Development of Enhanced Molecular Diagnostic Tools for Protein Detection and Analysis

TONGE EBAI
Abstract

Improved diagnosis, prognosis and disease follow-up is a fundamental procedure and a constant challenge in medicine. Among the different molecular biomarkers, proteins are the essential regulatory component in blood; hence, by developing enhanced specific and sensitive molecular tools will give great insight into the different processes in disease treatment. In this thesis, we build on the proximity ligation assay to develop and apply new adaptable methods to facilitate protein detection.

In paper I, I present a variant of the proximity ligation assay (we call PLARCA) using micro titer plate for detection and quantification of protein using optical density as readout in the fluorometer. PLARCA detected femtomolar levels of these proteins in patient samples, which was considerably below the detection threshold for ELISA.

In paper II, we developed and adapted a new method into the in situ PLA methods for detection and identification of extracellular vesicles (EVs) using flow cytometry as readout (a method we call ExoPLA). We identified five target proteins on the surface of the EVs and using three colors, we identified the EV using flow cytometer.

In paper III, we aim to improve the efficiency of in situ PLA by creating and developing new designs and versions of the assay we called Unfold probes. Through comparison of detection of protein using in situ PLA versus Unfold probes, we observed considerable decrease in non-specific signals, and also a lower detection threshold.

In paper IV, we describe the development of a solid phase proximity extension (sp-PEA) assay for protein detection and quantification. We compared detection of IL-8, TNF-alpha, IL-10 and IL-6 using spPEA and PEA; spPEA demonstrations over 2 orders of magnitudes in the lower detection concentrations by decreased in background noise.

Keywords: protein detection, proximity ligation assays, proximity extension assay, rolling circle amplification, ELISA, flow cytometry, fluorescence microscopy

Tonge Ebai, Department of Immunology, Genetics and Pathology, Molecular tools, Rudbecklaboratoriet, Uppsala University, SE-751 85 Uppsala, Sweden.

© Tonge Ebai 2017

ISSN 1651-6206
ISBN 978-91-554-9930-3
urn:nbn:se:uu:diva-320380 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-320380)
“Never regard study as a duty, but as the enviable opportunity to learn to know the liberating influence of beauty in the realm of the spirit for your own personal joy and to the profit of the community to which your later work belongs”  Albert Einstein

To my Family and Friends
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the respective publishers.
Related work by author

Review Articles


Book Chapter

Contents

Part 1 ..........................................................................................................................11

Introduction ..............................................................................................................12

1. Human proteome .................................................................................................14
   1.1 Basic building blocks ....................................................................................14
   1.2 Bodily fluid ....................................................................................................17
   1.3 Biomarkers .....................................................................................................21
   1.4 Analytical guidelines and tests of method validation .....................................23

2 Affinity reagents and applications .......................................................................26
   2.1 Affinity reagents ............................................................................................27
   2.2 Labeling of affinity reagents ..........................................................................30

3. Proteomics .............................................................................................................32
   3.1 Proteomics ........................................................................................................32
   3.2 Looking into organelles ....................................................................................33

4 Historical, current and emerging proteomic technologies .................................35
   4.1 History; an analytical approach ......................................................................35
   4.2 Overview on current non-targeted proteomic technologies .........................36
   4.3 Overview of targeted proteomics technologies ............................................37
      4.3.1 DNA assisted immunoassays ..................................................................44
      4.3.2 Proximity-based DNA assisted immunoassay .......................................46
   4.4 Proximity based IA: Still lagging .....................................................................47

Part II .........................................................................................................................51

5 Present Investigations ............................................................................................52
   5.1 Sensitive protein detection in microtiter plates by proximity ligation with rolling circle amplification (PLARCA) - Paper I ...................53
      5.1.1 Introduction ...............................................................................................53
      5.1.2 Aim of study ..............................................................................................54
      5.1.3 Summary of finding ..................................................................................54
5.2 Detecting individual extracellular vesicles using a multicolor \textit{in situ} proximity ligation assay with flow cytometric readout - Paper II

5.2.1 Introduction
5.2.2 Aim of study
5.2.3 Summary of finding

5.3 Enhanced \textit{in situ} proximity ligation assays via Unfolding PLA probes - Paper III

5.3.1 Introduction
5.3.2 Aim of study
5.3.3 Summary of finding

5.4 Protein detection by sensitive magnetic bead-based proximity extension assays - Paper IV

III Conclusions and Perspectives

IV Acknowledgments

V References
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of variation</td>
</tr>
<tr>
<td>DBCO</td>
<td>Dibenzocyclooctyne</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DARPin</td>
<td>Designed ankyrin repeats proteins</td>
</tr>
<tr>
<td>EndoIV</td>
<td>Endonuclease IV</td>
</tr>
<tr>
<td>EIA</td>
<td>Enzyme immunoassay</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>EV</td>
<td>Extracellular vesicles</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence activated cell sorter</td>
</tr>
<tr>
<td>FITC</td>
<td>Fluorescein isothiocyanate</td>
</tr>
<tr>
<td>GDF-15</td>
<td>Growth differentiation factor 15</td>
</tr>
<tr>
<td>HRP</td>
<td>Horse radish peroxidase</td>
</tr>
<tr>
<td>LOD</td>
<td>Limit of detection</td>
</tr>
<tr>
<td>IF</td>
<td>Immunofluorescence</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>IRCA</td>
<td>Immuno rolling circle amplification</td>
</tr>
<tr>
<td>K_D</td>
<td>Dissociation constant</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PLA</td>
<td>Proximity ligation assay</td>
</tr>
<tr>
<td>PTM</td>
<td>Posttranslational modification</td>
</tr>
<tr>
<td>RCA</td>
<td>Rolling circle amplification</td>
</tr>
<tr>
<td>RCP</td>
<td>Rolling circle product</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>SP-PEA</td>
<td>Solid phase proximity extension assay</td>
</tr>
<tr>
<td>SP-PLA</td>
<td>Solid phase proximity ligation assay</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular endothelial growth factor</td>
</tr>
<tr>
<td>UNG</td>
<td>Uracil-DNA glycosylase</td>
</tr>
</tbody>
</table>
Part 1
All I wanted growing up was to be part of something that could help others. I wanted to be part of the solution; I wanted to be the change that made someone else’s life better. Developing diagnostic tools that could predict early on the course of an illness and prevent curable diseases became something I wanted to be part of. To be able to develop these diagnostics tools we need to understand the molecules (RNA, DNA, and proteins) that work cooperatively to fulfill different biological processes. The human body is a complex mass of cells interacting with each other and secreting enormous numbers of molecules that vary in size, concentrations, and functions but all acting together to perform different regulatory roles in the body. These secreted molecules make up the human plasma proteome; composed of millions of antibodies and a multitude of other classes of proteins, which are involved in different biological activities, ranging from modulation of pathological conditions (including cancer, autoimmune diseases), the mediator of cellular responses, to modulation of receptor-mediated signal transduction. The plasma proteome has for decades been a rich source of biomolecules that represent the health of the human body and used as a diagnostic medium. The plasma proteome has an enormous role in circulation; it contains proteins that serve as messengers between organs such as hormones, and can indicate disease state of organs when leaked into blood, for instance, in cases of myocardial infarction, prostate cancer, and also proteins molecular machinery, which carry out several structural, catalytic, metabolic processes in living systems. Therefore, quantifying the relative abundance of these proteins gives us an opportunity to characterize little errors related to diseases and follow disease progressions in individual patients. To understand, discover, detect, characterize, validate and analyze these biomolecules in the human body, strategies and tools are required. This thesis presented here aimed at developing new diagnostic tools to address the issue of detecting, quantifying and analyzing proteins. The concept of this work was a target-based approach to use the existing tools and improve the performance by creating new designs, detections platforms, and application of enhanced tools in clinically relevant diseases. Herein presented are four new molecular tools that were developed for this thesis.

Paper I describes the combination of the proximity ligation and in situ assay systems to develop a new assay system. This variant assay system uses readout mode which complements already existing instrumentations used in
hospital laboratories and clinics. Paper II present a new method called ExoPLA for detection of extracellular vesicles (EVs). Extracellular vesicles are membrane-bound vesicles released from cells, may be relevant biomarkers and functions as cell-to-cell communicator agents carrying RNA, DNA, and proteins.

Investigating spatial localization of proteins in cells and tissue in situ is an excellent way to understand signaling events and protein-protein interactions. The primary goal of Paper III was to increase the efficiency of in situ protein detection. We had realized that the design for the current in situ protein detection method suffered from the limited efficiency that arises from non-circular templated that cannot be amplified. My role in this project was to measure quantitatively how much efficiency gained with this new method that we call UnFold. UnFold was used to evaluated the measurement cytokines spiked in non-human serum in comparison with in situ.

Another tool that can measure and quantify protein in serum, plasma, and blood is the proximity extension assay (PEA). PEA is a powerful assay as up to 92 proteins, and four controls measured in just one µL of blood with great sensitivity. However, low abundant proteins are usually missed because there are few molecules to be detected in such small volumes, and because of contributions from background signals. Paper IV founded on the hypothesis that by using a solid phase, increasing sample volume, provide a greater number of molecules to be detected; also excess reagents may be removed by washes, and the use of sets of three, rather than two antibodies, decrease the risk of nonspecific reactions. All these effects can help increasing the detection of the low abundant protein.

The tools presented in this thesis will contribute to the detection, quantification, and analysis of proteins, extracellular vesicles and applied to different diseases as a diagnostic toolbox both for researchers in academia and industry and finally also in clinical laboratories in hospitals and clinics.
1. Human proteome

In *De Rerum Natura*, Lucretius wrote: “Nothing comes from nothing.” In the central dogma of molecular biology (1) put forward by Francis Crick in 1970, which detailed the allocation of information from DNA to RNA and then to protein but also stating unequivocally that the information in protein can not transferred to DNA or RNA. In deciding on how to write this thesis, I imagined the readers to be people ranging from non-science, non-biotechnology background to colleagues and others with clinical, biological and biotechnological expertise. So, in this part of the thesis, a basic overview of the biology and technical aspect will be discussed. This first part will focus on biomolecules, their disease reporting characteristics, challenges and analytical criteria when developing tools to detect and analyze these biomolecules.

1.1 Basic building blocks

In the 50s, the idea that DNA was the carrier of genetic information astounded biologist with the insight of how the simple molecular structure with two parallel chains consisting of four bases Adenine, Thymine, Cytosine, Guanine, each linked together covalently by sugars that are attached to phosphate groups in DNA. The structure and its implications were described by Watson and Crick (2). Langridge and colleagues in 1957 made refinement to the structure via further x-ray crystallography studies (3) where they confirmed double helix model brought forth by Watson and Crick. Meselson and Stahl confirmed this transfer of information from parental to daughter generations with experiments where they radio-labeled bacteria with N^{15} and grew them for 14 generations and using ultraviolet absorption photography. They observed that the nitrogen of the DNA was divided equally between subunits. It remained intact throughout many generations, also each daughter molecules had a copy of the parental subunit after replication, there was a doubling in the molecular machinery, and they showed differences of the heavy, light and mixed chains by ultracentrifugation as an effect of semiconservative replication (4).

As described in the central dogma of molecular biology, the information embodied in protein transferred from DNA. Nirenberg and Matthaei discovered the key to breaking the genetic code in 1960. As the saying goes, an
experiment is as good as its control, Nirenberg and Matthaei had developed an *in vitro* system for protein synthesis when they discovered that upon disruption of the cells, protein synthesis stopped. Trying to lengthen the short phase during which *in vitro* synthesis takes place, they added ribosomal RNA into the reaction where all 20 amino acids incorporated into the new proteins. Their control experiment was set up with mock RNA, and this proved more valuable as they equally had all 20 amino acid incorporated into the new protein. To confirm their discovery, they used the enzyme polynucleotide phosphorylase to synthesize random RNA molecules from available precursors without a template to form a mock RNA only consisting of U residues as the polyuridylic acid from UTP (5, 6). Protein synthesis occurred when they added the poly U into new cell suspension that was disrupted and interestingly, $^{14}$C labeled phenylalanine was incorporated into the protein. This results confirmed earlier work by Brenner and colleagues (7) hypothesizing that ribosomes cannot distinguish mock RNA from naturally occurring RNA. The mock mRNA carried the genetic code for phenylalanine as UUU, and the ribosomes read it with high efficiency. Additional genetic code for the other amino acid where rapidly annotated (8-10). This approach leads to the synthesis of synthetic mRNA thereby deciphering the full genetic code. Ten years earlier, George Gamow had proposed the genetic code (11) in his Diamond code whereby several triplets selected in a specific manner coded for any given amino acid, and these codes were degenerate and overlapping. In the 60s, Brenner, Crick, Barnett, and Watts-Tobin formulated that the three letter stands for a word, which was the codon instructing the incorporation of one amino acid.

The cell-free protein synthesis experiments carried out by Matthaei and Nirenberg concluded that DNA was not directly involved in protein synthesis, but RNA (Figure 1.1) was responsible for incorporating amino acids to form proteins and finally that this amino acid was representative of the genetic code, which occurred in triplets and that some of the codons were degenerate. They are the most complex of all biomolecules; there are 20 amino acids in proteins with each amino acid having a distinct structural and chemical composition. The word *protein* stems from the Greek word *proteios* meaning ‘the most important one’ or ‘first one’ (12), the name protein coined and adopted in the 18th and 19th centuries by JJ Berzelius and GJ Mulder. Proteins are made up of a string of amino acids each linked together by covalent peptide bonds. HE Fischer first proposed the word peptide. Proteins are polypeptides which form repeating units along the polypeptide backbone forming different structure (13). The structure grouped as (a) primary (linear sequence of amino acid); secondary (alpha helixes and beta sheets structures that are stabilized by hydrogen bonds); tertiary structures (depicts the formation of globular structure via folding of polypeptide chain); quaternary structure (showing the arrangement of multiple polypeptide subunits stacked together). F Sanger in 1951 determined the primary
structure of insulin (14) which further established Fischer’s proposal for the polypeptide nature of proteins. Proteins are the molecular machinery that is responsible for carrying out catalytic, regulatory and metabolic functions (enzymes, hormones), structural (bones), signaling transduction, cell cycle control, gene transcription, and translation. There exist about 20,500 protein-coding genes in the human genome (15-17), which give rise to the unknown number of protein variants (Figure 1.1). An enormous majority of the human genes are subjected to differential transcriptional start, variable splicing (18, 19), giving rise to proteins translated from different splice variants, processing and posttranslational modifications (20) (such as glycosylation, and phosphorylation). All these different protein exhibit large functions and are of the several diagnostic significances (21-23). Therefore, by characterizing and measuring proteins, we get a plethora of information about the health and disease state of individuals.

