



universität
wien

MASTERARBEIT

Titel der Masterarbeit

„Evaluation of protein microarray technology for tumor
autoantibody screening in colon cancer“

Verfasserin

Khulan Sergelen BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2012

Studienkennzahl lt. Studienblatt:

A 066 834

Studienrichtung lt. Studienblatt:

Molekulare Biologie

Betreuer:

Priv.-Doz. Dr. Bernd Mayer

Acknowledgements

First and foremost, I would like to thank PD, Dr. Andreas Weinhäusel from the Department of Molecular Diagnostics, AIT Austrian Institute of Technology for his constant guidance and supervision and making this thesis possible.

I would like to express my gratitude to PhD, Priv.-Doz, Bernd Mayer from the Institute for Theoretical Chemistry, University of Vienna for his supervision and valuable advices for my thesis.

My deepest gratitude goes to my colleagues from AIT, Parvez Syed, PhD for his constant support and introduction to the protein microarray world and Ronald Kulovics for his assistance in protein expression, purification and microarray printing. I am thankful to DI. Michael Stierschneider from AIT for printing the protein microarrays. I am grateful for DI. Klemens Vierlinger for his advices and Stefanie Brezina for her translations.

I would like to thank A.o.Univ.Prof.Mag.Dr. Andrea Gsur and Philipp Hofer, MSc from the Institute of Cancer Research, Medical University of Vienna for providing the plasma samples.

I want to thank all my colleagues at AIT for their technical assistance and support.

I would like to convey my deepest gratitude for my family for their love and support in every way and every step.

CONTENTS

Acknowledgements	2
1 ABSTRACT	5
2 ZUSAMMENFASSUNG	7
3 BACKGROUND	9
3.1 Colon Cancer	9
3.1.1 Cancer epidemiology	9
3.1.2 Multiple step carcinogenesis	11
3.1.3 Colon Carcinogenesis	12
3.1.4 Epidemiology, screening and treatment	15
3.2 Tumor-associated antigens	18
3.3 Protein Microarray Technology	20
3.4 Protein microarray application for TAA discovery	21
4 AIMS	24
5 MATERIALS	26
5.1 Chemicals	26
5.1.1 Sample preparation	26
5.1.2 Protein expression and purification	26
5.1.3 Protein array spotting and probing	27
5.1.4 Instruments	28
6 METHODS	30
6.1 Clinical samples and IgG purification	30
6.1.1 Clinical samples	30
6.1.2 IgG purification	31
6.1.3 Protein expression and purification	33
6.1.4 Protein quantification	38
6.1.5 SDS-PAGE	39
6.1.6 Protein microarray production	39

6.1.7	<i>Protein microarray assay</i>	43
6.1.8	<i>Image processing and data analysis</i>	45
7	Results	50
7.1	Tumor autoantibody candidate marker screening using Protocol 1	50
7.2	Protein microarray assay optimizations	58
7.2.1	<i>Surface chemistries and blocking conditions</i>	58
7.2.2	<i>Assay conditions</i>	60
7.2.3	<i>Blocking agent in sample and detection antibody dilutions</i>	64
7.2.4	<i>Comparison of Buffer Detergents</i>	66
7.2.5	<i>Performance of serum vs plasma for autoantibody screening</i>	68
7.2.6	<i>Summary of optimizations</i>	72
7.3	Tumor autoantibody candidate marker screening using Protocol 2	73
7.3.1	<i>Control vs Carcinoma groups</i>	77
7.3.2	<i>Control vs Low risk vs High risk vs Carcinoma</i>	82
7.3.3	<i>Summary of class prediction analyses and selection of Colon Cancer Candidate array classifiers</i>	84
8	DISCUSSION	86
9	CONCLUSIONS	89
10	BIBLIOGRAPHY	90

1 ABSTRACT

Colorectal cancer is the third ranking cancer type worldwide with increasing incidences in developed countries. Progression of the disease takes several years and early detection and diagnosis following population based screenings has increased survival rate considerably. The screening methods in clinical application such as faecal occult blood test (FOBT) lack sensitivity and colonoscopy is unpleasant and invasive. Therefore, serum based minimal invasive methods are in great demand.

The concepts of specific molecular signatures in the different stages of tumorigenesis, and from there generation of tumor-associated antigens are highly anticipated as biomarkers for applications in diagnostics. Tumor associated antigens such as CEA and CA 19.9 are in clinical applications, but exhibit low sensitivity and specificity. Moreover, many studies have presented that panel of tumor biomarkers show higher sensitivity and specificity in detecting cancer than individual biomarkers.

