523.

Poetz, O., Ostendorp, R., Brocks, B., Schwenk, J. M., Stoll, D., Joos, T. O., and Templin, M. F. (2005). Protein microarrays for antibody profiling: specificity and affinity determination on a chip. *Proteomics*, 5(9):2402–2411.

Ponten, F., Schwenk, J. M., Asplund, A., and Edqvist, P.-H. D. (2011). The Human Protein Atlas as a proteomic resource for biomarker discovery. *Journal of Internal Medicine*, 270(5):428–446.

Pordeus, V., Szyper-Kravitz, M., Levy, R. A., Vaz, N. M., and Shoenfeld, Y. (2008). Infections and Autoimmunity: A Panorama. *Clinical Reviews in Allergy & Immunology*, 34(3):283–299.

Porter, R. R. (1959). The hydrolysis of rabbit γ-globulin and antibodies with crystalline papain. *Biochemical Journal*, 73(1):119.

Prechl, J., Papp, K., and Erdei, A. (2010). Antigen microarrays: descriptive chemistry or functional immunomics? *Trends in Immunology*, 31(4):133–137.

Prechl, J., Szittner, Z., and Papp, K. (2012). Complementing antibody profiles: assessing antibody function on antigen microarrays. *Immunology Letters*, 143(1):101–105.

Price, J. V., Haddon, D. J., Kemmer, D., Delepine, G., Mandelbaum, G., Jarrell, J. A., Gupta, R., Balboni, I., Chakravarty, E. F., Sokolove, J., Shum, A. K., Anderson, M. S., Cheng, M. H., Robinson, W. H., Browne, S. K., Holland, S. M., Baechler, E. C., and Utz, P. J. (2013a). Protein microarray analysis reveals BAFF-binding autoantibodies in systemic lupus erythematosus. *The Journal of Clinical Investigation*, 123(12):5135–5145.

Price, J. V., Jarrell, J. A., Furman, D., Kattah, N. H., Newell, E., Dekker, C. L., Davis, M. M., and Utz, P. J. (2013b). Characterization of influenza vaccine immunogenicity using influenza antigen microarrays. *PloS ONE*, 8(5):e64555.

Price, J. V., Tangsombatvisit, S., Xu, G., Yu, J., Levy, D., Baechler, E. C., Gozani, O., Varma, M., Utz, P. J., and Liu, C. L. (2012). On silico peptide microarrays for high-resolution mapping of antibody epitopes and diverse protein-protein interactions. *Nature Medicine*, 18(9):1434–1440.

Prudent, M., Tissot, J.-D., and Lion, N. (2011). Proteomics of blood and derived products: what's next? *Expert Review of Proteomics*, 8(6):717–737.

Ptolemy, A. S. and Rifai, N. (2010). What is a biomarker? Research investments and lack of clinical integration necessitate a review of biomarker terminology and validation schema. *Scandinavian Journal of Clinical & Laboratory Investigation*, 70(s242):6–14.

Puchois, P., Miranda, L. B., and van Gool, A. (2013). The Cardinal Role of Biobanks and Human Biospecimen Collections in Biomarker Validation: Issues Impeding Impact of Biomarker Research Outcomes. In *Comprehensive Biomarker Discovery and Validation for Clinical Application*, pages 73–110. Royal Society of Chemistry, Cambridge.

Querol, L., Clark, P. L., Bailey, M. A., Cotsapas, C., Cross, A. H., Hafler, D. A., Kleinstein, S. H., Lee, J.-Y., Yaari, G., Willis, S. N., and O'Connor, K. C. (2013). Protein array-based profiling of CSF identifies RBPJ as an autoantigen in multiple sclerosis. *Neurology*, 81(11):956–963.

Quintana, F. J., Farez, M. F., Viglietta, V., Iglesias, A. H., Merbl, Y., Izquierdo, G., Lucas, M., Basso, A. S., Khoury, S. J., Lucchinetti, C. F., Cohen, I. R., and Weiner, H. L. (2008). Antigen microarrays identify unique serum autoantibody signatures in clinical and pathologic subtypes of multiple sclerosis. *Proceedings of the National Academy of Sciences of the United States of America*, 105(48):18889–18894.

Quintana, F. J. F., Farez, M. F. M., Izquierdo, G. G., Lucas, M. M., Cohen, I. R. I., and Weiner, H. L. H. (2012). Antigen microarrays identify CNS-produced autoantibodies in RRMS. *Neurology*, 78(8):532–539.

Qundos, U., Hong, M.-G., Tybring, G., Divers, M., Odeberg, J., Uhlén, M., Nilsson, P., and Schwenk, J. M. (2013). Profiling post-centrifugation delay of serum and plasma with antibody bead arrays. *Journal of Proteomics*, 95:46–54.

Qundos, U., Johannesson, H., Fredolini, C., O'Hurley, G., Branca, R., Uhlén, M., Wiklund, F., Bjartell, A., Nilsson, P., and Schwenk, J. M. (2014). Analysis of plasma from prostate cancer patients links decreased carnosine dipeptidase 1 levels to lymph node metastasis. *Translational Proteomics*, 2:14–24.

Racanelli, V., Prete, M., Musaraj, G., Dammacco, F., and Perosa, F. (2011). Autoimmunity Reviews. *Autoimmunity Reviews*, 10(8):503–508.

Rai, A. J., Gelfand, C. A., Haywood, B. C., Warunek, D. J., Yi, J., Schuchard, M. D., Mehigh, R. J., Cockrill, S. L., Scott, G. B. I., Tammen, H., Schulz-Knappe, P., Speicher, D. W., Vitzthum, F., Haab, B. B., Siest, G., and Chan, D. W. (2005). HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics*, 5(13):3262–3277.

Rai, A. J. and Vitzthum, F. (2006). Effects of preanalytical variables on peptide and protein measurements in human serum and plasma: implications for clinical proteomics. *Expert Review of Proteomics*, 3(4):409–426.

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A. Y., Walter, J. C., and LaBaer, J. (2004). Self-assembling protein microarrays. *Science*, 305(5680):86–90.

Ramachandran, N., Raphael, J. V., Hainsworth, E., Demirkan, G., Fuentes, M. G., Rolfs, A., Hu, Y., and LaBaer, J. (2008). Next-generation high-density self-assembling functional protein arrays. *Nature Methods*, 5(6):535–538.

Ramirez, A. B., Loch, C. M., Zhang, Y., Liu, Y., Wang, X., Wayner, E. A., Sargent, J. E., Sibani, S., Hainsworth, E., Mendoza, E. A., Eugene, R., LaBaer, J., Urban, N. D., McIntosh, M. W., and Lampe, P. D. (2010). Use of a single-chain antibody library for ovarian cancer biomarker discovery. *Molecular & Cellular Proteomics: MCP*, 9(7):1449–1460.