Proteins that are only transcribed when needed like regulatory proteins (24), pheromones in yeast (25); time-dependent genes like cell cycle genes in yeast (26) and humans (27); bacterial proteins optimized for genes for fast translation within short operons (28) and also in codon-containing genes (29).

Figure 1.1: Mechanism of protein synthesis from a single gene in a cell in eukaryotes.
In the nucleus, cellular machinery identifies the promoter and transcribes a gene into mRNA, which contain both introns and exons. Before transportation of the mRNA from the nucleus to the cytoplasm, the introns are removed by RNA splicing. The transcripts are also 5’ capped, spliced and 3’ polyadenylated and for each gene structurally distinct several primary transcripts are translated into proteins. Posttranslational modification, compartmentalization and proteolysis then regulate proteins function.

Huge initiatives at the Human Proteome Organization (HUPO), called the Human Proteome Project (HPP), aims to revolutionize our understanding of the relative abundance, protein localization and interacting partners of the human proteome by characterizing every protein from every predicted gene (30). Other project includes the Human Protein Atlas (HPA) whose main aim is to map protein distribution in healthy and cancer tissue (31).

1.2 Bodily fluid

In 1878, the French Physiologist Claude Bernard (also known as the founder of modern experimental physiology) wrote, "The stability of the internal environment (the *milieu intérieur*) is the condition for the free and independent life" a state that was later characterized as homeostasis by WB Cannon in 1939. All vital biochemical, mechanical and physiological function of humans happen in an environment with multiple functions, consisting of cardiovascular and vascular (capillaries, arteries, and veins) system. This environment sustain the body fluid, which makes up approximately 3/5th of the adult human body and divided into compartments. These compartments are mainly the intracellular and extracellular fluids. All compartments are vital but the extracellular one is of great importance when it comes to proteome mining, and the application opportunities that the tools presented in this thesis can provide. The extracellular fluid is made up of the macroenvironment (blood) and the microenvironment (interstitial tissue fluid: extracellular and cellular elements). Blood is the single most populated macroenvironment in the human body. Blood comprises about 8% of the total human body weight, and it is an exceptionally complex systemic fluid, which functions mainly as the transporter (nutrients, waste products, gas) systems, while it maintains homeostasis of ions, water, and pH in the body. Over the average lifespan of a human individual, the heartbeats over 2.5 billion times injecting over 200 million liters of blood. An average adult (of about 70 kg) has about 6 liters of blood that flows continuously throughout the body to sustain the body’s physiological functions. These physiological functions include the transfer of nutrients (oxygen, electrolytes, enzymes, hormones, carbon dioxide), to balance and maintain pH and control temperature and chemical composition within the intracellular and extracellular elements of the tissue microenvironments. The body’s arsenal for fighting all infectious agents and its de-
fense mechanism widely found in blood. The milieu intérieur brought forth by Bernard included the bodily fluid and maintenance therein, which led Starling to write the “Wisdom of the body” in which he acknowledged “that living organism preserve the constancy of their internal milieu notwithstanding the significant variations in food, water intake and other environmental” (32). About half of the whole blood volume is made up of different cell types including red blood cells (erythrocytes), white blood cells (leucocytes) and platelets (thrombocytes). The latter two are involved in the defense systems of the body. Blood is the body’s transporter, and about 55% of its volume is made up of a liquid fluid called serum and plasma. Plasma and serum prepared from blood by different means. Plasma is collected by treating blood with an anticoagulant (EDTA, heparin or sodium citrate) and removing the blood cells by centrifugation. In the collection of serum, no anticoagulant used and it is collected by removing all cells and clotting proteins. Plasma and serum are different not just in their mode of production but also in their qualitative content. The serum is void of numerous coagulation factors and fibrinogens, and it contains a higher amount of abundant proteins such as globulins (33) thus making plasma the fluid of choice in proteomics. Studies carried out by the HUPO/HPPP (34, 35) also concluded that plasma is preferable to serum due to less degradation. However, care should be taken during plasma preparation, as the choice of anticoagulant should be based on the intended end point protein analysis. In this thesis, plasma samples have been used to validate the tools developed as discussed in the paper I & II.

The plasma proteome is in constant communication with the tissue microenvironment delivering and receiving nutrients and signals via the lymphatic system. In a paper by Liotta et al. they wrote: “every cell in the body leaves a record of its physiological state in the products sheds into the blood” (36). Anderson and Anderson elaborated (37) the Putnam’s classification of the function of proteins by adding proteins that are (i) secreted from solid tissue (like liver and intestines), (ii) antibodies, (iii) proteins that act away from site of production like hormones, (iv) receptor ligands that function in mediating local responses and may have short residual time in plasma like cytokines, (v) tissue leakages protein that may arise as a result of cell death for example creatine kinase (38), (vi) aberration secretions, which mainly include proteins, cells released from cancer cells into plasma and (vii) foreign pathogen that infect and release pathogens into the blood. As explained by Liotta and Anderson, the plasma proteome is important for revealing the pathological and physiological state of humans and used extensively in biomedical research and clinically for diagnosis and prognosis of diseases. To further elaborate on the utility of these different class of protein, Leigh Anderson looked at a subset of an FDA-cleared or approved group of protein classified into the same category above as follows (i) proteins with function in plasma 45%, (ii) proteins leaked from tissue 25%, (ii) receptor ligand
proteins like cytokines 18%, (iii) aberrant secretions from cancer tissues 6% and (iv) immunoglobulin 6% (39). Blood is predominantly rich in red blood cells (carried hemoglobin for oxygen transfer), white blood cells (which includes basophils, eosinophils, neutrophils and monocytes) and platelets (important in homeostasis by preventing blood clot and loss). Questions often raised about the plasma proteome include; how many proteins are there?? What types of proteins are present? How much communication does the other proximal bodily fluid have with plasma? Anderson and Anderson in their epic tale on plasma proteome quotes the dynamic range of protein concentrations in plasma varies from albumin (most abundant ≈ 40 mg/mL (mM)) to interleukin-6 (in the pg/mL (pM) range) in the order of $10^{10}$ with IL-6 being the least in concentration (37, 39). As mentioned earlier, about the complexity and difference of the proteome and genome, the well-know human protein-coding gene has been approximated to about 20,500 (15), which does not indicate the different splice variants for each gene, the function of the different proteins and their utility in diagnostics or as biomarkers. As written by Landegren et al. “the question about how many plasma protein variants to distinguish for diagnostic purposes is unlikely to receive a clear-cut answer anytime soon” (40). Plasma communicates directly with the tissue microenvironment (tissue secretion or damage) and other proximal bodily fluids from body cells, tissues, and organs. To give a basic picture of the other bodily fluid, I list them in Figure 1.2 with their organs of origins. The plasma proteome provides opportunities such as the availability of most accessible soluble proteins that is acquired by noninvasive means and contains tissue-derived proteins. Some specific drawbacks include that it is extremely complex with an enormous uncharacterized dynamic range and unstandardized protocol. Addressing the problems of the broad dynamic range of the plasma proteome, the paper I demonstrates that with new molecular tools, the dynamic range can be increased by $10^2$ orders of magnitude when compared with current state of the art technology for detecting and analyzing protein in plasma. Also, the protein content and biochemical properties of these fluids are different e.g. the pH ranges from 1.7 – 8.2 from the gastric to the bile respectively. Several studies have been performed to characterize the proteome of saliva (41), tears (42) and CSF (43) that provides useful biomarker candidates for the diluted proteins in the plasma proteome. While some of these alternative sources are very reachable and noninvasive, such as urine and saliva, other sources are less accessible and more invasive methods are required to collect, e.g. biopsies. Tissue compartment, known as the tissue proteome is another source for protein analysis where detail information at the subcellular level shows protein distribution patterns, expression protein profiles, protein localization, functional aspects of the protein with focus to protein-protein interactions, posttranslational modifications, signal transduction, membrane-bound proteins and interaction between nuclear and cytoplasmic proteins. The drawback with these sample types is that
surgical procedure is needed and the samples have a non-homogenous representation due to cell-cell variations. The human protein atlas, an online resource where antibodies with peroxidases are used to map out spatial distribution of different proteins (44). In research setting, cell lines, fixed tissues are used to model experiments to understand the tissue proteome and **Paper III** in this thesis talk about novel tools for *in situ* detection and quantification.

**Figure 1.2: Bodily fluids**

Cell to cell communication is an important function in all organisms, and this exchange of information among cells occurs in the soluble matrix by direct interactions (45). Another class of biomolecules that is indicative of the state of humans is extracellular vesicles (EVs). These EVs are produced by many eukaryotic cells and contain mRNA, non-coding RNA, and proteins, which can be transported and delivered to other cell types and even different species with alternative functionalities (46). EVs are found in eukaryotes and prokaryotes (47, 48). Recently, lots of studies have been carried out to evaluate the potential of this class of potential biomarkers and their implications in diseases. EVs have been isolated from body fluids, and there are ever increasing evidence of the role of EVs in cell maintenance like in propagating growth in hematopoietic progenitor cells and genetic information transfer (49), stimulate tissue repair (50) and blood coagulation by activating platelets (51). EVs are classified based on their biogenesis and cellular origin. Based on biogenesis, EVs can be exosomes, microvesicles and apoptotic (52) bodies. For the scope of this thesis, I will briefly talk about different EVs classes. In the late 70s, membrane-bound vesicles were discovered in the prostatic fluid (53-55) and the name exosomes were coined by RM Johnstone in 1987 (56). Exosomes are derived from endocytic pathways (57) and they vary in size from 30 to 200 nm in diameter. On the other hand, microvesicles are budded off from the plasma membrane (58) and
their size may be up to 2,000 nm in diameter. Recently, various tissue specific and cell type exosomes and microvesicles have been described. Examples of those include (a) prostatosomes, which are nano-sized microvesicles that are secreted by the acinar epithelial cells in prostate gland: they function as intercellular communication between the cells of the acinar cells of the prostate gland and the spermatozoa (59), (b) Cardiosomes are exosomes and/or microvesicles from the cardiomyocytes (60), and (c) Vexosomes are associated with adenovirus vectors and can be exosomes or microvesicles; with proposed function being good delivery tool (61). Additional types included Ectosomes (which are vesicles secreted from monocytes and neutrophils), microparticles (shed from platelets in endothelial cells in blood), and Tolerosomes (are vesicles that are purified from the serum of antigen-fed mice). Some of these tissue and cell specific EVs are emerging biomarker targets (62). In cancer cells during apoptosis, many vesicles are released from the cells. In prostate and ovarian cancer, small vesicles released from the organ and they share specific and similar signatures to their tissue, and they can reveal the original tumor cell (63, 64).

1.3 Biomarkers

In 1844, there was an accident by Alexander McBean, a London grocer, while on vacation. He had fallen in a cave and immediately felt as if something had given way in his chest, and he was unable to stir and in extreme pain; his physician, William Macintyre, recorded this observation (65). He was diagnosed and was treated for myeloma but in 1846, he had a relapse from myeloma, was functioning well but in extreme pain, and he later died. An autopsy was carried out by Dr. Macintyre and in the presence of Dr. Thomas Watson, and he described McBean’s bone marrow as “blood-red and gelatiniform” and microscopy coherent with plasma cell qualities. An additional test was done (examination of the physicals and chemical properties of the urine) revealed that he died from “atrophy from albuminuria.” The two physicians sent his samples collected during the fall and autopsy results to Dr. Bence Jones (also know as the “father of clinical chemistry”) at the St. Johns hospital in London who was interested in chemical experiments on albumin in urine. In 1847, he published the study (66) based on McBean’s urine sample linking the presence of proteins in urine to myeloma (proteinuria detection in multiple myeloma).

The term biomarker was used and published in 1980 by Paone and colleagues in a study where they showed that serum galactosyl transferase could be a potential marker to follow up treatment for cancer and breast cancer recurrence (67). The NIH in 1998 suggested the definition of a biomarker as “a characteristic that is objectively measured and evaluated as an indication of a normal biologic process, a pathogenic process, or a pharmacologic re-
response to a therapeutic intervention.” More than ten decades later Dr. Bence Jones’ discovery led Korngold and Lipari to identify and characterize the kappa and lambda free light chain (FLS) (68), which was later approved by the FDA in 2001 as diagnostic, prognostic, monitoring biomarker for diseases.

For the rest of this thesis, I will refer to biomarker as a molecular biomarker rather than the physiological (e.g. heart rate in cardiac arrest or body temperature in high fevers) or physical (change in eye color like in yellow fever) aspects of the definition. Biomarkers are used at different clinical stages and settings.

In this section of the thesis, I will give a brief overview of the different category of biomarkers, the biomolecules highly explored, the clinical relevance, opportunities, challenges and the market status. Keep in mind that the focus of this thesis is to develop molecular tools for protein biomarkers. Other forms of disease indicators include DNA, RNA, extracellular vesicles, and metabolites.