It is recognized that abnormal expression of proteins in tumors exhibits antigenic ability and are recognized by the immune system, consequently producing tumor autoantibodies (TAA). TAAs in combination with protein microarray technology are a promising approach for tumor biomarker discovery. However, due to the complex nature of proteins, protein microarray experiments have low reproducibility and require careful optimization.

This thesis presents screening for tumor autoantibodies using colon cancer plasma samples of healthy controls, low risk polyps, high risk polyps and colon carcinoma groups utilizing protein microarrays containing several thousand proteins, aiming at identifying specific TAAs for generating a candidate marker array for subsequent validation in larger sample set. Due to the low performance of conventional assay protocols, several optimizations were carried out to establish a standard operating procedure for the particular type of protein microarray. Several aspects important for protein microarray processing were addressed and tested for the possible alternatives, such as protein microarray surfaces, blocking and buffer chemistries, as well as reaction conditions of the assays. Most importantly, the possibility of using purified IgG for tumor autoantibody marker screening was tested.

As a result, several changes to the protocol were made: probing in rotating chambers rather than horizontal humidity chambers, extension of sample incubation from 2 hours to 4 hours, addition of milk powder to samples, and optimization of detection antibody dilutions, replacement of the detergent Tween20 with Triton X-100 in buffers and using purified IgG of samples rather than serum. Tumor autoantibody candidate marker screening performance could be significantly improved from previous screening, with respect to sensitivities and specificities, increasing from ~54% to 97% for distinguishing patients versus controls using 25 greedy-pairs gene selection criteria for class prediction.

With this improved protocol, 632 classifier clones representing 593 genes were deduced from class prediction analyses as the most predictive TAAs. Additionally, 100 genes were selected from published literature on screening for tumor autoantibodies in colorectal cancer using protein microarray technology. This total of 732 clones will comprise the candidate marker array and will be applied in future studies for validation with 384 samples of 4 sample groups: healthy controls, low risk polyps, high risk polyps and colon carcinoma.

The current Master's thesis was carried out under the supervision of Dr. Andreas Weinhäusel, at AIT, Molecular Diagnostics within the research project "Evaluation of Serum-Autoantibody-Biomarkers for early diagnostic testing of Colon and Prostate Cancers" funded by "Life Science Krems", Austria.

2 ZUSAMMENFASSUNG

Dickdarmkrebs ist die dritthäufigste Krebserkrankung weltweit, mit zunehmender Häufigkeit an Neuerkrankungen in Industrieländern. Die Progression der Erkrankung erstreckt sich über mehrere Jahre, Früherkennung und Diagnose innerhalb eines bevölkerungsweiten Screenings haben die Überlebensrate beträchtlich gesteigert. Da die üblichen klinischen Screening-Verfahren wie fäkaler okkultter Bluttest (FOBT) geringe Sensitivität aufweisen bzw. Koloskopie eine invasive und für den Patient unangenehme Methode darstellen, ist die Entwicklung von serumbasierten minimal invasiven Methoden von hohem Interesse.

Das Konzept von spezifischen molekularen Signaturen in unterschiedlichen Phasen der Tumorgenese und die sich daraus ergebenden tumorassoziierten Antigenen können voraussichtlich als Biomarker in der Krebsdiagnose fungieren.

Tumorassoziierte Antigene wie z.B.: CEA und CA 19.9 finden schon klinische Anwendung, zeigen jedoch niedrige Sensitivität und Spezifität. Außerdem zeigen Studien, dass die Detektion von Krebserkrankungen mit Hilfe eines Tumorbiomarkerpanels höhere Sensitivität und Selektivität zeigt als bei Verwendung einzelner individueller Biomarker.

Es ist bekannt, dass abnorme Expression von Proteinen innerhalb von Tumoren antigenische Eigenschaften aufweist. Diese werden durch das Immunsystem erkannt und daraus folgend Tumorautoantikörper (TAA) gebildet. In Kombination mit Protein-Microarray Technologie sind TAAs ein vielversprechender Ansatz für die Entdeckung von Tumorbiomarkern. Jedoch auf Grund der komplexen Natur von Proteinen, zeigen Experimente mit Protein-Microarrays eine niedrige Reproduzierbarkeit, verglichen mit DNA-Microarrays, und bedürfen daher sorgfältiger Optimierung.