Ransohoff, D. F. (2005). Bias as a threat to the validity of cancer molecular-marker research. *Nature Reviews Cancer*, 5(2):142–149.

Raven, P., Johnson, G., Mason, K., Losos, J., and Singer, S. (2010). *Biology*. McGraw-Hill Education, 9th edition.

Ray, S., Mehta, G., and Srivastava, S. (2010). Label-free detection techniques for protein microarrays: prospects, merits and challenges. *Proteomics*, 10(4):731–748.

Reddy, M. M., Wilson, R., Wilson, J., Connell, S., Gocke, A., Hynan, L., German, D., and Kodadek, T. (2011). Identification of candidate IgG biomarkers for Alzheimer's disease via combinatorial library screening. *Cell*, 144(1):132–142.

Reeves, W. H. and Satoh, M. (1996). Features of autoantigens. *Molecular Biology Reports*, 23(3-4):217–226.

Ricklin, D., Hajishengallis, G., Yang, K., and Lambris, J. D. (2010). Complement: a key system for immune surveillance and homeostasis. *Nature Immunology*, 11(9):785–797.

Riegman, P. H. J., Morente, M. M., Betsou, F., de Blasio, P., Geary, P., and on Biobanking for Biomedical Research, M. A. I. W. G. (2008). Biobanking for better healthcare. *Molecular Oncology*, 2(3):213–222.

Rifai, N., Gillette, M. A., and Carr, S. A. (2006). Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnology*, 24(8):971–983.

Rimini, R., Schwenk, J. M., Sundberg, M., Sjöberg, R., Klevebring, D., Gry, M., Uhlén, M., and Nilsson, P. (2009). Validation of serum protein profiles by a dual antibody array approach. *Journal of Proteomics*, 73(2):252–266.

Ringnér, M. (2008). What is principal component analysis? *Nature Biotechnology*, 26(3):303–304.

Rissin, D. M., Kan, C. W., Campbell, T. G., Howes, S. C., Fournier, D. R., Song, L., Piech, T., Patel, P. P., Chang, L., Rivnak, A. J., Ferrell, E. P., Randall, J. D., Provuncher, G. K., Walt, D. R., and Duffy, D. C. (2010). Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nature Biotechnology*, 28(6):595–599.

Rissin, D. M. and Walt, D. R. (2006). Digital concentration readout of single enzyme molecules using femtoliter arrays and Poisson statistics. *Nano letters*, 6(3):520–523.

Robinson, W. H., DiGennaro, C., Hueber, W., Haab, B. B., Kamachi, M., Dean, E. J., Fournel, S., Fong, D., Genovese, M. C., de Vegvar, H. E. N., Skriner, K., Hirschberg, D. L., Morris, R. I., Muller, S., Pruijn, G. J., van Venrooij, W. J., Smolen, J. S., Brown, P. O., Steinman, L., and Utz, P. J. (2002). Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nature Medicine*, 8(3):295–301.

Robinson, W. H., Fontoura, P., Lee, B. J., de Vegvar, H. E. N., Tom, J., Pedotti, R., DiGennaro, C. D., Mitchell, D. J., Fong, D., Ho, P. P.-K., Ruiz, P. J., Maverakis, E., Stevens, D. B., Bernard, C. C. A., Martin, R., Kuchroo, V. K., van Noort, J. M., Genain, C. P., Amor, S., Olsson, T., Utz, P. J., Garren, H., and Steinman, L. (2003a). Protein microarrays guide tolerizing DNA vaccine treatment of autoimmune encephalomyelitis. *Nature Biotechnology*, 21(9):1033–1039.

Robinson, W. H., Steinman, L., and Utz, P. J. (2003b). Protein arrays for autoantibody profiling and fine-specificity mapping. *Proteomics*, 3(11):2077–2084.

Roche, S., Gabelle, A., and Lehmann, S. (2008). Clinical proteomics of the cerebrospinal fluid: Towards the discovery of new biomarkers. *Proteomics - Clinical Applications*, 2(3):428–436.

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics: MCP*, 3(12):1154–1169.

Rusmini, F., Zhong, Z., and Feijen, J. (2007). Protein immobilization strategies for protein biochips. *Biomacromolecules*, 8(6):1775–1789.

Saah, A. J. and Hoover, D. R. (1997). "Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings. *Annals of internal medicine*, 126(1):91–94.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839):487–491.

Sakaguchi, S. (2000). Regulatory T cells: key controllers of immunologic self-tolerance. *Cell*, 101(5):455–458.

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(1):93–103.

Sandström, A., Andersson, R., Segersvärd, R., Löhr, M., Borrebaeck, C. A. K., and Wingren, C. (2012). Serum proteome profiling of pancreatitis using recombinant antibody microarrays reveals disease-associated biomarker signatures. *Proteomics - Clinical Applications*, 6(9-10):486–496.

Sanger, F. (1945). The free amino groups of insulin. *Biochemical Journal*, 39:507–515.

Sano, T., Smith, C., and Cantor, C. (1992). Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science*, 258(5079):120–122.

Sattlecker, M., Kiddle, S. J., Newhouse, S., Proitsi, P., Nelson, S., Williams, S., Johnston, C., Killick, R., Simmons, A., Westman, E., Hodges, A., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Lovestone, S., the AddNeuroMed Consortium, and Dobson, R. J. B. (2014). Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimer's & Dementia*.

Schellekens, G. A., de Jong, B. A., van den Hoogen, F. H., van de Putte, L. B., and van Venrooij, W. J. (1998). Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. *The Journal of Clinical Investigation*, 101(1):273–281.

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(5235):467–470.

Schenk, S., Schoenhals, G. J., de Souza, G., and Mann, M. (2008). A high confidence, manually validated human blood plasma protein reference set. *BMC Medical Genomics*, 1:41.

Schiess, R., Wollscheid, B., and Aebersold, R. (2009). Targeted proteomic strategy for clinical biomarker discovery. *Molecular Oncology*, 3(1):33–44.

Schirmer, L., Srivastava, R., and Hemmer, B. (2014). To look for a needle in a haystack: the search for autoantibodies in multiple sclerosis. *Multiple Sclerosis*, 20(3):271–279.

Schmidt, R., Cook, E. A., Kastelic, D., Taussig, M. J., and Stoevesandt, O. (2013). Optimised 'on demand' protein arraying from DNA by cell free expression with the 'DNA to Protein Array' (DAPA) technology. *Journal of Proteomics*, 88:141–148.