To shed some light on a first category (diagnostic, prognostic and predictive) of biomarkers, I will use this case study. With a one-month history of difficulty speaking and imbalance, a 65-year-old woman went to the hospital. Two years earlier, immunohistochemistry analysis had revealed E-cadherin, progesterone receptors (PR) and estrogen receptors (ER) were positive, and so was her human epidermal growth factor 2 (Her2) levels by fluorescent in situ hybridization (FISH) on breast tissue. ER is a diagnostic marker for metastatic breast carcinoma because it only expressed in breast tissue. E-cadherin, on the other hand is a protein that is expressed in adherens junctions in epithelial tissue and it is a diagnostic marker for metastatic carcinoma (69). In normal brain tissue and or primary brain tumors, there is no expression of E-cadherin; hence the presence of this protein indicated an external epithelial primary site (70, 71). Diagnostic biomarkers help to detect and identify the disease state and stage. This woman was diagnosed with metastatic breast cancer and solitary brain lesion that was malignant. These two biomarkers due to their biology are also significant prognostic markers. A prognostic marker indicates diseases outcome and is used in clinical trials to stratify patients for treatment but they do not necessarily predict the response to the treatment. Certain prognostic markers indicate favorable outcomes and others do not. In her case, she had a positive ER and PR, and these are associated with a favorable outcome as these group of patients have a lower mortality rate after diagnosis (72), not taking into account other confounding factors like age, diseases state, race, tumor grade and so on. She had treatment with surgery, chemotherapy, and radiotherapy. To access the efficacy of the treatment, predictive biomarkers are used. With her being positive for ER and Her2, her follow-up treatment with chemotherapy included Herceptin and Tamoxifen (73, 74). Predictive (or responsive) biomarkers are used to provide information on the effect of a particular
treatment and facilitates targeted therapy (75). A positive ER and Her2 overexpression are favorable predictive biomarkers because studies show that such cancers are sensitive to Tamoxifen and Herceptin (74). As discussed above, some biomarkers that act as diagnostic, prognostic and predictive markers like the ER protein.

Other biomarkers used includes (i) safety biomarkers (76) used predominantly in the preclinical toxicological validation of new drugs to predict and monitor the early onset of drug toxicity. (ii) Efficacy biomarkers, (iii) Pharmacodynamics biomarkers are used to characterize pharmacology models to demonstrate pharmacokinetic and pharmacodynamics relation with a drug under development, (iv) Surrogate biomarkers are defined as “laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions, or survives and is expected to predict the effect of the therapy” (77). Surrogate biomarkers are used in the early phase of drug development to obtain information that may be critical for further development (78) and (v) Validation biomarkers mostly used in drug development.

Biomarkers market reports projects growth of about 13.8% over the next decade. These projections arise from continuous growth of $29.3 billion to $53.34 billion from 2016 to 2021. The trends include an increase in the number of cancers, growing investments in R&D and the increasing usage of biomarkers in the pharmaceutical industry (Biomarkers Market Analysis and Trends - Product, Type, Disease Indication, and Application - Forecast to 2025 Report, Oct 2016. ID3951909).

Certain guidelines are put in place monitor and validate the characterization, detection, quantification, and analyzes of biomarkers. There are different regulatory institutions like the FDA, via the Good Laboratory Practice (GLP), the International Conference on Harmonization (ICH) and the ISO/IEC 17025.

1.4 Analytical guidelines and tests of method validation

As the title of this thesis suggests, I set out to develop new molecular tools that could improve the depth at which we detect proteins. While there are many methods available, there are some guidelines and terminologies that all molecular tools adhere. In this part of the thesis, I will briefly describe those terminologies, and my hope is that it will be helpful when we progress into the projects mentioned in this thesis. Figure 1.4 illustrates the observed standard curve for protein detection and some analytical properties.

IHC defines Specificity/Selectivity as “the ability to assess the analyte unequivocally in the presence of components which may be expected to be present. Typically this might include impurities, degradants, matrix, and so
forth.” In simple terms, specificity is the ability of a molecular tool to recognize all negative samples as negative and positive samples as positive. **Sensitivity** is the assays ability to measure the actual signal. However, depending on applications, sensitivity in the molecular sense refers to the capability of the assay to measure minute amount of analytes, while in clinical sensitivity, the ability to detect and identify all positive events. Three different means are used to characterize sensitivity; (i) how little molecules can be detected (ii) how small changes can be detected and (iii) how many of the true positives identified.

![Figure 1.4: Schematic illustration of a real protein detection standard curve](image)

Increase in protein concentration ideally should correspond to an increase in detection signal. However, there is background noise that arises from non-specific binding of detection reagents to surfaces or aggregations, non-specific adsorption to media and or cross reactivity to non-target analytes. A and B represents the analytical detection range (dynamic range) which is includes the lower and upper detection limits. A also represents the sensitivity of the assay. Also, above some level of input no further increase of signal is seen.

ICH defines **Precision** as “the precision of an analytical procedure is defined as the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling of the same homogeneous sample under the prescribed conditions.” Precision may be considered at three levels: repeatability, intermediate precision and reproducibility.

ICH defines **Accuracy** as “the closeness of agreement between the conventional true value or an accepted reference value and the value found.”
ICH defines **Linearity** as “an analytical procedure as its ability (within a given range) to obtain test results that are directly proportional to the concentration (amount) of analyte in the sample.”

ICH defines **Range** as “the interval from the upper to the lower concentration (amounts) of analyte in the sample (including these concentrations) for which it has been demonstrated that the analytical procedure has a suitable level of precision, accuracy and linearity.” In simple terms, the range is the ratio between the highest and the lowest measurement within the linear phase of the curve.

ICH defines **Limit of detection (LOD)** as “the lowest amount of analyte in a sample, which can be detected but not necessarily quantitated as an exact value.” Based on the level of significance in the assay, the LOD for a particular analyte can be calculated as such that the concentration of an analyte that corresponds to a signal that is three standard deviation above the background signal.

ICH defines **Limit of quantification (LOQ)** as “the lowest amount of analyte in a sample, which can be quantitatively determined with suitable precision and accuracy.”

ICH defines **Robustness** as “a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters. It provides an indication of the procedure’s reliability during normal usage.”

**Coefficient of variation** is the level of spread or variation between experiments, its calculated as the standard deviation by the mean and it has no units.
For molecules to be detected, there needs to be a tool that does the detection in a specific manner. In 1959, Yalow and Berson reported (79) for the first time the use of antibodies for the detection of human insulin. Antibodies have since then played a pivotal role as affinity reagents in *in vitro* and *in vivo* experiments in both academia and industry. Three terms usually associated with affinity reagents includes affinity, selectivity, and specificity. Affinity in the biological context refers to the strength of binding of two molecules A and B. Selectivity is the ability of an affinity reagent to bind a molecule (e.g. A) over other molecules (e.g. B or C or D) in the systems. Finally, specificity is one of the most widely used terms in this field refers to the ability of an affinity reagent to bind to its target molecule with zero cross-reactivity towards non-target molecules.

The interaction between A and B are influenced by parameters such as temperature and pH but the basic system is as represented below:

\[
[A] + [B] \rightleftharpoons [AB]
\]

\[
K_{eq} = K_A/K_D = [AB]/[A][B]
\]

Where [A] and [B] are concentrations of affinity reagent (A) and target (B) molecule in equilibrium with [AB]. \(K_A\) and \(K_D\) are associations, and dissociations constants of A and B. The dissociation constant \(K_D\) has units M, and it is the preferred measured to the affinity of affinity reagents. The common range for dissociations is in the micro- to nano-molar scale, however stronger interactions have \(K_D\) in the pico- to the femtomolar range. The different tools to measure affinity include but are not limited to Surface Plasmon Resonance (SPR) and Isothermal Titration Calorimetry (ITC) (80, 81).

In this part of this thesis, I will give an overview of the different affinity reagents, their applications, the opportunities, challenges, and applications in life sciences with focus on antibodies since that is what has been used in this work.
2.1 Affinity reagents

*Antibodies* are the most widely used affinity reagents used in the life sciences. They are naturally occurring biomolecules known as immunoglobulins that are produced by B cells to protect the body against foreign pathogens. They are large biomolecules made up of four polypeptide chain, having a molecular weight of 150 kDa. The are made of two light (L) chains (containing about 220 amino acid) and two identical heavy (H) chains (containing about 440 amino acid), which are linked to each other covalently by disulphide bonds forming a Y-shaped structure. Each of the four polypeptide chains is made up of the variable (V) and constant (C) regions; VL and VH confer specificity to the antibody and make up the antigen-binding site and bestows antibody avidity. The hinge region of the heavy chains ensures good antigen binding flexibility for binding. There are five classes of immunoglobulins in humans IgG, IgA, IgD, IgM, and IgE. IgG and IgA have subclasses, which are classified based on the unique sequence of the hinge and Fc region (Figure 2.1)

![Figure 2.1: Structure of an antibody.](image)

An IgG antibody consists of four polypeptides chain, two heavy (denoted C_H, V_H) and two light chains (denoted C_L, V_L) and both chains are linked by disulphide bond. The terminal of both chains have variable antigen determining regions. The tail end of the antibody is called the Fc fragment (or effector region) and consists of portions of the heavy chains. There are five different types of heavy chain (α, δ, ε, γ and μ), which determine the class of IgG (IgA, IgD, IgE, IgD and IgM). There are two type of light chains κ and λ.
In 1984, Köhler, Milstein, and Jerne won the Nobel Prize. In 1975, Köhler and Milstein reported a hybridoma technology that could be used to produce mouse monoclonal antibodies (82). Fusing myeloma cells with spleen cells from mice immunized with the target antigen gave rise to hybridoma cells that produce monoclonal antibodies with the desired specificity. Each of these hybridoma cells produces only one antibody, which were purified from their supernatant with all the antibodies directed against the same specific epitope on the antigen (82). On the other hand, polyclonal antibodies are produced via immunization animals such as mouse, rabbit, goat, sheep, and donkey with an antigen (83). The serum collected after the inoculation contains a mixture of IgGs that are produced by different B cell. Some of which recognize different parts of the antigen (84). The polyclonal antibodies are then purified using specific antigen immobilized on affinity chromatography matrices to produce specific antibodies (85) or using protein A/G columns (86). Monoclonal and polyclonal antibodies are widely used in basic research and diagnostics. However, when compared to polyclonal antibodies, preparation of monoclonal antibodies requires extra expertise, cumbersome work and long times (87) (4 – 8 weeks for polyclonal compared to 3 – 6 months for monoclonal antibodies production). Details on the different applications of antibodies will be discussed in the technologies section of this thesis.

The above-described affinity reagents are based on a natural selection systems and they have some limitations. In the in vivo selection systems, the antigen used to immunize animals cannot be pathogenic, toxic antigens (like drugs), unstable proteins or highly conserved proteins such as histones (88). All these constraints have led to the development of alternative affinity reagent with in vitro systems. Progress in recombinant affinity reagents has resulted in the development of platforms whereby high throughput recombinant affinity reagents are generated with distinctive characteristics such as library sizes, means of selection, and classes of reagents. These affinity reagents can be modified to include tags that will facilitate modifications, purification and they are usually small in size. Variants of recombinant antibodies include the antigen-binding fragments (Fab), single-chain variable fragment (scFv), Nanobodies and Yumabs. These antibody fragments produced by introducing genes into vectors for in vitro display systems that encode the V domain from the heavy and light chain of an antibody. These two domains can then be joined by disulphide bonds to form the antigen binding fragments or joined with an oligopeptide linker to form single-chain variable fragment (89-91) Another class of recombinant antibody fragment is a single monomeric variable antibody domain called nanobodies. They are about 12-14 kDa in size and are produced in camelids whose antibodies lack the light chain (92, 93). Alternatively, recombinant antibody fragments can be expressed in a form that preserves some of the natural characteristics of the antibody. Expressing the Fc region of an antibody from a certain species and
merging it with the ScFv fragments that are selected by an in vitro system, ScFv-Fc (94) fragments are created that are called Yumabs (http://yumab.com).

Though there are natural antibodies and antibodies fragments that are widely used in research and industries; there is an expanding range of alternatives that are non-immunoglobulin derived affinity reagents. These are protein scaffolds such as DARPinPs, affibodies, and anticalins. The designed ankyrin repeat proteins (DARPinPs) are about 14 kDa in size, are they are derived from natural ankyrin protein consisting of four to five repeats of motifs of these proteins (95). The ankyrin proteins are involved in different biological processes such as inflammation and cell signaling. Affibodies, on the other hand are 6 kDa size scaffold based on the Z domain from staphylococcal protein A (96). DARPinPs and nanobodies are very stable compared to other affinity reagents. Another class of non-antibody affinity reagents is the nucleic acid based affinity reagents like aptamers and SOMAmers. Aptamers are single-stranded DNA or RNA binders (97) while SOMAmers are so-called slow off rate-modified aptamers, which are generated via a selection process based on the slow off rate on target antigens, and using chemically modified nucleotides (98, 99). These new alternatives provide opportunities whereby the need for animal immunization is avoided. With all this advances, there is still a great need to harmonize, characterize, and validate the existing affinity reagents for research.

In this regard, there is ongoing efforts and consortium with the aim to systematically produce affinity reagents against all proteins encoded by the 20,500 human genes, build databases, and portals to catalog the characterized affinity reagents. These programs include the Human Protein Project (HPP) by the Human Protein Organization (HUPO; https://www.hupo.org), AFFINOMICS Project (http://www.affinomics.org), ProteomeBinder (http://www.proteomebinders.org), NIH Protein Capture Reagent Program (https://proteincapture.org), and Human Protein Atlas (http://www.protein-atlas.org).