Diese Arbeit präsentiert ein Tumor-Autoantibody Screening von Dickdarmkrebsproben (gesunder Kontrollgruppe, Darmpolypen mit niedrigem Risiko Darmpolypen, Darmpolypen mit hohem Risiko und Dickdarmkarzinom) mittels Protein-Microarrays bestehend aus mehreren tausend Proteinen. Ziel ist es spezifische TAAs zu identifizieren, um ein Kandidatenmarker-Array herstellen und dieses mit einem größeren Probenet zu validieren. Aufgrund der niedrigen

Effizienz des konventionellen Assay-Protokolls wurden einige Optimierungen vorgenommen, um so eine standardisierte Arbeitsvorschrift für den Umgang mit den vorliegenden Protein-Microarrays zu erstellen. Wichtige Aspekte bei der Prozessierung der Protein-Microarrays wurden berücksichtigt und auf mögliche Alternativen getestet, wie beispielsweise Protein-Microarray Oberflächen, Blockier- und Puffer chemikalien und experimentelle Bedingungen. Am wichtigsten war es, die Verwendbarkeit von aufgereinigtem IgG für das Tumorausantikörper Screening zu testen.

Letztendlich wurden einige Veränderungen des Protokolls vorgenommen: Probestestung in rotierenden Kammern bevorzugt gegenüber statischer Feuchtigkeitskammer, Verlängerung der Probestestzeit von zwei auf vier Stunden, Zusatz von Milchpulver zu den Proben, Optimierung der Verdünnung des Detektions-Antikörpers, Ersetzen des Detergenz Tween20 mit Triton X-100 im Puffer und Verwendung von aufgereinigtem IgG besser als Serum. Die Leistung des Tumorantikörper-Kandidatscreenings konnte signifikant, im Bezug auf Sensitivitäten und Spezifitäten, gegenüber früheren Screenings verbessert werden. Eine Steigerung von ca. 54% auf 97% für die Unterscheidung zwischen Patienten gegen Kontrollen wurde unter der Verwendung von 25 greedy-pairs gene selection für statistische Klassenvorhersage erreicht.

Mit diesem verbesserten Protokoll wurden mittels class prediction analyses 632 Klassifikatorklone, welche 593 Gene repräsentieren, als die am besten vorhersagenden TAAs ausgewählt. Zusätzlich wurden weitere 100 Gene aus publizierten Tumor-Antikörper Screenings mittel Protein-Microarrays in Dickdarmkrebs herangezogen. Die insgesamt 732 Klone werden auf einem Kandidatenmarker-Array zusammengefasst und in der Zukunft für eine Leistungsvalidierung des Arrays mit 384 Proben (unterteilt in 4 Probengruppen) verwendet.

Die vorliegende Masterarbeit wurde von PD Dr. Andreas Weinhäusel, am AIT in der Forschungsgruppe „Molecular Diagnostics“, im Zuge des Forschungsprojekts „Evaluation of Serum-Autoantibody-Biomarkers for early diagnostic testing of Colon and Prostate Cancers“, finanziert durch „Life Science Krems“, betreut.

3 BACKGROUND

3.1 Colon Cancer

3.1.1 Cancer epidemiology

In 2008, the World Health Organization reported that cancer accounted for 13% or 7.6 million of deaths worldwide annually, making it a leading cause of death. This figure is expected to rise to about 13 million by 2030 (Ferlay J, 2010). The top most common cancer types are breast, lung, colon, prostate and cervical cancers, but differ among gender and socio-economic status (Figure 1).