Schofield, D. J., Pope, A. R., Clementel, V., Buckell, J., Chapple, S. D., Clarke, K. F., Conquer, J. S., Crofts, A. M., Crowther, S. R. E., Dyson, M. R., Flack, G., Griffin, G. J., Hooks, Y., Howat, W. J., Kolb-Kokocinski, A., Kunze, S., Martin, C. D., Maslen, G. L., Mitchell, J. N., O'Sullivan, M., Perera, R. L., Roake, W., Shadbolt, S. P., Vincent, K. J., Warford, A., Wilson, W. E., Xie, J., Young, J. L., and McCafferty, J. (2007). Application of phage display to high throughput antibody generation and characterization. *Genome Biology*, 8(11):R254.

Schreiber, R., Uliyakina, I., Kongsuphol, P., Warth, R., Mirza, M., Martins, J. R., and Kunzelmann, K. (2010). Expression and function of epithelial anoctamins. *The Journal of Biological Chemistry*, 285(10):7838–7845.

Schröder, C., Jacob, A., Tonack, S., Radon, T. P., Sill, M., Zucknick, M., Rüffer, S., Costello, E., Neoptolemos, J. P., Crnogorac-Jurcevic, T., Bauer, A., Fellenberg, K., and Hoheisel, J. D. (2010). Dual-color proteomic profiling of complex samples with a microarray of 810 cancer-related antibodies. *Molecular & Cellular Proteomics: MCP*, 9(6):1271–1280.

Schrohl, A. S., Wurtz, S., Kohn, E., Banks, R. E., Nielsen, H. J., Sweep, F. C. G. J., and Brunner, N. (2008). Banking of Biological Fluids for Studies of Disease-associated Protein Biomarkers. *Molecular & Cellular Proteomics: MCP*, 7(10):2061–2066.

Schutzer, S. E., Liu, T., Natelson, B. H., Angel, T. E., Schepmoes, A. A., Purvine, S. O., Hixson, K. K., Lipton, M. S., Camp, D. G., Coyle, P. K., Smith, R. D., and Bergquist, J. (2010). Establishing the proteome of normal human cerebrospinal fluid. *PloS ONE*, 5(6):e10980.

Schwartz, M. and Cohen, I. R. (2000). Autoimmunity can benefit self-maintenance. *Immunology Today*, 21(6):265–268.

Schwartz, R. H. (1989). Acquisition of immunologic self-tolerance. *Cell*, 57(7):1073–1081.

Schweitzer, B., Roberts, S., Grimwade, B., Shao, W., Wang, M., Fu, Q., Shu, Q., Laroche, I., Zhou, Z., Tchernev, V. T., Christiansen, J., Velleca, M., and Kingsmore, S. F. (2002). Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nature Biotechnology*, 20(4):359–365.

Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S., Kukanskis, K., Zhu, Z., Kingsmore, S., Lizardi, P., and Ward, D. (2000). Immunoassays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen detection. *Proceedings of the National Academy of Sciences of the United States of America*, 97(18):10113.

Schwenk, J. M., Gry, M., Rimini, R., Uhlén, M., and Nilsson, P. (2008). Antibody suspension bead arrays within serum proteomics. *Journal of Proteome Research*, 7(8):3168–3179.

Schwenk, J. M., Igel, U., Kato, B. S., Nicholson, G., Karpe, F., Uhlén, M., and Nilsson, P. (2010a). Comparative protein profiling of serum and plasma using an antibody suspension bead array approach. *Proteomics*, 10(3):532–540.

Schwenk, J. M., Igel, U., Neiman, M., Langen, H., Becker, C., Bjartell, A., Ponten, F., Wiklund, F., Grönberg, H., Nilsson, P., and Uhlén, M. (2010b). Toward next generation plasma profiling via heat-induced epitope retrieval and array-based assays. *Molecular & Cellular Proteomics: MCP*, 9(11):2497–2507.

Schwenk, J. M., Lindberg, J., Sundberg, M., Uhlén, M., and Nilsson, P. (2007). Determination of binding specificities in highly multiplexed bead-based assays for antibody proteomics. *Molecular & Cellular Proteomics: MCP*, 6(1):125–132.

Schwenk, J. M., Stoll, D., Templin, M. F., and Joos, T. O. (2002). Cell microarrays: an emerging technology for the characterization of antibodies. *BioTechniques*, Suppl:54–61.

Sebbag, M., Moinard, N., Auger, I., Clavel, C., Arnaud, J., Nogueira, L., Roudier, J., and Serre, G. (2006). Epitopes of human fibrin recognized by the rheumatoid arthritis-specific autoantibodies to citrullinated proteins. *European Journal of Immunology*, 36(8):2250–2263.

Sedlaczek, P., Frydecka, I., Gabryś, M., van Dalen, A., Einarsson, R., and Harłozińska, A. (2002). Comparative analysis of CA125, tissue polypeptide specific antigen, and soluble interleukin-2 receptor alpha levels in sera, cyst, and ascitic fluids from patients with ovarian carcinoma. *Cancer*, 95(9):1886–1893.

Sela, M., Schechter, B., Schechter, I., and Borek, F. (1967). Antibodies to Sequential and Conformational Determinants. *Cold Spring Harbor Symposia on Quantitative Biology*, 32(0):537–545.

Serneels, S. and Verdonck, T. (2008). Principal component analysis for data containing outliers and missing elements. *Computational Statistics & Data Analysis*, 52(3):1712–1727.

Sevecka, M., Wolf-Yadlin, A., and MacBeath, G. (2011). Lysate microarrays enable high-throughput, quantitative investigations of cellular signaling. *Molecular & Cellular Proteomics: MCP*, 10(4):M110.005363.

Shafer, M. W., Mangold, L., Partin, A. W., and Haab, B. B. (2007). Antibody array profiling reveals serum TSP-1 as a marker to distinguish benign from malignant prostatic disease. *The Prostate*, 67(3):255–267.

Shao, W., Zhou, Z., Laroche, I., Lu, H., Zong, Q., Patel, D. D., Kingsmore, S., and Piccoli, S. P. (2003). Optimization of Rolling-Circle Amplified Protein Microarrays for Multiplexed Protein Profiling. *Journal of biomedicine & biotechnology*, 2003(5):299–307.

Sheehan, K. M., Calvert, V. S., Kay, E. W., Lu, Y., Fishman, D., Espina, V., Aquino, J., Speer, R., Araujo, R., Mills, G. B., Liotta, L. A., Petricoin, E. F., and Wulfkuhle, J. D. (2005). Use of reverse phase protein microarrays and reference standard development for molecular network analysis of metastatic ovarian carcinoma. *Molecular & Cellular Proteomics: MCP*, 4(4):346–355.