Antibody application in therapy has evolved immensely in the last decades with novel super specific antibodies used in treatments alone and in combination with chemotherapy as well. In therapeutics, affinity reagents have been successfully used in therapy for treatment of many cancers such as the FDA approved alemtuzumab, Trastuzumab, Ibritumomab tiuxetan Ipilimumab (100-102) and Nivolumab (103) for melanoma and bladder cancer respectively, which are monoclonal antibodies. RNA aptamers Macugen™ approved (104) by the FDA in the treatment of age-related macular degeneration. Other potential therapeutic reagents include Yumabs, DARPinPs and nanobodies (105, 106).
2.2 Labeling of affinity reagents

Labeling of affinity reagent has increased tremendously with antibody-based applications in research, diagnostics and clinical applications. From the discovery of electrophoresis for protein separation by Arne Tiselius in 1937 to setting up the first radio-immunosorbent technique for antigen quantification by Wilde and Porath in 1966, labeled of antibodies via different methods. 80 years later, labeling of affinity reagents is as relevant in studying of protein biochemistry, pharmacokinetics, protein localization and expression and protein-protein interaction. In this section, I will briefly discuss the different labeling methods for proteins and antibodies with the focus on antibody labeling.

For better understanding, I will divide the labeling types in vivo and in vitro chemical labeling.

The most common labeling approach for antibodies and proteins in vitro is via primary amino groups in lysine residues or N-termini of the protein sequences. Small molecules called N-hydroxysuccinimide ester (NHS) serves as a linker between the antibody and the molecules that need to be attached. When performing this labeling method, care should be taken as the NHS-ester readily hydrolyzes under alkaline conditions. Examples of other NHS-ester based coupling reactions includes attaching a biotin or a fluorophore to an amine group via biotin- and fluorophore – NHS-ester. Papers I, II and IV in this thesis use this type of conjugation to couple oligonucleotide to primary and or secondary antibodies. Carboxylic groups in side chains of aspartic or glutamic acid in the Fc region of antibodies or at their C-termini can also be labeled by using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), which activates the COOH to allow for coupling of the desired molecule. In the case of antibody labeling, this method is more suitable as it allows for the preservation of the antigen-binding site of the antibody thereby preventing reduced specificity and efficiency of binding. Another frequently used labeling method is a thiol or sulphydryl coupling. This type of labeling is specific to label cysteine side chain in proteins and antibodies. Here, succinimidyl-4-[N-maleimidomethyl]cyclohexane-1-carboxylate (SMCC) acts a heterobifunctional crosslinker. The cysteine-containing (16) biomolecule is reduced with DTT or another reducing agent, which allows for the coupling to the reactive maleimide-labeled derivative to form a thioester linkage. This technique though widely used, is the most unfavorable for labeling antibodies and proteins. In antibodies, both the heavy and light chain of the antibody is linked via S-S bond and reducing the antibody may cause instability via the hinge region also disruption of the Fab region of the antibody thereby reducing affinity and avidity. Paper III employ the succinimidyl 6-hydrazinonicotinate acetone hydrazone (SANH) coupling chemistry to couple oligos to secondary antibodies.
The other labeling approach is *in vivo*. With the increase in recombinant affinity reagents, this mode of labeling has become a standard. Site-directed labeling is carried out during the production of reagents via recombinant DNA technologies like the fusion of targeted gene to protein in different expression systems. Examples of *in vivo* labeling including *in vivo* coupling of biotin to lysine in the biotin carboxyl-carrier protein subunit of acetyl-CoA carboxylase in E coli using the endogenous enzymes BirA (107), Sortase tags (108) and a wide variety of other tags.
3. Proteomics

3.1 Proteomics

Understanding and exploration of proteins dates back to the days of Louis Pasteur in the middle of 1800s; he was a chemist but also interested in the basic question of understanding what was unique about the chemistry of living systems and their involvement in the treatment of human diseases. Marc Wilkins only devised the term proteome in 1994 and defined it as the study of protein on a large scale. The term PROTEOME was explained by Wilkins as PROTEin expressed in the genOME via RNA (Figure 1.1). However, there is no one to one relationship between the proteome and the genome. Besides, the terminologies involved at the level of the proteome is intrinsically far more complex than in the genome from the basic alphabet (20 amino acid compared to 4 nucleotides for DNA) to genes that can be spliced forming a plethora of protein products. Hence, the study of proteins can be categorized (Figure 3.1) to include; approaches to identifying and measuring their relative abundances, identifying and characterizing protein-protein interactions, identifying and characterizing post translations modification, analyzing protein function and localizations, signaling pathways, detection and measurement in different sample types just to name a few.

Figure 3.1: Type of proteomics, applications in biology and technologies
Looking back to Pasteur’s interest, this thesis is focus in developing molecular tools to detect and measure proteins in a complex biological material that facilitates diseases diagnosis and future treatment.

3.2 Looking into organelles

Other techniques developed for studying protein localization, expressions, posttranslational modifications (PTMs) include yeast 2-hybrid, Western blot (WB), Immunofluorescence (IF), immunohistochemistry (IHC) and in situ PLA (will be discussed later). The yeast 2-hybrid (Y2H) is techniques used for studying protein-protein interaction, was developed by Fields and Song in 1989 (109). In Western blot, proteins blotted from an electrophoresis gel into a membrane. WB was first described by Burnett in 1981 as the tool to analyze and detect protein(110). WB can be used in a direct mode whereby the protein is detected directly with an antibody linked to a fluorophore or enzymes to create the signal whereas an indirect WB utilizes a secondary antibody that carries the detection moiety and is directed against the primary antibody. Advantages of direct WB include shorter time, less cross-reactivity from a secondary antibody, can multiplex while indirect WB includes signal amplification by requiring secondary, no labeling of primary antibody hence less distortion of Fab and one secondary antibody can be used against many primaries. Drawbacks with direct WB include, primary antibodies are expensive, and labeling may reduce immunoreactivity, low signal amplification while increase cross-reactive and additional step in indirect WB. Immunofluorescence and immunohistochemistry are usually used to investigate endogenous proteins in situ that is in cells and tissue (111). Both technologies utilize as affinity reagents mainly antibodies conjugated to fluorophore or enzymes (112) and the signal generated is visualized by confocal or wide-field microscopy. The signal produced by these methods depends on the strength of the specific fluorophore signal compared to the auto fluorescence arising from the background of the sample. IF and IHC is applied in numerous clinical applications (113).

Flow cytometry is another powerful tool that measures multiple (up to 20) physical parameters in a single cell, or other particles such as a microorganism, virus, chromosome or nuclei by allowing them to travel past a light source in the fluid stream. Also, it provides a great opportunity in diagnostics in a diverse majority of cancer and blood-related diseases with lots of molecular and cellular information within a given cell. Fulwyler reported the first flow cytometer in 1965 when he demonstrated the separation of cells by volume (114). The flow cytometry principle is based on its ability to measure different characteristics and fluorescence of the cells. A computer visualizes the signals and displays the intensity of each cell as a dot plot or histogram. Light reaching the cell diffracts into a pattern (scattering) that is de-
tected. This scattering of light reflects different properties of the cell such as size, shape, the presence of a nucleus, membrane, and granularity, and so on. Two type of scattering light detected; the forward (FS), and side (SS) scatter. The FS range of light scattering is about 0.5 – 10°, and it is proportional to the cell size. SS has a scattering range of about 90° and is affected by the shape and complexity of the cell. Herzenberg et al. presented the first fluorescence-activating cell sorter (FACS) to sort and identify fetal cells in the maternal system (115, 116). In FACS, the cells flow through a microfluidic droplet system where the size determined by electrical (positive or negative) charge, pre-set on the computer. When fluorescence labeled probes are used, they are excited, and they emit at longer wavelengths that are detected. In paper II, we used flow cytometry as the readout of the cleaved exosomes with rolling circle amplification products that were labeled with a different fluorophore. One of the many advantages of flow cytometry is its multi-parametric function, which makes it possible to look at a subpopulation of cells in a heterogeneous population a vital tool in cancer diagnostics. Flow cytometry is greatly used in research for detection of DNA damage, protein expression cell viability and molecular biology, immunology and pathology. It is a powerful instrument in medicine with FDA approved usage in hematology, tumor immunology and many others.
4 Historical, current and emerging proteomic technologies

4.1 History; an analytical approach

Detection, separation, and characterization of protein have grown significantly from the days of immunochemistry by SA Arrhenius in 1859 to sequential technological progressions to the invention of ultracentrifugation by T Svedberg in 1924 and electrophoresis by AW Tiselius in 1937. These men paved the way for more advanced analytical technologies for protein separation and quantification. UK Laemmli was the first in 1970 to describe protein separation in polyacrylamide gels (117) based on the molecular mass using the one-dimension electrophoresis (1-DE). Though a pretty simple, reliable and reproducible method, it had limitations in that the protein to be characterized were first purified and could only resolve protein with a mass of 10 – 300 kDa. O’Farrell developed an enhanced protein separation technique called the 2-dimension electrophoresis (2-DE) where the protein is separated based on the isoelectric point and molecular mass. He could resolve with high sensitivity about 5000 different proteins from E. coli (118). This work was built on work done by AW Tiselius (electrophoresis), UK Laemmli with the 1-DE (117), isoelectric focusing and gel gradient electrophoresis by KG Kendrick and J Margolis (119) and the stacking system with SDS described in 1964 by L. Ornstein and BJ Davis (120, 121). With the 2D technique, protein modified in vivo systems from any biological system could be resolved. This lead to the analysis and the separation of proteins in other animals such as work done by J Klose. He mapped proteins in serum in mouse tissue and was the first to point out the relevance of 2-DE in identifying mutations (122) and GA Scheele who using the exocrine pancreas from guinea pig separated about 19 distinct with a molecular weight that was larger than 10 kDa in 1974 (123). Although the 2-DE is still of importance today, there are some shortcomings to the technology. It is labor intensive, hydrophilic or hydrophobic (and larger) proteins with 3>pH<10 cannot enter the gel for the first dimensions, different solubilization (pH gradient conditions) of the protein (124), low copy number proteins are not properly represented and also the dynamic range of protein could not be covered (125). With the above limitations peculiar to 2-DE, alternative protein separation and quantifications technologies were developed. Alternative advancement
to the existing tools geared towards improving the sensitivity of identifying the proteins that were separated by the gel. Edman sequencing developed in 1949 by P Edman (126) was the earliest protein-sequencing tool to determine the amino acid composition in intact proteins or enzymatic degradation (127). It was a powerful tool, which provided the linked between the separated and purified proteins with their amino acid sequence compositions of the gene of origin via the N-terminal sequence of the protein. Edman sequencing through a reliable and an automated protein sequencing technology had low sensitivity in that it could mostly sequence relatively high abundance proteins and with the emergency of mass spectrometric methods, it became obsolete.

4.2 Overview on current non-targeted proteomic technologies

Mass spectrometry is one of the greatest tools in proteomic specifically in biomarkers discovery and validation. Flashback to 1886, E Goldstein a German physicist observed that the rays in gas discharged under low pressure travelled in the opposite positive direction (from anode) to the negatively charged cathode ray. In 1898, W Wien, showed that the rays could deflect strong electric and magnetic field and that the mass-to-charge ratio of the particles have opposite polarity. JJ Thomson a British chemist in 1897 demonstrated that the cathode rays where made up of unknown negatively charged particles and he went on to measure the charge-to-mass ratio of these molecules and he created the first mass spectrograph. AJ Dempster (1918) and FW Aston (1919) developed the mass spectrometer, which allowed for identification of the mass and isotopic composition of elements in samples and was a great tool to physicist identifying isotopes. Up on to 1980, MS was a great analytical power horse tool to chemist and physicist and in 1984, JB Fenn & colleagues demonstrated the use of electrospray ionization (ESI; the molecules are dissolved and analyzed in solvent) for peptide and protein fragmentation via ionization without the need for excessive fragmentation. That earned Fenn a shared Nobel Prize in Chemistry in 2002 with K. Tanaka. F. Hillenkamp, M. Karas and colleagues in 1985 described matrix-assisted laser adsorption ionization (MALDI; the peptide are mixed & adsorbed in a matrix and a short laser is directed towards the surface to ionize both matrix & sample). Other ionization methods include ion trap, soft laser desorption and SELDI. Because MS is not the main focus of this thesis, I will give a brief overview on the technology. Herein, I have decided to categorize MS into non-targeted & targeted. Non-targeted MS is mainly attractive for unbiased identification of protein in discovery-designed studies. Samples for MS are first treated with enzymes to generate short peptides
before different fractionation is done followed by MS. Here MALDI or ESI is employed on the samples to generate ionized molecules without destroying them, which are then detected in the mass spectrometer. Non-targeted MS has low throughput, high abundant protein mask the low abundant proteins (unless an enrichment or depletion step is incorporated before peptides are produced) and the sensitivity is in the µg/mL to mg/mL range. On the other hand, targeted MS provides a more sensitive and quantitative detection and analysis of protein and is widely used in biomedicine. Here, the targeted molecules are selected prior to MS. The most commonly utilized MS is the selective reaction monitoring (SRM). Here, the masses are first filtered via the triple quadruple mass spectrometer followed by the selection of the specific analytes as ionized fragments or molecular ions. There is also the multiple reaction monitoring (MRM), Stable Isotope Standard and Capture by Anti-Peptide Antibodies (SISCAPA) where specific peptides are enriched and capture with antibodies before MS. Over the years, MS achieves high specificity and has demonstrated high multiplexing power, with this potential, reproducibility (arise from the numerous steps involved from fractionation to identification of proteins) is still a problem.