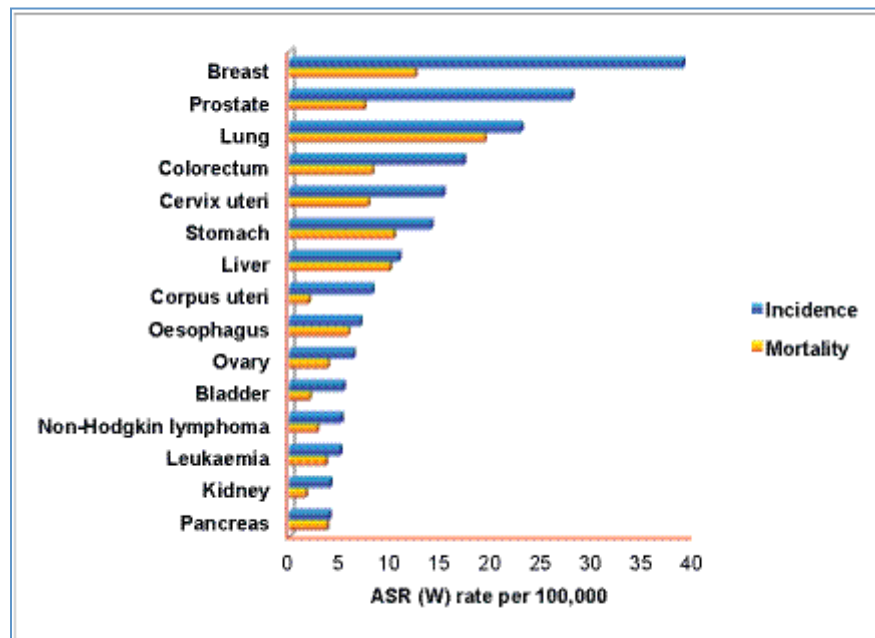


Figure 1. Estimated age-standardized incidence and mortality rates for cancer. Adapted from GLOBOCAN 2008 (Ferlay J, 2010).

According to the *Jahrbuch der Gesundheitsstatistik 2010, Statistik Austria*, in the year 2009 328,000 cancer patients in total were reported in Austria, with 37,039 incidences. 19,757 cancer deaths were registered in the same year and comprised a quarter of the total deaths. In Austria, following prostate and lung cancers, colon cancer is the third most frequent cancer among men with 13.6% incidence and 7.2% mortality rates. Colon cancer is also the third most frequent cancer type with 11.3% cases and 7.7% of deaths among women (Figure 2). Since

1999, incidences of colon cancer among men and women declined by 14.4% and 22.1%, mortality by 30.8% and 38.1%. Incidence and mortality rise significantly around the age of 50, incidence is highest around 60 and mortality around 80 years (Figure 3, Statistik Austria, Cancer incidence and mortality in Austria, 2010).

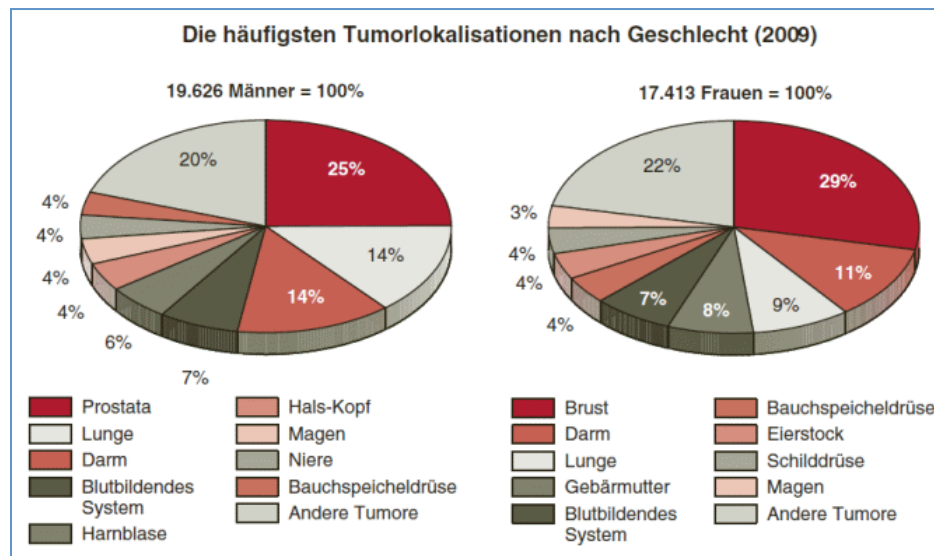


Figure 2. Common cancer prevalence by gender in Austria. (Statistik Austria, Health statistics yearbook, 2010)