Shen, Y., Kim, J., Strittmatter, E. F., Jacobs, J. M., Camp, D. G., Fang, R., Tolié, N., Moore, R. J., and Smith, R. D. (2005). Characterization of the human blood plasma proteome. *Proteomics*, 5(15):4034–4045.

Shieh, A. D. and Hung, Y. S. (2009). Detecting Outlier Samples in Microarray Data. *Statistical Applications in Genetics and Molecular Biology*, 8(1):1–24.

Shin, D.-S., Lee, K.-N., Yoo, B.-W., Kim, J., Kim, M., Kim, Y.-K., and Lee, Y.-S. (2010). Automated maskless photolithography system for peptide microarray synthesis on a chip. *Journal of combinatorial chemistry*, 12(4):463–471.

Shlomchik, M. J., Marshak-Rothstein, A., Wolfowicz, C. B., Rothstein, T. L., and Weigert, M. G. (1987). The role of clonal selection and somatic mutation in autoimmunity. *Nature*, 328(6133):805–811.

Shoenfeld, Y. (2013). Everything is autoimmune until proven otherwise. *Clinical Reviews in Allergy & Immunology*, 45(2):149–151.

Shoenfeld, Y., Selmi, C., Zimlichman, E., and Gershwin, M. E. (2008a). The autoimmunologist: geoepidemiology, a new center of gravity, and prime time for autoimmunity. *Journal of Autoimmunity*, 31(4):325–330.

Shoenfeld, Y. and Toubi, E. (2005). Protective autoantibodies: Role in homeostasis, clinical importance, and therapeutic potential. *Arthritis and Rheumatism*, 52(9):2599–2606.

Shoenfeld, Y., Zandman-Goddard, G., Stojanovich, L., Cutolo, M., Amital, H., Levy, Y., Abu-Shakra, M., Barzilai, O., Berkun, Y., Blank, M., de Carvalho, J. F., Doria, A., Gilburd, B., Katz, U., Krause, I., Langevitz, P., Orbach, H., Pordeus, V., Ram, M., Toubi, E., and Sherer, Y. (2008b). The mosaic of autoimmunity: hormonal and environmental factors involved in autoimmune diseases–2008. *The Israel Medical Association journal: IMAJ*, 10(1):8–12.

Shovman, O., Gilburd, B., Barzilai, O., Shinar, E., Larida, B., Zandman-Goddard, G., Binder, S. R., and Shoenfeld, Y. (2005). Evaluation of the BioPlex 2200 ANA screen: analysis of 510 healthy subjects: incidence of natural/predictive autoantibodies. *Annals of the New York Academy of Sciences*, 1050:380–388.

Shreffler, W. G. (2011). Microarrayed recombinant allergens for diagnostic testing. *Journal of Allergy and Clinical Immunology*, 127(4):843–849.

Sidhu, S. S. (2012). Antibodies for all: The case for genome-wide affinity reagents. *FEBS Letters*, 586(17):2778–2779.

Silman, A. J. and Pearson, J. E. (2002). Epidemiology and genetics of rheumatoid arthritis. *Arthritis Research*, 4 Suppl 3:S265–72.

Silverstein, A. M. (1991). The dynamics of conceptual change in twentieth century immunology. *Cellular Immunology*, 132(2):515–531.

Silverstein, A. M. (2001a). Autoimmunity versus horror autotoxicus: The struggle for recognition. *Nature Immunology*, 2(4):279–281.

Silverstein, A. M. (2001b). *Paul Ehrlich's Receptor Immunology: The Magnificent Obsession*. Academic Press, 1st edition.

Silverstein, A. M. (2009). *A History of Immunology*. Elsevier, 2nd edition.

Simon, M., Girbal, E., Sebbag, M., Gomès-Daudrix, V., Vincent, C., Salama, G., and Serre, G. (1993). The cytokeratin filament-aggregating protein filaggrin is the target of the so-called "antikeratin antibodies," autoantibodies specific for rheumatoid arthritis. *The Journal of Clinical Investigation*, 92(3):1387–1393.

Singh-Gasson, S., Green, R. D., Yue, Y., Nelson, C., Blattner, F., Sussman, M. R., and Cerrina, F. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nature Biotechnology*, 17(10):974–978.

Sjöberg, R., Sundberg, M., Gundberg, A., Sivertsson, A., Schwenk, J. M., Uhlén, M., and Nilsson, P. (2012). Validation of affinity reagents using antigen microarrays. *New Biotechnology*, 29(5):555–563.

Sjölander, A., Nygren, P.-Å., Ståhl, S., Berzins, K., Uhlén, M., Perlmann, P., and Andersson, R. (1997). The serum albumin-binding region of streptococcal protein G: a bacterial fusion partner with carrier-related properties. *Journal of Immunological Methods*, 201(1):115–123.

Skates, S. J., Gillette, M. A., LaBaer, J., Carr, S. A., Anderson, L., Liebler, D. C., Ransohoff, D., Rifai, N., Kondratovich, M., Težak, Ž., Mansfield, E., Oberg, A. L., Wright, I., Barnes, G., Gail, M., Mesri, M., Kinsinger, C. R., Rodriguez, H., and Boja, E. S. (2013). Statistical Design for Biospecimen Cohort Size in Proteomics-based Biomarker Discovery and Verification Studies. *Journal of Proteome Research*, 12(12):5383–5394.

Skerra, A. (2000). Lipocalins as a scaffold. *Biochimica et Biophysica Acta*, 1482(1-2):337–350.

Slaastad, H., Wu, W., Goullart, L., Kanderova, V., Tjønnfjord, G., Stuchly, J., Kalina, T., Holm, A., and Lund-Johansen, F. (2011). Multiplexed immuno-precipitation with 1725 commercially available antibodies to cellular proteins. *Proteomics*, 11(23):4578–4582.

Smit, S., Hoefsloot, H. C. J., and Smilde, A. K. (2008). Statistical data processing in clinical proteomics. *Journal of Chromatography B*, 866(1-2):77–88.

Smith, K. A. (2004). Demystified... recombinant antibodies. *Journal of Clinical Pathology*, 57(9):912–917.

Smith, L. M., Kelleher, N. L., and Consortium for Top Down Proteomics (2013). Proteoform: a single term describing protein complexity. *Nature Methods*, 10(3):186–187.

Söderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N., Alexeiva, V., Aberg, A. M., Nilsson, A., Jansson, B., Ohlin, M., Wingren, C., Danielsson, L., Carlsson, R., and Borrebaeck, C. A. (2000). Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nature Biotechnology*, 18(8):852–856.