4.3 Overview of targeted proteomics technologies

In this section of the thesis, I will look at the existing targeted affinity based technologies with this thought in mind, how are proteins identified and reported? Also, I will take you briefly down the memory lane of immunoassay development history, assay formats, assay types, limitations and challenges, commercially available immunoassays and I will focus more on the proximity DNA assisted assays. Figure 4.3.1 illustrates the timeline of affinity-based immunoassays and some properties.
Figure 4.3.1: Affinity-based immunoassays formats development timeline and some properties

The table below demonstrates (a) effects of cross reactivity (b) capability of looking at multiple targets (c) background signal effect from non-specific adsorption to reagents and surfaces and (d) Increased specificity and sensitivity from additional proofreading.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplexing</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
</tr>
<tr>
<td>Non-specific adsorption</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
</tr>
<tr>
<td>Additional proofreading</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
<td>+++++</td>
</tr>
</tbody>
</table>

In 1959, Yalow and Berson described the detection of insulin using a competitive radioactive radioisotope antigen in plasma, which they called the radioimmunoassay (RIA) (79). In 1967, Wilde and colleagues described the detection of IgE using a radioimmunoassay method they called radioallergosorbent test (128). Miles and Hales in 1968 modified the technique by using a radioactive isotope-labeled antibody to measure insulin in plasma. In 1971, Engvall and Perlmann demonstrated that using alkaline phosphatase as a reporter molecule, they could detect IgG in serum and the enzyme-linked immunosorbent assay (ELISA) was born (129). This assay was performed using a single antibody against the target. In 1981, Uotila et al. described the first sandwich ELISA for alpha-fetoprotein with a capture monoclonal antibody immobilized and detection antibody coupled to an enzyme (130). Roger Ekins in 1989 proposed the ambient analytes immunoassay which only meant that the immobilized molecule on micro spots would exhibit higher signal intensities by area thereby reducing the signal to noise ratio (131). George Feinberg and colleagues did early work on microspots-based immunoassays in the 60s for the diagnosis of autoimmune diseases and proposed its importance to diagnostics. In the years after the sequencing of the human genome, DNA microarray technology was developed and used extensively and profoundly in genomic research. DNA microarray is now an established and reliable method for analyzing gene expression profiling (132), and FDA approved for clinical testing (AmpliChip CYP450 microarray by Roche). Protein analysis using this assay format is still lagging behind even though cellular information of the protein is more important for phenotypic characterization. Moreover, the rise in the use of planar or microarray format for protein only came post genome-wide study. This type of assay format is out
of the scope of this thesis, but I will mention some characteristics and applications of the system. The early use of microarray for protein demonstrated in the early 2000s by Macbeath and Schreiber (133) where they showed protein-protein interaction, identified small molecules target for protein and substrate for protein kinases. This paved the ways for more protein microarray by Haab and colleagues (134) where they used the robot to spot their antibodies on microarrays and quantified over 100 proteins in a complex matrix. Over the years many research groups have used the microarray format to characterize hundreds of proteins using different surfaces. In protein microarray, surfaces for coupling chemistries include adsorption, covalent and affinity binding. In comparing the success of DNA microarray to protein microarray, there is still a long way to go for proteins. The main problem associated with the stability of the protein (their hydrophobic and hydrophilic nature), the functionality of the proteins after immobilization (which depends on their tertiary and quaternary structures) and the interactions in different media (ionic, electrostatic interactions, Van der Waals, and hydrogen bond). Antibodies and or proteins are spotted on arrays (usually glass slides), using different chemistries with the sole aim to have maximum binding irrespective of how they bind (specific or non-specific), and a blocking step is always carried out. Storage of the arrays is a huge issue as maintaining antibody functionality is of importance. Antibody microarray has been applied in autoimmunity, allergy, some cancers, proteins expressions and signaling (135). Commercially available proteomic microarrays included ProtArray®

Human Protein Microarray (https://www.thermofisher.com), IMMray™ (http://immunovia.com) and Proteome Profiler™ Antibody Arrays (https://www.rndsystems.com) just to name a few. Another immunoassay format is the bead-based immunoassays also know as suspension assays. In this assay format, affinity reagents immobilized on microparticles that are later used to capture targets allowing for washes. Just like microarrays, the microspheres can be polystyrene (non-magnetic) and paramagnetic which are pre-functionalized with different chemistries and reactive groups. The most commonly used today are streptavidin, carboxylic and epoxy surfaces. The sizes of these microspheres also vary from 1 µm to 6.5 µm in diameter. For the rest this section, the terms microspheres and paramagnetic particles will be use interchangeably. In Paper IV in this thesis, we use streptavidin functionalize paramagnetic particles to develop the assay for protein detection and quantification method. The use of microspheres based immunoassay goes back to 1977 when Horan and Wheeless demonstrated its use in single cell sorting and analysis by flow cytometry (136). Subsequent work with this support system was done primarily for readout by flow cytometry. McHugh and colleagues later demonstrated the used of three sizes of polystyrene microsphere to set up an immunoassay where they immobilized antigen on the microspheres and used antibody labeled fluorophore to detect and quantify
proteins from *Candida albicans* using flow cytometry (137). The method and application described in **Paper II** take advantage of paramagnetic particles as a solid support to perform the assay and detection products with readout using flow cytometry. Alternative strategies to produce microspheres was implemented when different fluorophore was introduced into the core of the microspheres, which then allowed for multiple analytes detection via varying spectral ranges allowing distinction of 64 different particles (138). The company Luminex has now developed 100- to 500-plex immunoassays (xMAP® and FLEXMAP 3D®) using this technology. Other companies like Singulex and Quanterix have developed ultra-sensitive protein detection and quantification technologies using bead-based immunoassay systems. Comparison and evaluation of the performance of bead-based immunoassays and the traditional sandwich immunoassays (ELISAs) have been carried; with several groups consistently reporting increased assay sensitivities (139, 140). Planar microarray-based formats have some limitations such as requiring spotting instrumentation, images analysis, software, scanners, data collection and experienced users, different from those for bead-based assays. Also, to the solid support (microarray or beads) assays, *solution phase (homogenous) immunoassays* are also in play. Here there is no need for washes of unbound reagents and everything happens in an aqueous mixture, hence making it preferable over solid support based assays that require either slides or beads. In addition, there is no need to immobilize affinity reagents, which maintains high functionality of the Fab regions, which sometimes can alter during treatment with small molecules for immobilization. With all these advantages, solution phase immunoassays still have limitations. Since there are no washes involved, there is an increase in the background noise, which reduces the assay sensitivity. Several solution phase immunoassays developed for protein detection and quantification in different diseases (141-143). In **Paper IV**, we have developed a solid phase version of a solution phase assay, using paramagnetic particles to increase the performance of the assay by introducing washing steps and increasing the analytes assay volume, while also increase the numbers of antibodies needed for detection from two to three.

Perlmann and Engvall used alkaline phosphatase instead of radioactive labels for protein detection and created a new era in the diagnostics field. The safety concerns in the utilization of the radioactive substance was no longer an issue, and the analytical performance of ELISA was equivalent to that of RIA. **ELISA** is the most widely used protein detection tool in clinics, hospitals, industry and academic research for diagnostics and follow-up and is considered the gold standard. Over the years, various readout format of ELISA has been developed, in addition to the fluorescence optical density, chemiluminescence and electrochemiluminescent a commonly used. There are reverse-phase, forward-phase, direct, and indirect immunoassays (IA) (Figure 4.3.2). **Reverse-phase IA**, the antigen is first absorbed in a non-
specific manner in the plate, and the detection antibody (primary antibody coupled to a reporter molecule) is used directly on the antigen, unbound detection antibody is washed off, and the enzyme substrate is added to detect the antibody enzymes interaction. Forward-phase includes immobilization of affinity reagent, and which then captures the target antigen. Direct IA involves a capture affinity reagent on solid support. This then captures the target antigen and is detected by a primary antibody coupled to an enzyme or fluorophore, which reports the targets. In indirect IA, a capture antibody is first immobilized on the surface to capture the antigen; an unlabeled primary antibody then incubated with the captured antigen. A secondary antibody coupled to a reporter recognizes the primary antibody. Since multiple secondary antibodies can recognize a single primary antibody, there is signal amplification thereby increasing the sensitivity of detection. Most commercial ELISAs, utilize the indirect assay system where the primary antibody that recognizes the target coupled to biotin followed by streptavidin-linked to the reporter molecule. Here, binding by multiple reporters result in signal amplification and increased detection sensitivity. However, background issues still arise as detection reagents can bind non-specifically to the solid support. Commercial ELISA has sensitivity in the pg/mL (or pM) levels for analytes like IL-6 with dynamic ranges of about three orders of magnitudes, and coefficient of variations below 5%. The need to analyze molecules that are present at low concentrations is ever present, as a tool to monitor diseases and for early diagnosis. This has led to the development of ultrasensitive immunoassays. In 1979 Harris et al. reported the first ultrasensitive immunoassay for detection of cholera toxin and rotavirus using RIA (116) and Shalev et al. in 1980 reported a high-sensitive ELISA (HS-ELISA) for IgG detection in mouse (144): both in zeptomolar concentrations. In 2010, Rissin et al. demonstrated sub-femtomolar detection ranges in a single molecule detection ELISA format and used this technology they detected PSA in patients who had undergone prostatectomy. This enabled monitoring of these patients with prostate cancer for reoccurrence (145, 146). Here, magnetic particles are used to capture target protein, and individual beads confined in femtoliter wells, which allows for digital counting of signals from even single labeled antibodies in the individual wells. The company Quanterix Corporation has commercialized this SiMoA technology. Another category of emerging single molecule detection methods is the Erenna immunoassay systems by the company Singulex. Here, magnetic microparticles are also used to capture target protein, followed by detection by fluorescence-labeled antibodies. The labeled antibodies are then chemically eluted from the magnetic microparticles, and single fluorescent molecules are counted using a capillary flow system. The assay has been applied to detect troponin, a biomarker for acute myocardial infarction (147). In contrast to over the counter ELISAs where the readout is an analogy measurement of total bulk fluorescence, these single-molecule protein detection assays offer a unique ad-
vantage in that individually labeled antibodies are detected, and the precision is increased significantly if sufficient numbers of molecules are counted. Also, the background signal is minimized as even single labeled antibodies are clearly detected above signals from plates or assay medium as in the case of the SiMoA and Erenna systems. Even though these emerging technologies offer significant analytical sensitivities, there are limitations as with any sandwich IA. They have limited multiplexing abilities as cross-reactivity of detection molecules increases as the number of molecules to be detected increases (148). It is worth mentioning that although these IA are reported as single molecule assays, to achieve detection above background hundreds of protein molecules are needed in the sample. Also, since this IA utilize the same molecular architecture as regular sandwich ELISA, limitations in specificity becomes an issue as non-specific adsorption of detection reagents give rise to signals, and false positive signals may also arise from antibody pairs detecting irrelevant proteins non-specifically (4.3.1).
Figure 4.3.2: Schematic overview of affinity-based immunoassay format
(a) Represents immunoassays fluorophore conjugated to antibodies and (b) DNA-assisted immunoassays

So far, all the technologies I have explained above utilize fluorophore or enzymes directly coupled to antibodies (or other affinity reagents), which serve as detection moieties. In the next part, I will talk about DNA-based (aptamers and SOMAmers), and DNA-assisted immunoassay (Immuno polymerase chain reaction (iPCR), Immuno rolling circle amplification (iRCA), proximity ligation assay (PLA) and proximity extension assay (PEA). All the papers I, II, III, IV in this thesis irrespective of the assay format and other features mentioned above are based on the proximity DNA-assisted technology. Also, all immunoassay systems described above and subsequent discussions are targeted rather than fishing expeditions.

Aptamers and SOMAmers based immunoassays belong to the DNA-based immunoassays class of affinity reagents. Aptamers are versatile short single-stranded DNA (or RNA) oligonucleotide affinity molecules that bind to spe-
cific target molecules sometimes termed chemical antibodies. They are produced via in vitro selection from large random libraries using their ability to discriminate false targets via the systemic evolution of ligands by exponential enrichment (SELEX) system (97). SOMAmers are Slow Off-rate Modified Aptamers that are DNA aptamers with modified bases that upon binding to specific target proteins fold into a complex 3D structure (149). The specificity is ensured by the in vitro selection using large DNA libraries and SELEX to isolate reagents with slow dissociation rate. SOMAmers is the core behind the SOMAscan proteomic platform of the company SomaLogic whose aim is to set up highly multiplex protein detection and analysis assays for personalized, preventive diagnosis and treatment. This technology combines the SOMAmers molecular recognition elements and DNA microarray to read out target proteins in biological samples. SOMAscan has been used in clinical studies for diagnostic biomarker discovery in non-small lung cancer (150) among others. They possess many appealing qualities such as the size of the reagents, which is between 6-30 kDa compared to 150 kDa for antibodies; this makes them an attractive reagent for tissue penetration, they can have high affinity, and multivalent aptamers can have even higher affinity due to the selection process.

4.3.1 DNA assisted immunoassays

In 1992, Sano and colleagues developed and introduced the first DNA-assisted immunoassays using PCR as the readout, which they called immunoPCR (151) (Figure 4.3.2). Quantification of the resulting DNA amplicon was analyzed with traditional gel electrophoresis, which made it quite laborious and not so precise. The invention of the real-time PCR advanced the growth of the DNA-assisted immunoassays in the late 1990s. Other variants of this form of immunoassay were developed such as the immuno-rolling circle amplifications (152) (Figure 4.3.2). Another class of DNA immunoassay is called the bio-barcode assay developed in the Mirkin’s group. Here, target proteins are captured by magnetic nanoparticles and detected by antibodies that are co-immobilized with barcode DNA on gold nanoparticles. The DNA molecules are then released upon antibody binding and captured on DNA microarrays. The detection was done via a reduction reaction of silver ions, catalyzed by the gold particles to silver, resulting in light scattering of the developed silver spots (153).

These particular categories of immunoassays take advantages of the properties of DNA including their specificity, DNA hybridization, opportunities to amplify DNA, the flexibility of ligations, and other enzymatic capabilities and the ability to store an enormous amount of information to enable barcoding.