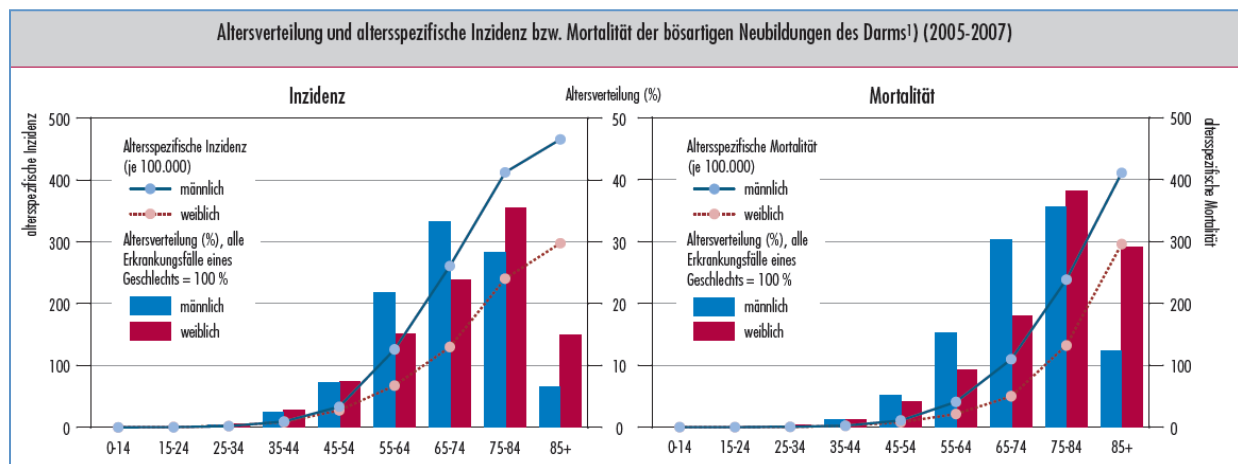


Figure 3. Age- and gender-standardized Incidence and Mortality of Colon Cancer in Austria (Statistik Austria, Cancer incidence and mortality in Austria, 2010).

3.1.2 Multiple step carcinogenesis

Cancer is an aberrant and uncontrolled growth of cells, also referred to as neoplastic diseases, as a result of malfunctioning of the controls governing normal cell proliferation and homeostasis. Classification and naming of cancers depend on their origin, with four main groups, epithelial, mesenchymal, hematopoietic and neuroectodermal. Carcinomas are the most common type of cancer and originate from the epithelium. Tumors can localize to the tissue and be non-invasive or metastatic and invasive, termed benign and malignant, respectively. The transition between non-invasive to invasive tumors arises due to traveling of some cancer cells from the primary tumor site to other tissues via the blood and lymphatic vessels. These kinds of metastatic or secondary tumors cause 90% of cancer deaths (Weinberg, 2007). In principle, carcinogenesis occurs through a three-stage process: initiation, promotion and progression. An irreversible damage to the DNA initiates neoplastic transformation followed by the expression of the mutant product mediating promotion. In the last stage, progression, cells are irreversibly transformed into malignant growing cells (Henry C. Pitot, 1993). It is suggested that for above stages to progress, cells need to require self sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of apoptosis, limitless replicative potential, sustained angiogenesis and tissue evasion and metastasis, as depicted in Figure 4. These capabilities are mainly acquired through gain of function mutations of oncogenes and loss of function mutations of tumor suppressor genes (Hanahan & Weinberg, 2000).

Normal cells of eukaryotes have several mechanisms to prevent aberrant growth of cells, by mechanisms that are controlled through the products of tumor suppressor genes and evading activation of possible oncogenes. Tumor suppressor genes (TGS) encode proteins that regulate normal cellular activities like cell cycle checkpoint control, transcriptional signaling, DNA damage and other stress responses. The main processes are apoptosis and cell cycle arrest or senescence. Majority of determined tumor suppressor genes are recessive and obey Knudson's "two-hit" hypothesis, meaning both gene copies must undergo mutation to suppress their activity. Only a few haploinsufficient TSGs are recorded, that can be inactivated through

mutation in only one copy. Aside from the famously studied p53, there are several other TSGs that are found in general or are cancer specific (Sherr, 2004).

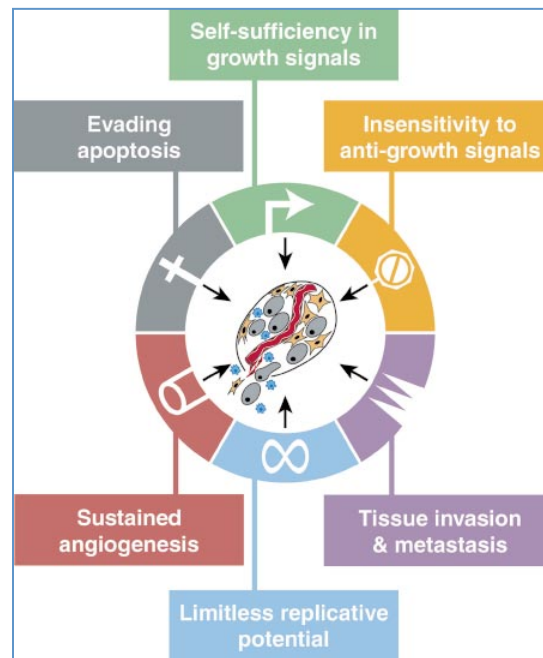


Figure 4. Capabilities acquired by cells for neoplastic transformation (Hanahan & Weinberg, 2000).