Song, Q., Liu, G., Hu, S., Zhang, Y., Tao, Y., Han, Y., Zeng, H., Huang, W., Li, F., Chen, P., Zhu, J., Hu, C., Zhang, S., Li, Y., Zhu, H., and Wu, L. (2010). Novel autoimmune hepatitis-specific autoantigens identified using protein microarray technology. *Journal of Proteome Research*, 9(1):30–39.

Sospedra, M. and Martin, R. (2005). Immunology of Multiple Sclerosis. *Annual Review of Immunology*, 23(1):683–747.

Speer, R., Wulfkuhle, J., Espina, V., Aurajo, R., Edmiston, K. H., Liotta, L. A., and Petricoin, E. F. (2007). Development of reverse phase protein microarrays for clinical applications and patient-tailored therapy. *Cancer genomics & proteomics*, 4(3):157–164.

Spellman, F. R. (2007). *Biology for Nonbiologists*. Government Institutes, 1st edition.

Spirin, A. S. and Swartz, J. R. (2008). Cell-Free Protein Synthesis Systems: Historical Landmarks, Classification, and General Methods. *Wiley*, pages 1–34.

Sreekumar, A., Nyati, M. K., Varambally, S., Barrette, T. R., Ghosh, D., Lawrence, T. S., and Chinnaiyan, A. M. (2001). Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Research*, 61(20):7585–7593.

Srinivasan, H., Allory, Y., Sill, M., Vordos, D., Alhamdani, M. S. S., Radvanyi, F., Hoheisel, J. D., and Schröder, C. (2014). Prediction of recurrence of non muscle-invasive bladder cancer by means of a protein signature identified by antibody microarray analyses. *Proteomics*, 14(11):1333–1342.

Srivastava, R. R., Aslam, M. M., Kalluri, S. R. S., Schirmer, L. L., Buck, D. D., Tackenberg, B. B., Rothhammer, V. V., Chan, A. A., Gold, R. R., Berthele, A. A., Bennett, J. L. J., Korn, T. T., and Hemmer, B. B. (2012). Potassium channel KIR4.1 as an immune target in multiple sclerosis. *The New England Journal of Medicine*, 367(2):115–123.

Stadler, V., Felgenhauer, T., Beyer, M., Fernandez, S., Leibe, K., Güttler, S., Gröning, M., König, K., Torralba, G., Hausmann, M., Lindenstruth, V., Nesterov, A., Block, I., Pipkorn, R., Poustka, A., Bischoff, F. R., and Breitling, F. (2008). Combinatorial synthesis of peptide arrays with a laser printer. *Angewandte Chemie (International ed. in English)*, 47(37):7132–7135.

Stafford, P., Cichacz, Z., Woodbury, N. W., and Johnston, S. A. (2014). Immunosignature system for diagnosis of cancer. *Proceedings of the National Academy of Sciences of the United States of America*.

States, D. J., Omenn, G. S., Blackwell, T. W., Fermin, D., Eng, J., Speicher, D. W., and Hanash, S. M. (2006). Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nature Biotechnology*, 24(3):333–338.

Steen, H. and Mann, M. (2004). The abc's (and xyz's) of peptide sequencing. *Nature Reviews Molecular Cell Biology*, 5(9):699–711.

Steinman, L. (1996). Multiple sclerosis: a coordinated immunological attack against myelin in the central nervous system. *Cell*, 85(3):299–302.

Stoevesandt, O. and Taussig, M. J. (2012a). Affinity proteomics: the role of specific binding reagents in human proteome analysis. *Expert Review of Proteomics*, 9(4):401–414.

Stoevesandt, O. and Taussig, M. J. (2012b). European and international collaboration in affinity proteomics. *New Biotechnology*, 29(5):511–514.

Stoevesandt, O., Vetter, M., Kastelic, D., Palmer, E. A., He, M., and Taussig, M. J. (2011). Cell free expression put on the spot: advances in repeatable protein arraying from DNA (DAPA). *New Biotechnology*, 28(3):282–290.

Sutandy, F. X. R., Qian, J., Chen, C.-S., and Zhu, H. (2013). Overview of protein microarrays. *Current Protocols in Protein Science*, Chapter 27:Unit 27.1.

Svedberg, T. and Rinde, H. (1924). The ultra-centrifuge, a new instrument for the determination of size and distribution of size of particle in amicroscopic colloids. *Journal of the American Chemical Society*.

Szodoray, P., Alex, P., Brun, J. G., Centola, M., and Jonsson, R. (2004). Circulating Cytokines in Primary Sjogren's Syndrome Determined by a Multiplex Cytokine Array System. *Scandinavian Journal of Immunology*, 59(6):592–599.

Tam, S. W., Wiese, R., Lee, S., Gilmore, J., and Kumble, K. D. (2002). Simultaneous analysis of eight human Th1/Th2 cytokines using microarrays. *Journal of Immunological Methods*, 261(1-2):157–165.

Tan, E. M. and Zhang, J. (2008). Autoantibodies to tumor-associated antigens: reporters from the immune system. *Immunological Reviews*, 222:328–340.

Tao, S.-C. and Zhu, H. (2006). Protein chip fabrication by capture of nascent polypeptides. *Nature Biotechnology*, 24(10):1253–1254.

Taussig, M. J., Schmidt, R., Cook, E. A., and Stoevesandt, O. (2013). Development of proteome-wide binding reagents for research and diagnostics. *Proteomics - Clinical Applications*, 7(11-12):756–766.

Taussig, M. J., Stoevesandt, O., Borrebaeck, C. A. K., Bradbury, A. R., Cahill, D., Cambillau, C., de Daruvar, A., Dübel, S., Eichler, J., Frank, R., Gibson, T. J., Gloriam, D., Gold, L., Herberg, F. W., Hermjakob, H., Hoheisel, J. D., Joos, T. O., Kallioniemi, O., Koegll, M., Konthur, Z., Korn, B., Kremmer, E., Krobitsch, S., Landegren, U., van der Maarel, S., McCafferty, J., Muyldermans, S., Nygren, P.-Å., Palcy, S., Plückthun, A., Polic, B., Przybylski, M., Saviranta, P., Sawyer, A., Sherman, D. J., Skerra, A., Templin, M., Ueffing, M., and Uhlén, M. (2007). ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nature Methods*, 4(1):13–17.

Tegel, H., Steen, J., Konrad, A., Nikdin, H., Pettersson, K., Stenvall, M., Tourle, S., Wrethagen, U., Xu, L., Yderland, L., Uhlén, M., Hober, S., and Ottosson, J. (2009). High-throughput protein production - Lessons from scaling up from 10 to 288 recombinant proteins per week. *Biotechnology Journal*, 4(1):51–57.