I will use a few sentence to explain the different advantages that DNA confers to immunoassays. DNA hybridization based on the Watson and
Crick base-pairing specific nucleic acid complementarity (G=C and A=T/U). All DNA microarray technology utilizes hybridization to capture, detect and report molecules. In 1988, Landegren and colleagues developed an assay to discriminate single nucleotide substitution by oligonucleotide ligation assay (OLA) using the substrate fidelity of T4 DNA ligase (154). Ligation is an enzymatic process whereby a phosphodiester bond formed between the 3’-OH and the 5’-PO_4 in the DNA strands. Lehman first used the reaction in the lab in 1973. DNA repair utilizes ligase and it has been used extensively in molecular biology for cloning. In 1994, Nilsson and colleagues described a process where DNA was detected by circularization of oligonucleotide probes called Padlock probes (155). These are single-stranded DNA molecules that comprise two target-complementary end sequences and a non-hybridizing backbone, which includes sequences required for detection. The target complementary arms recognize its target, and with a ligase enzyme, the two arms are joined to form a circle, which can then be amplified using rolling circle amplification (RCA). Rolling circle amplification is an isothermal linear DNA amplification process first described as rolling circle replication by Fire and Xu (156) where they demonstrated that “rolling circle synthesis in a simple enzymatic system that can produce tandem repeats of monomers as short as 34 nucleotides”. RCA has been used extensively in in situ experiments for detection of protein localization, interactions, and post-translational modification (157, 158). Banér et al. demonstrated that using Phi29 DNA polymerase, 100 nucleotides of circular DNA can produce strands containing up to 900 copies of complements of the circles in 60 min of rolling time by approximating the polymerization rate at 1500 nucleotide incorporations per minute (159). The combination of target-dependent ligation of padlocks and amplification by RCA makes them a powerfully specific and sensitive tools for molecular diagnostics (160). Also, the ability to design specific target-complementary region and utilize the non-hybridization backbone (for tagging, barcoding, and detection), makes padlock probes a powerful tool to investigate multiple analytes at the same time. Papers I, II and III makes used of this power tools in showcasing their enhanced abilities as a molecular diagnostic tool for protein detection in vitro and in vivo. Up to this point, all technologies I have discussed utilize the same molecular architecture as the traditional ELISA, which is that of dual recognition of target protein molecules. As mentioned above, even with their great analytical sensitivities among other great features, they share similar setbacks such as increased non-specific signals from assay medium, increased non-specific absorption of detection reagents and increased the cross-reactivity of detection reagents against target proteins (Figure 4.3.1).
4.3.2 Proximity-based DNA assisted immunoassay

Proximity-based DNA-assisted immunoassays include PLA and PEA, they have a different molecular architecture compared to the other IA categories (Figure 4.3.2). Fredriksson et al. first reported this assay format in 2002 when they used aptamer as PLA reagents to detect and quantify platelet growth factor (PDGF) in solutions and could detect zeptomolar concentrations (141). In this assay setup, affinity reagents (that can be pairs of monoclonal or single polyclonal antibody preparations) functionalized with DNA oligonucleotides that together can bind to different sites on target proteins. The juxtaposition of DNA-coupled affinity reagents binding the same target allows for the DNA strands to be brought close together which allows templated ligation of the single stranded DNA on the antibodies, or an extension by polymerization in the case of proximity extension assays (PEA) (161). These assays have been performed in homogenous and solid support formats in case of PLA (162). The ligation and or extension gives rise to templates that can be amplified by PCR or in the case of in situ assays via RCA (157). To address the issue of background signal from non-specific binding of detection reagents and the non-specific signal from assay medium which occurs in the traditional ELISA, signal generation in proximity-based IA only arise from amplifiable products of pairs of affinity reagents agreeing on the presence of a target protein. In addition, the prerequisite of requiring two affinity reagents to agree on a target in addition to ligation to form the amplifiable template improves the assay specificity and reduces cross-reactivity. To further enhance the specificity of PLA, the solid phase variant of the assay was developed using magnetic microparticles to act as a solid support with immobilize antibodies that then capture target proteins (163). The solid support format then allows for washing of unbound detection reagents, thereby reducing background signals. Even though the efficiency of each step in this assay is not 100%, however, triple recognition allows for more proofreading, which serves by increasing assay specificity. Advantages of this assay system, as with iPCR and iRCA is the flexibility of designing oligonucleotides for multiple target detection, the introduction of barcodes to identify different sample types and the versatility of enzymatic reaction to enable amplification. PLA has been adapted for various applications from protein visualization in situ (157, 164), protein-protein interaction (158, 165, 166), protein-DNA interactions (167), posttranslational modifications measurement of biomarkers in single (168, 169), multiple protein detection (170-172), detect infectious reagents (173) and applied to Western blot (174).

A more stringent PLA design requiring more than three antibodies have also been developed and demonstrated with triple binding (175) and quadruple antibodies (63) to generate DNA templates that can be amplified. PLA is a power technology, which allows for even the smallest amount of targets to be detected by the amplification via PCR and RCA. In Paper 1, we set out to
incorporate the PLA with RCA readout on microtiter plates with optical density readout. A variant, which can be easily, adapted to already existing fluorescence readout instruments in clinics and hospitals.

As mentioned earlier, PEA is another variant of the proximity-based assays and has been commercialized by the company Olink Proteomics. Here, they only require 1 µL or 5 µL of sample (plasma or serum) for analyzing sets of 92 proteins without any washes and with moderate background (171, 176). In Paper IV, we have developed a solid phase variant of the assay where use increased amounts of reagents and washes serve to improve assay sensitivity further. An additional capture antibody on beads will also increase assay-proofreading functions to enhance specificity. Furthermore, the adaptation and application of PLA for in situ protein detection on fixed cells and tissue section is a powerful tool, which allows for PTM studies, intercellular interactions, and localization. The Duolink assay is using PLA technology and represents a commercial product of this technology for in situ protein detection, being licensed and sold by Sigma-Aldrich. A detection efficiency of 100% of all target molecules in a sample is the ultimate goal. In in situ PLA, some factors that prevent us from achieving this ultimate aim. These include the limited efficiencies of the different enzymatic steps, oligonucleotide design systems, and the dissociation constant of the affinity reagents. In Paper III, we set out to develop a new system with better oligonucleotide design and enzymatic conditions. We call the new assay system UnFolding probes, which increase the assay efficiency of the traditional in situ PLA design.

4.4 Proximity based IA: Still lagging

After all said, proximity based immunoassays also comes with limitations. In the final part of this segment, I will address some of the concerns of proximity based IAs. In complex or dense biological milieu, the proximity-based detection system can give rise to false positive signals. This is particularly true when performing in situ experiments on tissues and cell lines. Limiting the minimum distance needed to give rise to signals can reduce this effect, and so will the concentration of the detection reagents. However, this is usually difficult when doing in situ experiments. In homogenous assay system where no washes are involved, signals can be generated by false proximity between affinity reagents irrespective of the presences of the target antigen. Here, the concentration of detection reagents, incubations times and kinetics are critical parameters to reducing false signals. In PLA experiments based on solid support, we still experience some background noise in the absence of protein. This is mainly due to sticking of PLA probes to the beads surface and or to the antibody immobilized on the solid support. Improved blocking reagents may be the remedy for the nonspecific binding of detection reagents.
to beads, but one may need to try alternative binders to counter the effect of binding of detection probes to the immobilized antibodies.

About 15 years after the first publication of the proximity ligation assay, can we now say that PLA is a mature technology that can compete with conventional or emerging technologies? Looking at the different areas of applications of PLA, one can safely say that we are almost there. Commercially available assays like PEA from Olink and Duolink from Sigma put this technology at the forefront of current tools in proteomics. With this said, my answer to the above question is NO. As mentioned above the advantages of PLA, which included enhanced specificity, sensitivity, broader dynamic range and the ease multiplexing. However, when it comes to application in routine clinical applications, precision is another important parameter.

Currently, the coefficient of variation (CV) of PLA assay is between 15 – 40%, this is far from that of approved clinical immunoassays. A significant proportion of the CV arises from the PCR which depending on the number of molecule to be detected range from 5% to 20%. Improving the precision of PLA will not only make it more suitable for clinical use, but it will significantly increase the number of detectable molecules. Previous work to address this issue has used sequencing as readout. Although sequencing is a well-established technology, the enormous amount of data coupled with additional assay steps may make this less practical for routine use. In paper I, the use of RCA instead of PCR is a step towards achieving this goal, as the precision of this isothermal form of amplification is approximately 5%. Likewise, the use of already existing and approved instrumentation makes this approach more feasible with further optimization to meet with the different regulations.

Another limiting factor that comes to mind is the issue of specificity and sensitivity. The question regarding sensitivity is, are the assays detecting and reporting all the correct target molecules present in the sample? How much of the detection signal comes from true targets? What are the limiting steps in the assay? These questions though easy to formulate theoretically, are not so easy to answer practically. For example, the disparity between the observed vs. expected results can be summed up as assay inefficiency. This affects the sensitivity (often denoted as limits of detections and or lower limits of quantification) of the assay. The specificity, another key element of PLA, manifests itself at two levels, the multiple recognitions of affinity reagents on their target and the ligation of the pair of reagents. Can this multiple target recognition pose a problem for the assay? The simple answer is yes! When using single polyclonal antibody preparations that have been divided to make pairs of detection probes carrying the two DNA arms, there is a 50% chance that antibodies carrying the same DNA sequence can be bind to the same target hence decreasing efficiency. Also, by requiring two binding events to generate a signal, the likelihood of identifying a particular target
drops. Using monoclonal antibodies or “monospecific” affinity binders can amend this problem.
Part II
5 Present Investigations

Looking back at the title of this thesis, the projects discussed here are about the development of enhanced molecular diagnostic tools for protein detection and analysis. The overall objective was to develop tools that could potentially be used in routine clinical settings to detect and quantify low abundant protein. To introduce the work reported in this thesis, Figure 5.1 will present how present investigation fits in the proteomic puzzle.

Figure 5.1: Synopsis of the papers presented in the thesis
The projects presented in thesis can be applied into different categories of the proteome. (a) Papers I (PLARCA) & IV (SP-PEA) can be applied into proteome mining for detection and analysis, (b) while paper II (ExoPLA) into structural proteomics for organelle identification and characterizations and finally (c) paper III can be applied into detecting and quantifying protein expressions, PTMs and studying protein-protein interactions.
5.1 Sensitive protein detection in microtiter plates by proximity ligation with rolling circle amplification (PLARCA) - Paper I

5.1.1 Introduction

Proteins are crucial to all cellular processes from cell structure formation, to transport of molecules in and out of cells, to biochemical processes and to determine disease state and they are for diagnostic purposes. Antibody are versatile and powerful tools for the analysis of various molecular and cellular process as well as clinical and biomedical diagnostics (177). Antibody-based detection methods have greatly advanced with the introduction of the radio immunoassay and ELISA (79, 129) respectively. Since that era, several formats of immunoassays and read out techniques have emerged for protein detection. Several versions of immunoassays have been developed have been discussed in section 4.3 of this thesis.

DNA-assisted protein detection assays have been discussed in 4.3.1. Proximity ligation assay (141) detects protein with dual detection reagent recognitions (thereby increasing the specificity) and enhanced sensitivity. There is a need to adapt such a powerful technology in current protein detection methods used in clinics and hospital laboratories. We used the microtiter plate to immobilize monoclonal and or polyclonal antibodies, and incubate overnight at 4°C on a shaker. Unbound antibody was washed off and a blocking buffer added and incubated at room temperature. The plate was washed and the target antigen was diluted in appropriate buffer was added and incubated at room temperature. Unbound antigen was washed, and the PLA probes were added and incubated at RT, upon target recognition, this pair of PLA probes guides oligonucleotide ligation, followed by amplification via RCA. The RCA products were then detected by HRP-labeled oligonucleotides, followed by TMB substrate to generate colorimetric products, that is readout with an absorbance microplate reader (Figure 5.1.1).
Figure 5.1.1: Schematic description of PLARCA.

(I) Antibodies are immobilized via hydrophobic interaction in wells of microtiter plates, followed by washes and blocking of nonspecific binding. Bound antibodies are used to capture target proteins from added samples, followed by washes to remove unbound sample components. (II) Next, a pair of PLA probes (antibody-DNA conjugates) is added that can recognize captured target molecules, followed by washes. (III) DNA oligonucleotides are then added along with a ligase allowing the oligonucleotides to be ligated into DNA circles, templated by oligonucleotides on one of the PLA probes. (IV) A DNA polymerase is added to initiate replication of the DNA circles via RCA, primed from that antibody-conjugated oligonucleotide which did not template ligation. (V) Finally, the RCA products are detected by adding HRP-labeled oligonucleotides to the RCA product, followed after washes by addition of the TMB substrate.

5.1.2 Aim of study

The objective of this project was to develop a variant of PLA that are adapted easily to hospital laboratories and clinics that use the regular ELISA for diagnostic and prognosis.

5.1.3 Summary of finding

The performance of PLARCA showed no significant difference when IL-2, IL-8, IL-6, and VEGF spiked in serum compared to buffer (Figure 5.1.2).
Figure 5.1.2: Comparison of results from assays for IL-2 and IL-8 in buffer (blue) or 10% (red), or 50% (green) chicken serum. Y-axes show optical density (OD) measured at 450 nm while the X-axes show protein concentrations in molarity (top) and weight per volume (bottom). Averages of measurements done in triplicate are shown with standard deviations indicated with error bars.

Comparison of PLARCA with traditional ELISA revealed that PLARCA detected low femtomolar concentrations in IL-4, IL-6, and GDF-15 in serum with increased dynamic range (Figure 5.1.3). Analysis and detection of IL-4 and IL-6 in colorectal cancer tissue lysates and plasma from prostate cancer patients demonstrate PLARCA detecting IL-4 and IL-6, considerably below the detection limits of traditional ELISA (Figure 5.1.4).