Oncogenes, on the other hand, function through activation or overexpression of their protein products, involved in cell proliferation and apoptosis. Mechanisms that are activated include mutations, fusion with other genes, translocation of enhancer elements and by amplification. Products of such genes range from transcription factors, signal transducers, growth factors and their receptors to apoptosis regulators and chromatin remodelers. There are several therapeutic approaches that target oncogenes, with some clinically approved products (Croce, 2008).

3.1.3 Colon Carcinogenesis

Colorectal cancer is the result of uncontrolled growth of a cell of the colon and/or rectal epithelial layer. This single layer cells form the so-called crypt structure along the connective tissue. The epithelial stem cells are located at the very base of the crypts and are the origin of

neoplastic conversion due to a series of successive mutations (Figures 5 and 6) (Humphries & Wright, 2008).

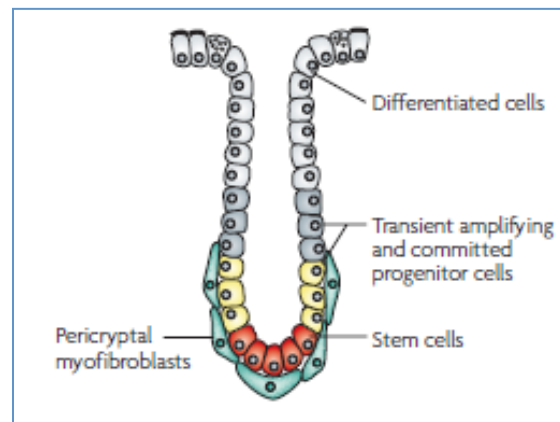


Figure 5. An illustration of the colonic crypt. Differentiation of stem cells is directed upwards from the base bordering the pericryptal myofibroblasts. Two intermediate cell stages: transient amplifying and committed progenitor cells (Humphries & Wright, 2008).

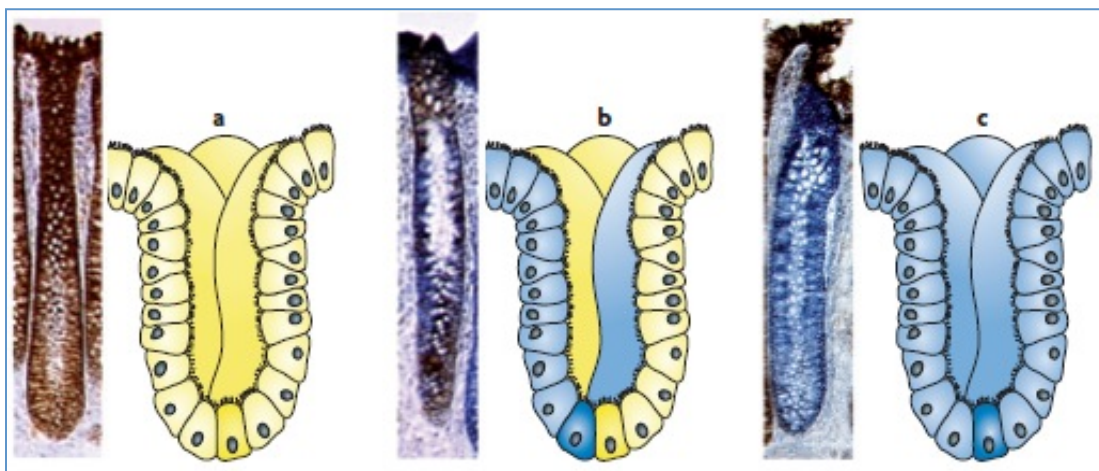


Figure 6. Clonal conversion with niche succession demonstrated on frozen section stained by double-enzyme histochemistry for cytochrome c oxidase (COX; brown) and for succinic dehydrogenase (SDH; blue), as a. wild type, b. mutant clones appear (blue) and c. the mutant clone succeeds throughout the crypt (Humphries & Wright, 2008).

In hyperplastic growth of cancerous cells, histology appears normal, however, epithelial cell division is faster and the tissue is slightly thicker. As a result of dysplastic growth, normal colonic mucosal cell layers are altered and have slightly different cell morphology than normal cells. Clear thickening of dysplastic cell growth in colon epithelia is termed a polyp or an

adenoma. Polyps that are attached to the colon wall are considered benign. A malignant polyp occurs when attachment is lost and growth breaks through the basement membrane into the stromal layers and smooth muscles. When the penetration is deep, cells migrate resulting metastasis (Weinberg, 2007).