Tegel, H., Tourle, S., Ottosson, J., and Persson, A. (2010). Increased levels of recombinant human proteins with the Escherichia coli strain Rosetta(DE3). *Protein Expression and Purification*, 69(2):159–167.

Teng, P.-n., Bateman, N. W., Hood, B. L., and Conrads, T. P. (2010). Advances in Proximal Fluid Proteomics for Disease Biomarker Discovery. *Journal of Proteome Research*, 9(12):6091–6100.

Tessler, L. A. and Mitra, R. D. (2011). Sensitive single-molecule protein quantification and protein complex detection in a microarray format. *Proteomics*, 11(24):4731–4735.

Thomson, J. J. (1897). Cathode Rays. *Philosophical Magazine*, 44(269):293–316.

Thongboonkerd, V. (2007). *Proteomics of human body fluids*. Humana Press, 1st edition.

Tighe, P., Negm, O., Todd, I., and Fairclough, L. (2013). Utility, reliability and reproducibility of immunoassay multiplex kits. *Methods*, 61(1):23–29.

Tiselius, A. (1937). A new apparatus for electrophoretic analysis of colloidal mixtures. *Transactions of the Faraday Society*, 33:524–531.

Tiselius, A. and Kabat, E. A. (1939). An Electrophoretic Study of Immune Sera and Purified Antibody Preparations. *The Journal of Experimental Medicine*, 69(1):119–131.

Todd, J., Freese, B., Lu, A., Held, D., Morey, J., Livingston, R., and Goix, P. (2007). Ultrasensitive flow-based immunoassays using single-molecule counting. *Clinical Chemistry*, 53(11):1990–1995.

Tomizaki, K.-y., Usui, K., and Mihara, H. (2005). Protein-detecting microarrays: current accomplishments and requirements. *ChemBioChem*, 6(5):782–799.

Tuerk, C. and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249(4968):505–510.

Uhlén, M. (2008). Affinity as a tool in life science. *BioTechniques*, 44(5):649–654.

Uhlén, M., Björling, E., Agaton, C., Szigyarto, C. A.-K., Amini, B., Andersen, E., Andersson, A.-C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergström, K., Brumer, H., Cerjan, D., Ekström, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Björklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundeberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Sköllermo, A., Steen, J., Stenvall, M., Sterky, F., Strömberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Waldén, A., Wan, J., Wernérus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S., and Pontén, F. (2005). A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Molecular & Cellular Proteomics: MCP*, 4(12):1920–1932.

Uhlén, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernérus, H., Björling, L., and Ponten, F. (2010). Towards a knowledge-based Human Protein Atlas. *Nature Biotechnology*, 28(12):1248–1250.

Uhlén, M. and Ponten, F. (2005). Antibody-based proteomics for human tissue profiling. *Molecular & Cellular Proteomics: MCP*, 4(4):384–393.

Ummanni, R., Mannsperger, H. A., Sonntag, J., Oswald, M., Sharma, A. K., König, R., and Korf, U. (2014). Evaluation of reverse phase protein array (RPPA)-based pathway-activation profiling in 84 non-small cell lung cancer (NSCLC) cell lines as platform for cancer proteomics and biomarker discovery. *Biochimica et Biophysica Acta*, 1844(5):950–959.

Unlü, M., Morgan, M. E., and Minden, J. S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis*, 18(11):2071–2077.

Uotila, M., Ruoslahti, E., and Engvall, E. (1981). Two-site sandwich enzyme immunoassay with monoclonal antibodies to human alpha-fetoprotein. *Journal of Immunological Methods*, 42(1):11–15.

Utz, P. J. (2005). Protein arrays for studying blood cells and their secreted products. *Immunological Reviews*, 204:264–282.

van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., and van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7:142.

van der Woude, D., Syversen, S. W., van der Voort, E. I. H., Verpoort, K. N., Goll, G. L., van der Linden, M. P. M., van der Helm-van Mil, A. H. M., van der Heijde, D. M. F. M., Huizinga, T. W. J., Kvien, T. K., and Toes, R. E. M. (2010). The ACPA isotype profile reflects long-term radiographic progression in rheumatoid arthritis. *Annals of the Rheumatic Diseases*, 69(6):1110–1116.

Vanderlugt, C. L. and Miller, S. D. (2002). Epitope Spreading in Immune-Mediated Diseases: Implications for Immunotherapy. *Nature Reviews Immunology*, 2(2):85–95.

Vanmeter, A., Signore, M., Pierobon, M., Espina, V., Liotta, L. A., and Petricoin, E. F. (2007). Reverse-phase protein microarrays: application to biomarker discovery and translational medicine. *Expert Review of Molecular Diagnostics*, 7(5):625–633.

Venner, A. A., Ibañez, D., Gladman, D. D., Urowitz, M. B., MacKinnon, A., Blasutig, I. M., and Yip, P. M. (2013). Comparison of three anti-dsDNA assays: performance and correlation with systemic lupus erythematosus disease activity. *Clinical Biochemistry*, 46(4-5):317–320.

Vermeulen, N., de Béeck, K. O., Vermeire, S., Van Steen, K., Michiels, G., Ballet, V., Rutgeerts, P., and Bossuyt, X. (2011). Identification of a novel autoantigen in inflammatory bowel disease by protein microarray. *Inflammatory bowel diseases*, 17(6):1291–1300.

Vessman, J., Stefan, R. I., van Staden, J. F., Danzer, K., Lindner, W., Burns, D. T., Fajgelj, A., and Müller, H. (2001). Selectivity in analytical chemistry (IUPAC Recommendations 2001). *Pure and Applied Chemistry*, 73(8):1381–1386.

Vickery, H. B. (1950). The origin of the word protein. *The Yale journal of biology and medicine*, 22(5):387–393.

Vigil, A., Chen, C., Jain, A., Nakajima-Sasaki, R., Jasinskas, A., Pablo, J., Hendrix, L. R., Samuel, J. E., and Felgner, P. L. (2011). Profiling the humoral immune response of acute and chronic Q fever by protein microarray. *Molecular & Cellular Proteomics: MCP*, 10(10):M110.006304.

Vigil, A., Davies, D. H., and Felgner, P. L. (2010). Defining the humoral immune response to infectious agents using high-density protein microarrays. *Future Microbiology*, 5(2):241–251.

Vincent, A., Lily, O., and Palace, J. (1999). Pathogenic autoantibodies to neuronal proteins in neurological disorders. *Journal of Neuroimmunology*, 100(1-2):169–180.

Vitzthum, F., Behrens, F., Anderson, N. L., and Shaw, J. H. (2005). Proteomics: from basic research to diagnostic application. A review of requirements & needs. *Journal of Proteome Research*, 4(4):1086–1097.