Figure 5.1.3: Performance of PLARCA vs. ELISA
Comparison of PLARCA (squares) and ELISA (diamonds) for measuring levels of IL-4 and GDF-15. Averages of measurements done in triplicate are shown with standard deviations indicated with error bars.
Figure 5.1.4: Measurement of cytokine concentrations in colorectal and prostate cancer. Measurement of IL-4 and of IL-6 with PLARCA and ELISA in plasma samples from 25 prostate cancer patients and 24 healthy controls. PLARCA and ELISA limits of detection were calculated from the standards in Table 2 for IL-4 and IL-6 (see paper I). The age ranges for these patients and controls are illustrated in Supplemental Table 2 (see paper I). Each dot represents the mean of a triplicate. p values were calculated using a two-sample Wilcoxon rank sum test.

5.2 Detecting individual extracellular vesicles using a multicolor in situ proximity ligation assay with flow cytometric readout - Paper II

5.2.1 Introduction
Currently, discovery and characterization of extracellular vesicles (EVs) in a biological sample is of great importance (178). Exosomes and microvesicles are the two classes that can be distinguished based on their biogenesis. Exosomes are about 40 to 200 nm in diameters and have important roles in many biological processes and have been shown to be promising biomarkers for some diseases (179). Currently, the main methods used for exosome detection and characterization is by flow cytometry (180). However, their size makes it difficult to distinguish single EVs from the background of the microfluidics of the traditional flow cytometry instruments. We identified five target proteins on the surface of EVs and using three fluorescent dyes, and we successfully distinguished specific EVs.
Here, beads were functionalized with oligonucleotides containing uracil bases, monoclonal and or polyclonal antibodies conjugated to a complementary uracil-containing oligonucleotide that hybridizes to the oligonucleotides coupled to the beads. This hybrid system allows for enzymatic cleavage with UNG/EndoIV. Antibody captures exosomes via specific surface proteins, and unbound exosomes are washed off. Four PLA probes are used to detect three different surface proteins. We designed the same common probes that recognize dipeptidyl peptidase 4 (CD26) which is present in most EVs and paired with antibodies directed against Neprilysin (CD10), Aminopeptidase N (CD13), and Cathepsin B. Unbound probes are removed by washing and a ligation mixture is added into the reaction. After ligation, enzymatic digestion of the uracil containing hybrid system is done to release the detectable moieties into solution. Rolling circle amplification is then performed in solution-phase, followed by detection with three different fluorophores coupled to oligonucleotides used for detection (Figure 5.2.1). Finally, flow cytometry is used for quantification and analysis.

Figure 5.2.1: Illustrations and assay format of ExoPLA
5.2.2 Aim of study
Develop a multicolor *in situ* PLA method for extracellular vesicles detection and quantification with flow cytometry readout

5.2.3 Summary of finding
By analyzing exosomes in human plasma, we developed and adopted a new method called ExoPLA that can detect and characterize individual exosomes using the multi-color detection using flow cytometry.

Using purified prostasomes as a model, we could use specific surface protein markers to detect them and confirm the data with fluorescence microscopy analysis (Figure 5.2.2).

The performance of the ExoPLA was compared to the conventional bead-based assay. ExoPLA method could detect as little as 200 pg/mL of prostasomes compared to 20 µg/mL using the commercial bead-based method.

By isolating extracellular vesicles from myeloid cell line U937, breast cancer cell line MCF7 and prostasomes from seminal fluid, ExoPLA could also distinguish the different EV populations

![Figure 5.2.2: Detection of Evs from cell lines.](image-url)

Detection of Evs from isolated from U937 cells in a mix with prostasomes ratio 1:1, and 3:1, and a mix between EVs from U937 cells and MCF7 cells ratio 1:1, and 3:1, using the multicolor ExoPLA assay with the selective PLA probe against CD114, only present on EVs from U937 cells. The ratios are based on the total protein concentration of the EVs. The top rows are all events and the bottom rows are gated plots.

Moreover finally, in a complex biological matrix, 10 µg/mL or 200 µg/mL of prostasomes was spiked in 10% human plasma from a healthy female. With control experiments, ExoPLA was able to distinguish these populations
5.3 Enhanced \textit{in situ} proximity ligation assays via Unfolding PLA probes - Paper III

5.3.1 Introduction

Proteins are key regulatory biomolecules that control most cellular processes through signaling networks involving posttranslational modifications and interactions. These cellular processes define disease states hence developing tools efficient for identifying this processing therefore, is vital towards understanding diseases. \textit{In situ} PLA (157) is a sensitive and specific tool for detection of these cellular processes in cells. This method enables us to visualize protein interactions, modification, and co-localization \textit{in situ} (181, 182). However, there is an efficiency limitation to the current approach. We developed a new version of the reagents that we call UnFolding probes. In this version of the assay, the efficiency and convenience of detection are improved by designing the assays such that the necessary antibody-oligonucleotide conjugates carry with them all elements required for detection via rolling circle amplification. Only after proximal binding of pairs of probes and washes is an enzyme cocktail added that allows the probes to interact to form circular templates for amplification (Figure 5.3.1).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.3.1.png}
\caption{Schematic representation of Unfold}
\end{figure}

\begin{itemize}
\item I: Immobilize antibody on solid support and capture protein
\item II: Incubate with Unfold probes
\item III: UNG/EndoIV treatment of Unfold probes
\item IV: Ligation
\item V: RCA and detection
\end{itemize}
5.3.2 Aim of study
The aim was to improve the efficiency of the *in situ* proximity ligation assay by decreasing the rolling circle amplification of non-targeted circular DNA templates.

5.3.3 Summary of finding
Unfold probes show improved signal to noise ration over *in situ* PLA. Secondly, we improved the assay efficiency for Unfold probes compared to the *in situ*. In phosphorylation experiments, we compared Unfold and *in situ* in BJ-hTert cells that were stimulated with platelets-derived growth factor-BB. Unfold detected higher number of phosphorylation compared to *in situ* (Figure 5.3.2).
Figure 5.3.2 Phosphorylation of PDGF receptor in BJ hTert cells
A) Images comparing phosphorylation levels of the PDGF receptor β in serum starved BJ hTert cells (PDGF-BB -) to cells treated with PDGF-BB for 45 min on ice (PDGF-BB +). The RCA products were labeled with Cy3 (red dots in the merged images) and the nuclei were stained with Hoechst 33342 (Scale bar (white) = 50 μm).
B) Quantification of signal comparing \textit{in situ} PLA and UnFold at different probe concentrations. (n=3, Error bar represents SEM).

We also made comparison of Unfold and \textit{in situ} PLA on microtiter with absorbance readout, Unfold showed increased signal intensity that thereby increased the lower limits of detection about 10-fold (Figure 5.3.3)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5_3_3.png}
\caption{Performance comparison of S3, ELISA and Unfold for detection of IL-6}
\end{figure}

Comparison of ELISA (blue), S3 (red) and Unfold buffer (green) in 10% chicken serum. y-axis show optical density (OD) measured at 450 nm and the x-axis show protein concentrations in weight per volume. Averages of measurements done in triplicates are shown with standard deviations indicated with error bars.
5.4.1 Introduction
Detection of protein in the blood can serve as an indicator of a biological process in humans. Early detection of certain proteins in blood is vital for early disease diagnosis and prevention. Proximity extension assay (142) is a homogenous sensitive and specific protein detection and analysis method. The current assay format utilizes a microliter sample volume, and no washing of unbound detection reagents is required. The fact that unbound detection reagents are not removed by washes increase the background noise, and hence decreases the detection threshold of the assay. The assay is set up such that antibodies of interest are immobilized on magnetic particles and incubated at room temperature. The immobilized antibody then captures the target protein and unbound antigen is washed off. Next, detection reagents are added and incubated after unbound reagents have been washed off. The extension and qPCR mix is then prepared and added to the reaction (Figure 5.4.1).

Figure 5.4.1: Schematic description of spPEA.
A. Schematic representation of magnetic bead-based PEA assay (I) Antibodies are functionalized on magnetic particles, incubated and unbound antibodies are washed (II) Functionalized antibody is then used to capture antigen from sample, incubated and free antigens are washed off (II) Pairs of PEA probes are then added to the reaction, incubated and free (and or loosely bound) probes are washed off. (IV) Enzymes required for extension and polymerization are then added to the mix, incubated, followed by PCR amplification

5.4.2 Aim of study
In this study, we developrd a solid phase format of the proximity extension assay (PEA) for the analysis and quantification of clinical material. To improve the performance of the PEA (161, 171) we used magnetic particles,
which allowed us to wash off excess detection reagents and also increase the sample volume to introduce more detectable molecules.

5.4.3 Summary of finding

We developed a protocol for protein detection on magnetic particles. We compared detection of IL-8, IL-10 and IL-6 using spPEA and PEA. SpPEA demonstrated over 2 orders of magnitudes lower detection levels compare to the homogenous form of the assay (Figure 5.4.2).

**Figure 5.4.2: Comparison of bead-based PEA vs. solution-based PEA**

Assay performance comparison of spPEA (blue), and PEA (red). Comparison of detection of (A) IL-6, (B) IL-10 and (C) TNF-alpha in assay buffer. X-axis represents the concentration in mass per volume and y-axis represent Ct. All experiments were done in duplicates and the standard deviation is shown in the arrow bars.
III Conclusions and Perspectives

What is limiting the adaptation of new molecular tools to routine clinical use? I woke up to this thought every single day since I started this journey. As of this day that I write this section of my thesis, I checked on PubMed with the words ‘immunoassay and there were 483,413 articles published from 1967 – 2017. At the same time, I searched the FDA database and found 500 approved in vitro diagnostic assays (IVDs). The economics of the proportion of failed and unapproved methods comes to mind. In the same regard, concerning the agent by which immunoassays are built on (antibodies) as my mind wandered to the comment made by Bradbury, Plückthun and 110 co-signatories in 2015 “To save millions of dollars and dramatically improve reproducibility, protein-binding reagents must be defined by their sequences and produced as recombinant proteins”. Are poor quality antibodies the only reason we only have 1 approved IVD for every 966 reported in the literature? This can be debated (since antibodies are a crucial component in immunoassays), but in my opinion, the answer is no. Many parameters contribute to poor performance, antibodies, sample collections validation, control samples, substrates, analysis methods, and instrumentation. All the above parameters have an effect on the analytical characteristics, which includes sensitivity, specificity (or selectivity), throughput, dilution linearity, reproducibility, robustness, and precision.

In the words of Marie Curie, “One never notices what has been done, one can only see what remains to be done.” Looking at how far immunotherapy has come, and the increasing use of antibodies in different types of cancers treatment, one must say and agree that there is a bright light at the end of the tunnel for immunoassays in IVDs. Hundreds of millions of dollars are wasted each year in antibodies usage both in academia and industry. Numerous reports, letters, and publications have been written with proposals on improved means for validating antibodies. Should we keep on writing proposals and never getting to a point where we arrive at an agreement on what needs to be done? As the father of medicine Hipocrates put it “Declare the past, diagnose the present, foretell the future; practice these acts. As to diseases, make a habit of two things – to help, or at least to do no harm”. In 2016, Uhlen et al. wrote, “We convened an ad hoc International Working Group for Antibody Validation to formulate the best approaches for validating antibodies used in common research applications provide guidelines that ensure antibody reproducibility. We recommend five conceptual ‘pillars’ for antibody validation to be
employed in an application-specific manner”. No one solution from the different reports will fit all; researchers from academia and industry must come together in sync to put into practice all that is said and written. Alternative recombinant affinity binders have gained applause over the years. However, the demand to validate them for immunoassays (IVDs) is not as appealing as its application in pharmaceuticals. What do we do?

When it comes to protein detection and analysis, almost everybody utilizes and adapts their assay to the sandwich ELISA format. Once the scientific research community decides to come up with specific guidelines for affinity reagents evaluation and validations, then some of the problems mentioned above with double binder immunoassay will be solved. In the meantime, DNA-assisted proximity based immunoassays come forth with some solutions on specificity, sensitivity, and throughput. In this thesis, we got to compare the new tools developed and presented herein with existing tools for detecting protein from simple to complex matrix. In Paper I, we demonstrated improved detection limits, better dynamic range over traditional ELISA and single binders DNA-assisted method. This does not end here, having the best samples to show the relevance of these assays is as important as developing them.

Which led me to my next concern. Availability of well characterized, evaluated and stored samples material. In developing immunoassays, validation is an important aspect that enables other users to implement what has been developed. Sample stability can dramatically impact the results of analytes measurement. As we are aware, downstream testing of methods is significantly affected by poor quality pre-analytical sampling. Consequently, the investigation of different storage conditions, sample tubes, anticoagulants, and centrifugation conditions, among other factors can be vital. International and European efforts like the International Society for Biological and Environmental Repositories (ISBER) and Biobanking and Bimolecular Resources Research Infrastructure (BBMRI) have put in place policies on optimizing sample collection, storage, and analysis. Data about how these different factors affect assay validation should be made available to avoid variations from these sources. With everything in place, accessibility of optimized well-characterized samples should be made available for necessary assay validations without excessive bureaucratic procedures. We ‘immunoassay developers’ are all after the same thing, which is providing a better analytical tool for clinical diagnosis, prognosis and diseases monitoring. All these factors are central in decision-making by medical practitioners in hospitals. The availability of well-validated antibodies, samples and other resources that make immunoassay successful does not necessarily ensure a sensitive, accurate or multiplex immunoassay. As discussed in section 1.2, leakage of a few proteins from damaged tissues or proteins secreted from endocrine epithelial organs into blood get diluted, which creates a need for enhanced protein detection assays with improved sensitivity. This will enable early detection of proteins that may reveal diseases processes in organs that are difficult to access, they by facilitat-
ing early diagnosis, treatment or possibly disease prevention. Tools like ExoPLA can specifically distinguish different proteins on microvesicles surface with high specificity and sensitivity in a complex biological matrix and are therefore vital for early diagnosis. Over the past decade, lots of emerging protein assay technologies such as Simoa and Erenna assays have been developed with ‘single protein’ sensitivities. With the complexity of diseases, looking at proteins one at a time is not going to cut it. With still a long way to go from having highly validated antibodies, cross-reactivity for irrelevant proteins and sticking of detection reagents to surfaces becomes a central issue for these single binders assays. Not forgetting the Immuno RCA / PCR assays as well. They, on the other hand, can handle the investigation of multiple proteins due to their DNA-based architecture. In this regard, it is worth mentioning that DNA-assisted proximity bases immunoassays are efficient methods for sensitive, multiplex and specific protein detection in complex matrices. Somebody may argue that specificity or selectivity may be difficult to prove. I will have to disagree with that point. The molecular architecture of proximity ligations or extension assays is built on the two antibodies coupled to DNA molecules having to agree on the presence of a target. Now, I will take a second to go back to double binders’ immunoassays. They don’t have such a requirement, and anywhere a detection reagents sticks, absorbs on the surface or matrix will give rise to signal which merges with the actual signal. This is not the case with PLA, PEA or PLARCA systems as the probability of detecting a targeted decrease as the number of detection events needed to take place increases. Notwithstanding, in other work not described in this thesis, we are designing experiments to investigate the phenomenon of specificity. From an exploratory assay design perspective, alternative recombinant binders can be used to detect and distinguish proteins from the same family. This can also be a real advance in technology development and hopefully clinical adaptation.