Colon cancer carcinogenesis is one of the most studied in terms of the genetic events taking place. A mathematical approach to determine the model for colon carcinogenesis found it may have four stages (Figure 7). Initial rare events that transform colonic crypt stem cells probably involve mutations at the two APC (adenomatous polyposis coli) gene loci, override of the asymmetric division of the stem cell and localization due to space limitations. The following stage of initiated cell transformation into malignant cell was not described in detail, nevertheless, TP53 gene mutation and genomic instability are commonly observed (Luebeck & Moolgavkar, 2002).

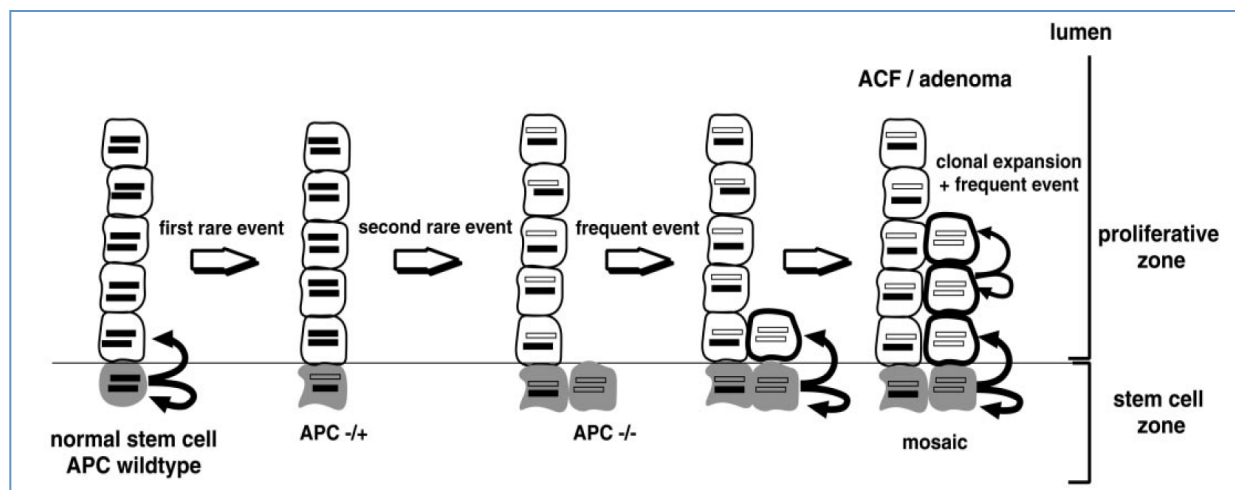


Figure 7. Four-stage model of adenoma formation from normal crypt (Luebeck & Moolgavkar, 2002).

It can be summarized from the numerous studies on the molecular basis of colorectal cancer, the concert of elements involved in genomic instability, tumor suppression gene inactivation through mutations, oncogenic activations and growth factor pathway alterations causing tumor progression. Aside from APC and P53 genes, many other genes and growth factors were found to be involved in the different stages of colon carcinogenesis (Figure 8) (Markowitz & Bertagnolli, 2009).