Volk, S., Schreiber, T. D., Eisen, D., Wiese, C., Planatscher, H., Pynn, C. J., Stoll, D., Templin, M. F., Joos, T. O., and Pötz, O. (2012). Combining ultracentrifugation and peptide termini group-specific immunoprecipitation for multiplex plasma protein analysis. *Molecular & Cellular Proteomics: MCP*, 11(7):O111.015438.

Wacholder, S., Silverman, D. T., McLaughlin, J. K., and Mandel, J. S. (1992). Selection of controls in case-control studies: II. Types of controls. *American Journal of Epidemiology*, 135(9):1029–1041.

Wallstrom, G., Anderson, K. S., and LaBaer, J. (2013). Biomarker Discovery for Heterogeneous Diseases. *Cancer Epidemiology, Biomarkers & Prevention*, 22(5):747–755.

Walport, M. J. (2001). Complement. First of Two Parts. *The New England Journal of Medicine*, 344(14):1058–1066.

Walther, T. C. and Mann, M. (2010). Mass spectrometry-based proteomics in cell biology. *The Journal of cell biology*, 190(4):491–500.

Wang, J., Barker, K., Steel, J., Park, J., Saul, J., Festa, F., Wallstrom, G., Yu, X., Bian, X., Anderson, K. S., Figueroa, J. D., LaBaer, J., and Qiu, J. (2013). A versatile protein microarray platform enabling antibody profiling against denatured proteins. *Proteomics - Clinical Applications*, 7(5-6):378–383.

Wang, X., Yu, J., Sreekumar, A., Varambally, S., Shen, R., Giacherio, D., Mehra, R., Montie, J., Pienta, K., and Sanda, M. (2005). Autoantibody signatures in prostate cancer. *The New England Journal of Medicine*, 353(12):1224.

Wardemann, H. (2003). Predominant Autoantibody Production by Early Human B Cell Precursors. *Science*, 301(5638):1374–1377.

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

Watson, J. D. and Crick, F. H. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356):737–738.

Weiß, F., van den Berg, B. H. J., Planatscher, H., Pynn, C. J., Joos, T. O., and Poetz, O. (2014). Catch and measure-mass spectrometry-based immunoassays in biomarker research. *Biochimica et Biophysica Acta*, 1844(5):927–932.

Whitacre, C. C. (2001). Sex differences in autoimmune disease. *Nature Immunology*, 2(9):777–780.

Wichmann, H.-E., Kuhn, K. A., Waldenberger, M., Schmelcher, D., Schuffenhauer, S., Meitinger, T., Wurst, S. H. R., Lamla, G., Fortier, I., Burton, P. R., Peltonen, L., Perola, M., Metspalu, A., Riegman, P., Landegren, U., Taussig, M. J., Litton, J.-E., Fransson, M. N., Eder, J., Cambon-Thomsen, A., Bovenberg, J., Dagher, G., van Ommen, G.-J., Griffith, M., Yuille, M., and Zatloukal, K. (2011). Comprehensive catalog of European biobanks. *Nature Biotechnology*, 29(9):795–797.

Wide, L., Bennich, H., and Johansson, S. G. O. (1967). Diagnosis of Allergy by an in-vitro Test for Allergen Antibodies. *The Lancet*, 290(7526):1105–1107.

Wiese, R., Belosludtsev, Y., Powdrill, T., Thompson, P., and Hogan, M. (2001). Simultaneous multianalyte ELISA performed on a microarray platform. *Clinical Chemistry*, 47(8):1451–1457.

Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmair, A., Faerber, F., and Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, 509(7502):582–587.

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., Williams, K. L., and Hochstrasser, D. F. (1996). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/technology (Nature Publishing Company)*, 14(1):61–65.

Wilson, D. L. (2000). *11th Hour: Introduction to Biology*. Blackwell Publishing, 1st edition.

Wilson, R. (2013). Sensitivity and specificity: twin goals of proteomics assays. Can they be combined? *Expert Review of Proteomics*, 10(2):135–149.

Wiltshire, S., O'Malley, S., Lambert, J., Kukanskis, K., Edgar, D., Kingsmore, S. F., and Schweitzer, B. (2000). Detection of multiple allergen-specific IgEs on microarrays by immunoassay with rolling circle amplification. *Clinical Chemistry*, 46(12):1990–1993.

Wingren, C. and Borrebaeck, C. A. (2008). Antibody microarray analysis of directly labelled complex proteomes. *Current Opinion In Biotechnology*, 19(1):55–61.

Wingren, C. and Borrebaeck, C. A. K. (2007). Progress in miniaturization of protein arrays—a step closer to high-density nanoarrays. *Drug Discovery Today*, 12(19-20):813–819.

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(17):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(6):1511–1517.

Wingren, C., Sandstrom, A., Segersvard, R., Carlsson, A., Andersson, R., Lohr, M., and Borrebaeck, C. A. K. (2012). Identification of Serum Biomarker Signatures Associated with Pancreatic Cancer. *Cancer Research*, 72(10):2481–2490.

Wingren, C., Steinhauer, C., Ingvarsson, J., Persson, E., Larsson, K., and Borrebaeck, C. A. K. (2005). Microarrays based on affinity-tagged single-chain Fv antibodies: sensitive detection of analyte in complex proteomes. *Proteomics*, 5(5):1281–1291.

Winters, R., Winters, A., and Amedee, R. G. (2010). Statistics: a brief overview. *The Ochsner journal*, 10(3):213–216.

Wolf-Yadlin, A., Sevecka, M., and MacBeath, G. (2009). Dissecting protein function and signaling using protein microarrays. *Current Opinion in Chemical Biology*, 13(4):398–405.

Won, J.-H., Goldberger, O., Shen-Orr, S. S., Davis, M. M., and Olshen, R. A. (2012). Significance analysis of xMap cytokine bead arrays. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8):2848–2853.

Wong, J., Sibani, S., Lokko, N. N., LaBaer, J., and Anderson, K. S. (2009). Rapid detection of antibodies in sera using multiplexed self-assembling bead arrays. *Journal of Immunological Methods*, 350(1-2):171–182.

Wood, P. J. (2006). *Understanding Immunology (Cell and molecular biology in action)* . Pearson Education Limited, 2nd edition.

Wright, C., Sibani, S., Trudgian, D., Fischer, R., Kessler, B., LaBaer, J., and Bowness, P. (2012). Detection of multiple autoantibodies in patients with ankylosing spondylitis using nucleic acid programmable protein arrays. *Molecular & Cellular Proteomics: MCP*, 11(2):M9.00384.