As I try finalizing my work discussed in this thesis, the question I ask myself is, where does all the work presented in this thesis fit in the proteomic tree? As shown in (Figure 5.1), the field of proteomics covers a multitude of domains, my niche for my thesis with the papers I, II, III and IV, extends from studying protein expression in cells and tissues to PTMs, protein-protein interactions, EVs, and proteome mining (target detection and quantification).

In section 1, I started with the quotes from Lucretius “Nothing comes from nothing.” The work presented in this thesis is built on work done by many across disciplines from the publishing of RIA in 1959 to PLA assay in 2002 and everything before, during and after. I still wake up with this question, what is limiting the adaptation of new molecular tools for routine clinical use?

Now, I know it is all work in progress as we continue to build on improving existing technologies.
Paulo Coelho de Souza in the Alchemist said, “When you want something, all the universe conspires in helping you to achieve it.” Throughout this journey, I have been fortunate to meet and work with some amazing but really diverse group of individuals. How could I have kept track of them all throughout these years? Here, I will acknowledge colleagues, supervisors, committee members, admin personnel, mentors, family, friends and family.

My utmost gratitude goes to my faculty opponent Prof. Peter Nilsson and my examination board Docent Ylva Ivarsson, Prof. Vladimir Tolmachev and Mats Inganäs and my chairperson Docent Anna Dimberg. Thank you all for accepting to read my thesis and your participation in the discussion and evaluation of my work.

As Christopher Pike (real name Kevin McFadden) in Sati puts it “A true teacher would never tell you what to do. But he would give you the knowledge with which you could decide what would be best for you to do.” That is what Ulf Landegren is in my opinion. It took a long, long, long time to warm up to his brilliance in everything science and otherwise. I usually will leave his office more confused in an amazing way (now in retrospect) than when I got in. There was always something to reflect on. Thank you for your unfading interest in pushing your ideas against traditional methods.

Tack för din värdefulla och entusiastiska handledning och för den inspirerande vetenskapliga och innovativa atmosfären i din forskargrupp.

“He has a right to criticize, who has the heart to help,” said by Abraham Lincoln. Thank you, Masood Kamali, for your support throughout this journey. Thank you for all the practical, technical and scientific help. Thank you for all the random knocks on your ever-open door for all sorts of questions even in the midst of your deadlines. Thank you for your enthusiasm. Thank you for tolerating my thinking out loud and being myself.
Ulf & Masood, thanks for not confusing me more with very different and diverse opinions. Thank you, both for teaching me how to listen & think when developing tools.

Thanks to the European Union Commission for funding the BioMaX program that I was a part of. Thanks to the fantastic team and my fellow colleagues **Delfina, Sole, Fabi, Deniz, Elena, Stefano, Tamara, Beda, Malte, Diego, Matteo, Rory** and **Donacien**. Thank you all for making the deadlines, meetings, and trips enjoyable. Exploring science and Europe with you all was an experience I will never forget. Thanks for all the fun memories (with lots of crazy pictures ☺). I pray and hope we have a reunion this time in Africa ☺

I would like to thank the IGP administration staff for their selfless work in ensuring the smooth life of this journey. **Christina Magnusson**, you are one of the most selfless people I me this far. Thank you for all you do, your efficiency, your warmth, and the pure light you elude. Thank you **Tuulikki** for your warmth ☺ and always speaking to me in Swedish. **Claes Wadelius** Thank for all your help and corrections with the defense application. Thanks for all the fun activities for both learning and fun organized by the **IGP Ph.D.** student council. The winter masquerade party and the symposiums were very lovely.

Moltools past and present, It will be tough to remember everything but in the words of Maya Angelou “People will forget what you said, people will forget what you did, but people will never forget how you made them feel.” To the present members of the group **Peter**, your knowledge in cell biology among other things always leads me to knocks on your door. Thank you for the fascinating scientific discussions and lunch social talks, you should laugh often ☺. **Phathu**, you may never get around to learn that pidgin English but thank you for always listening to crazy thoughts and imagination. Thank you for all the interesting discussions and the late night lab work and workouts. You are a great gym partner ☺. **Lei**, your drive to make all things easier gave me a different perspective. Thank you for all the discussion on DNA-related stuff, the lots of laughter and my many questions. Thanks to you & your wife for my little friend Estelle & the pure joy she plants in my heart when she calls my name ☺. **Felipe**, my office mate, I thank you for the energy you bring to work every day. Your love for star wars made me try to get into it, but that was not going to happen (I tried ☺). **Johan H**, you always had the best illustrations to describe your work. Thank you for the occasional discussion and your every willingness to help. And your Fika homemade is always delicious ☺. **Rasel** thanks, for sharing your enthusiasms at work and your love for all things Pharmaco ☺. **Sophie**, your extensive knowledge on all things statistics is fantastic. Thank you for all the discussions and chats. Thank you for sharing your Chinese medicines for my different headache.
problems and always with a smile ☺. **Hongxing** Thanks for encouraging me, and also taking pictures. Thank you for always reminding us about respect for others & also to take care of equipment in the lab. Your warmth & smile is soothing ☺. **Reddy**, thank for the interesting discussion about different variants of setting up PLA. **Caroline**, thank you for always sharing your broad knowledge of science. Thank you for all the new publications on all things single cells and technology development. Thanks also for scheduling group meetings throughout the years’ ☺. **Radiosa**, thank you for all the discussions, all the interesting questions. Your love & care makes you a fun person to be with ☺. **Johan O**, thank you for your endless help with my computer problems. Thank for teaching me new about the tech world. **Elin**, your organization skills beat it. Thank you for your warmth ☺. Thank you for organizing wonderful Christmas parties, putting up the Christmas tree, and little things that make work feel like home ☺. **Johanna**, thank you for keeping the lab for us. Thank you for the entire home made goodies by the coffee machine. Thank you for your warmth & good heart ☺. **Erik**, your knowledge in all things money-hunting ☺ is a valuable resource. Thank you for all the songs in all the parties. **Liza** your strength and energy in dealing with everything is a source of admiration. Thank you for all the beautiful times, meetings, lunches, and most importantly your smile ☺, priceless ☺. **Anne-Li** ☺, I never met any Swede who was so outgoing, carefree and smart at the same time. Thank you for sharing so much with me and making my stay here beautiful. Thanks for all the lunches, dinners outside work and dances at parties ☺. **Gucci**, thanks for the interesting discussions. **Rachel**, you introduced me into the lab and thought me most of the things I know today. Thank you for your endless help in dealing with my many questions ☺. Thank you for introducing me to PLA, C2CA, and making lovely images. **Spyros** thanks for always answering my questions, providing alternatives to solving problems and responding to my emails. **Junhong**, thanks for the interesting discussion about science. We had some much fun outside the lab, thank for introducing me to sushi, and other Chinese food ☺. **Maria H**, thanks for your enthusiasm and all things conjugations. **Johan V** thanks for introducing me to Luminex technology. **Christina** thanks for taking care of he lab and ensuring its smooth functioning. **Andries** thank for your tremendous insight into all things molecular biology and signaling. **Joakim** thank for your great sense humor and looking for better ways to preserve samples. **Agata**, you are what we call beauty and brains. Thanks for not only sharing your knowledge in setting up western blots but also your incredible fashion sense ☺. **Lotta W**, thank you for providing me with samples for my first paper and for sharing your passion for all things exosomes. **Di**, thank you for sharing your brilliance, smart way of thinking and the after work games & fun ☺.

Thanks to **Leisa, Femi, Yajun**, your interest in developing molecular tools is ever fresh. Thank for starting the spPEA projects, which I am now contin-
Hannah darling, ☺ thank you for the wonderful discussions, the after work lunches. Thanks for sharing your view about life and for understanding my frustrations ☺. Samaneh thanks for being my student, and helping me to learn as I teach.

Söderberg group: Ola, thank you for sharing your knowledge in cell biology, pathology, developing molecular tools. Thank you for the morning swimming lessons. Dorotheya thank you for all the discussions, your constant smile and positive energy ☺. Axel thanks for keeping the HPLC in working conditions and teaching me how to set it up. Thanks for the interesting discussions. Johan H thanks for all the scientific discussions. Bjorn thanks for the nice discussions. Linda, I will never forget my first kettle bell class, you are a tough cookie. Thank you for the interesting discussions over the years ☺. Karin G you made me feel at home at work. Thank you for all the activities like the before lunch running. Please do not lose your selflessness ☺☺. Carl-Magnus, thank you for trying to help me understand how depression works ☺. Your passion for charity is amazing. Gaëlle thanks for all the discussion and after work fun ☺.

To Nilsson (MolDia) group: Mats N, your positive energy made it so much easier to talk to you. Thank you for your optimistic view on science and life. Thanks to the member of the MolDia group past and present Elin L. your energy and presence were amazing, thanks for being a good dance partner at defense parties. Anja, Thank you for all discussions and introducing me into the forensic teaching course. Annika your warmth and positive energy I will never forget. Thomas thanks for always helping me with the microscope every time I was in Stockholm. Marco M thanks for all the laughter about all things. Thanks for always explaining your projects and thinking that I am crazy ☺. Malte, what a journey we had together, thank you for being my partner in BioMax adventure ☺. Camilla, your aim of stretching blobs, was always fascinating. David H. thanks you for all showing me how to design & experiment with all things padlocks. Thank for introducing me to macros on Excel. Tagrid, thank your for being a friend, a sister and you ever smiley face ☺ love you, lady ☺. Elin F thanks for introducing me to Perl programming. Lotta thank you for all things selectors and your critical opinion on things. Thanks to Mustapha Amal, Anna, Tomas Rongqin and Magda for the dynamic environment.

Thanks to Andries L. and the OncoTrack consortium for providing the samples for my work.

Thanks to the Dynal Thermo Fisher people in Oslo, Norway for teaching me the different ways of coupling antibodies to magnetic particles. A special thank goes to Marie and Daren for their hospitality made the stay incredible.
Thanks to the Mentor4Research program for the amazing opportunity to learn from my mentor Katarina B. Thank for making the time to teach me things on commercialization, managements and much more. Thank for your guidance, help, and I look forward learning more from your expertise.

In the words of Nellie McClung, “Women are going to form a chain, a greater sisterhood than the world has ever known” The “Girls Club” is that sisterhood that made my life so much easier. Aww, ladies, you all know how much I love you Ruta, Fernanda, Vicky, Maryam, and Zohren. Our lunches, diners, and hangout were some of the best moments throughout this journey. You all have a very special place in my heart forever 😊

To the one person who has always been a source of inspiration is Joyce B Endeley. Your drive and work ethics are just incredible. You make it possible for young girls like me to dream to be more in the society where women in higher education is scares

To other academic people, I have met. Fidelis Cho your enthusiasm about science and you believe in me. Thank you for your encouragements. Thanks to Lloyd R, thank you for giving me an opportunity to study at Oulu University for my masters. Jamie C, you accepted me into your lab for my master project and introduced me into the world of crystals. Aman R, thank you for your love for developing assays for resource limiting setting. Thank you for always encouraging me to be more.

Thank my family away from home Emmy, Marie, Vera, Marilyn, Flore, Ma Ju, Daddy Melvin, Mr. Ebot, and Etchu E

To my family Papa and Mami, you guys, controlled everything in my life with an iron fist but when it came to school and career choice that was up to me. Thanks for the solid foundations you set for me to achieve my dreams. Thanks for the numerous and endless sacrifices you had to make. Agbor, you are the most challenging but smart member of the family. I love you with all your madness😊. To my amazing not so baby sisters Mawa and Manyi😊 thank you for your continuous support and love. Thanks for making me the happiest aunt on the planet with the loves of my life Florian E😊, Declan M😊, Francis E😊, Nathan E😊 and Micah M😊. Thanks to my uncle after my heart, Uncle Steve, I wish people saw the world through your eyes. Your selflessness is mesmerizing. Thanks to the Ebai & Sona families for their love and support. Thanks to my in-laws for sharing their son with me.
To my partner, my lover, my bestie, my tech guy, my man crush, and my go-to-guy Eyong. Thank you for your continual support, understanding, and patience. You make it all so much easier. Thank you 😊
References


102. administration FaD. https://www.fda.gov/Drugs/default.htm [ ]


110. WN B. Western blotting": electrophoretic transfer of proteins from sodium dodecyl. Analytical biochemistry. 1981;112(0003-2697 (Print)):195-203.

111. The Demonstration of Pneumococcal Antigen in Tissues by the Use of Fluorescent Antibody. The Journal of Immunology. 1942;45(3):159.


129. Engvall E - Perlmann P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. 1971(0019-2791 (Print)).
147. Todd J, B F, Held D, Morey J, Livingston R, Goix P. Ultrasensitive flow-based immunoassays using single-molecule counting. 2007(0009-9147 (Print)).

81
A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)