Wright, D. B. (2003). Making friends with your data: improving how statistics are conducted and reported. *The British journal of educational psychology*, 73(Pt 1):123–136.

Wu, A. H. B., Fukushima, N., Puskas, R., Todd, J., and Goix, P. (2006). Development and preliminary clinical validation of a high sensitivity assay for cardiac troponin using a capillary flow (single molecule) fluorescence detector. *Clinical Chemistry*, 52(11):2157–2159.

Wu, W., Slåstad, H., de la Rosa Carrillo, D., Frey, T., Tjønnfjord, G., Boretti, E., Aasheim, H.-C., Horejsi, V., and Lund-Johansen, F. (2009). Antibody array analysis with label-based detection and resolution of protein size. *Molecular & Cellular Proteomics: MCP*, 8(2):245–257.

Wucherpfennig, K. W. (2001). Mechanisms for the induction of autoimmunity by infectious agents. *The Journal of Clinical Investigation*, 108(8):1097–1104.

Wulfkuhle, J. D., Aquino, J. A., Calvert, V. S., Fishman, D. A., Coukos, G., Liotta, L. A., and Petricoin, E. F. (2003). Signal pathway profiling of ovarian cancer from human tissue specimens using reverse-phase protein microarrays. *Proteomics*, 3(11):2085–2090.

Yalow, R. S. and Berson, S. A. (1959). Assay of Plasma Insulin in Human Subjects by Immunological Methods. *Nature*, 184(4699):1648–1649.

Yan, W., Apweiler, R., Balgley, B. M., Boontheung, P., Bundy, J. L., Cargile, B. J., Cole, S., Fang, X., Gonzalez-Begne, M., Griffin, T. J., Hagen, F., Hu, S., Wolinsky, L. E., Lee, C. S., Malamud, D., Melvin, J. E., Menon, R., Mueller, M., Oiao, R., Rhodus, N. L., Sevinsky, J. R., States, D., Stephenson, J. L. J., Than, S., Yates, J. R. I., Yu, W., Xie, H., Xie, Y., Omenn, G. S., Loo, J. A., and Wong, D. T. (2009). Systematic comparison of the human saliva and plasma proteomes. *Proteomics - Clinical Applications*, 3(1):116–134.

Yates, III, J. R. (2011). A century of mass spectrometry: from atoms to proteomes. *Nature Methods*, 8(8):633.

Yoshida, S. and Gershwin, M. E. (1993). Autoimmunity and selected environmental factors of disease induction. *Seminars in Arthritis and Rheumatism*, 22(6):399–419.

Yu, X., Schneiderhan-Marra, N., and Joos, T. O. (2010). Protein microarrays for personalized medicine. *Clinical Chemistry*, 56(3):376–387.

Yuille, M., van Ommen, G. J., Brechot, C., Cambon-Thomsen, A., Dagher, G., Landegren, U., Litton, J. E., Pasterk, M., Peltonen, L., Taussig, M., Wichmann, H.-E., and Zatloukal, K. (2007). Biobanking for Europe. *Briefings in Bioinformatics*, 9(1):14–24.

Zaenker, P. and Ziman, M. R. (2013). Serologic Autoantibodies as Diagnostic Cancer Biomarkers–a Review. *Cancer Epidemiology, Biomarkers & Prevention*, 22(12):2161–2181.

Zhao, X., Qureshi, F., Eastman, P. S., Manning, W. C., Alexander, C., Robinson, W. H., and Hesterberg, L. K. (2012). Pre-analytical effects of blood sampling and handling in quantitative immunoassays for rheumatoid arthritis. *Journal of Immunological Methods*, 378(1-2):72–80.

Zhong, L., Hidalgo, G. E., Stromberg, A. J., Khattar, N. H., Jett, J. R., and Hirschowitz, E. A. (2005). Using Protein Microarray as a Diagnostic Assay for Non–Small Cell Lung Cancer. *American Journal of Respiratory and Critical Care Medicine*, 172(10):1308–1314.

Zhou, C., Simpson, K. L., Lancashire, L. J., Walker, M. J., Dawson, M. J., Unwin, R. D., Rembielak, A., Price, P., West, C., Dive, C., and Whetton, A. D. (2012a). Statistical Considerations of Optimal Study Design for Human Plasma Proteomics and Biomarker Discovery. *Journal of Proteome Research*, 11(4):2103–2113.

Zhou, H., Bouwman, K., Schotanus, M., Verweij, C., Marrero, J. A., Dillon, D., Costa, J., Lizardi, P., and Haab, B. B. (2004). Two-color, rolling-circle amplification on antibody microarrays for sensitive, multiplexed serum-protein measurements. *Genome Biology*, 5(4):R28.

Zhou, L., Zhao, S. Z., Koh, S. K., Chen, L., Vaz, C., Tanavde, V., Li, X. R., and Beuerman, R. W. (2012b). In-depth analysis of the human tear proteome. *Journal of Proteomics*, 75(13):3877–3885.

Zhou, W., Ross, M. M., Tessitore, A., Ornstein, D., Vanmeter, A., Liotta, L. A., and Petricoin, E. F. (2009). An initial characterization of the serum phosphoproteome. *Journal of Proteome Research*, 8(12):5523–5531.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., and Snyder, M. (2001). Global analysis of protein activities using proteome chips. *Science*, 293(5537):2101–2105.

Zhu, H., Hu, S., Jona, G., Zhu, X., Kreiswirth, N., Willey, B. M., Mazzulli, T., Liu, G., Song, Q., Chen, P., Cameron, M., Tyler, A., Wang, J., Wen, J., Chen, W., Compton, S., and Snyder, M. (2006). Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray. *Proceedings of the National Academy of Sciences of the United States of America*, 103(11):4011–4016.

Zhu, H., Klemic, J. F., Chang, S., Bertone, P., Casamayor, A., Klemic, K. G., Smith, D., Gerstein, M., Reed, M. A., and Snyder, M. (2000). Analysis of yeast protein kinases using protein chips. *Nature Genetics*, 26(3):283–289.

Zhu, H. and Qian, J. (2012). Applications of functional protein microarrays in basic and clinical research. *Advances in genetics*, 79:123–155.

Zhu, H. and Snyder, M. (2003). Protein chip technology. *Current Opinion in Chemical Biology*, 7(1):55–63.

Zipfel, P. F. and Skerka, C. (2009). Complement regulators and inhibitory proteins. *Nature Reviews Immunology*, 9(10):729–740.

Zolg, W. (2006). The proteomic search for diagnostic biomarkers: lost in translation? *Molecular & Cellular Proteomics: MCP*, 5(10):1720–1726.

Part VI

Appendix: Included Articles