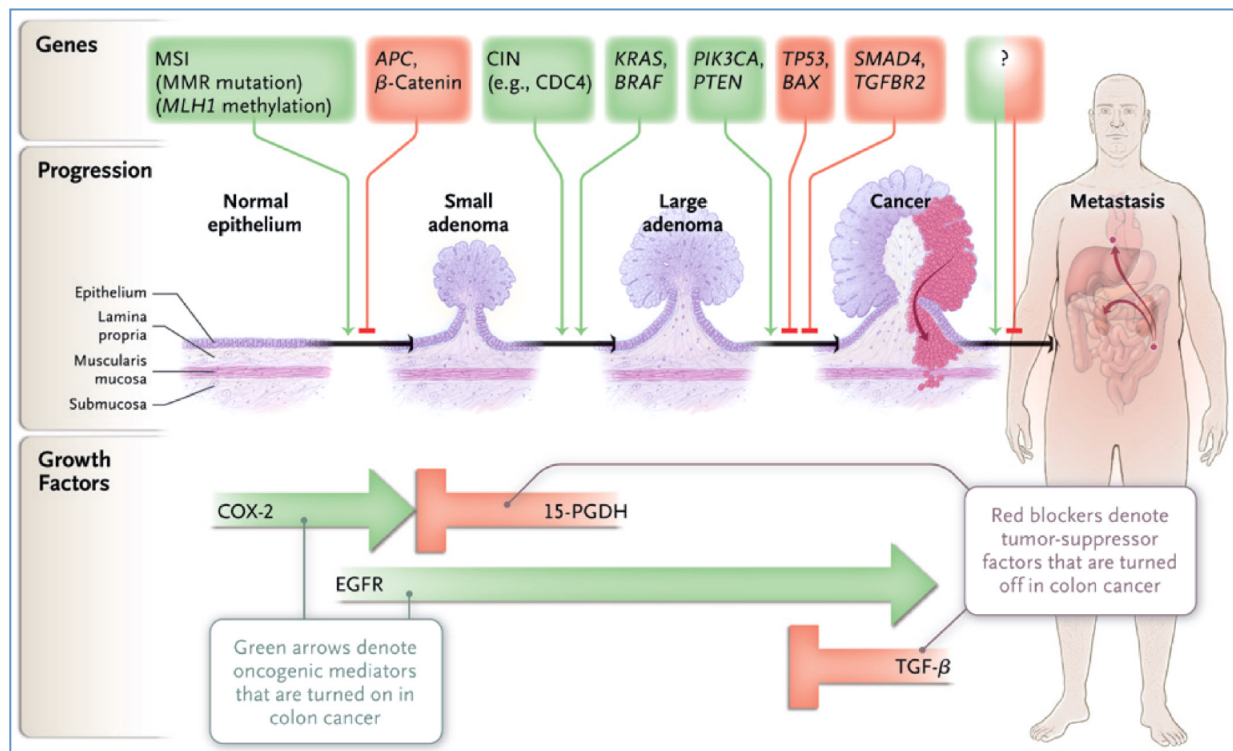


Figure 8. Colon cancer carcinogenesis associated genes and growth factors (Markowitz & Bertagnolli, 2009).

3.1.4 Epidemiology, screening and treatment

Colon cancer is largely age dependent, as incidence rises dramatically for individuals after 50 years of age with more tendentially men to be affected. Previous cases of polyps or cancer and family history of the diseases are another important risk factor. Lifestyle factors of the individual, such as diet, obesity, associated diseases like diabetes mellitus, smoking and high alcohol consumption are equal contributors for colorectal cancer development ("CDC - Colorectal Cancer Risk Factors," 2012).

The reduction of colon cancer incidences and increased survival in recent years can be attributed to diagnostic and treatment improvements. Around 80% of cancers are thought to arise from premalignant polyps or adenoma, therefore diagnosis during this stage is vital for prevention of malignancy. Screening of colon cancer can decrease mortality although incidence rates rise. The most commonly used method for screening is the faecal occult blood test (FOBT) that has low sensitivity, but could reduce the mortality rate by 16% if done every 2 years. The test basically detects blood in the stool. Conventional guaiac-based FOBT requires two samples

from three successive stools, whereas for immunochemical test, one of two stool samples is adequate. Immunochemical faecal test is not affected by animal blood from diet and is more sensitive for detection of CRC and advanced adenomas (Cunningham et al., 2010).

Screening by rectal imaging involves colonoscopy and flexible sigmoidoscopy. Both methods require bowel preparation through enema. Sigmoidoscopy is carried out with 60cm endoscope and can examine the sigmoid colon and rectum where most adenomas are found. Simple bowel preparation with self-administered enema is sufficient for sigmoidoscopy. The method is recommended every 5 years and can reduce incidence and mortality of CRC by around 60 to 80%. If small polyps are observed during the procedure, they are removed, however in the case of more or advanced polyps, colonoscopy is recommended. Important disadvantages associated with colonoscopy are risk of complications, requirement for sedation and complicated bowel preparation. Colonoscopy is recommended every 10 years. An alternative to the invasive imaging approaches, CT colonography, is comparable in sensitivity for cancer detection, however colonoscopy is required for confirmation (Cunningham et al., 2010).

Diagnosis of CRC is made based on the histopathological examination on biopsy sample obtained during colonoscopy or sigmoidoscopy. Localization of the tumor can be achieved with CT colonography or MRI, with MRI having advantage of measuring the extent. Diagnosis through histology or other methods are used to stage the tumor with the TNM (tumor, node and metastasis) system (Figure 9) (Cunningham et al., 2010).

Panel 4: TMN classification of colon cancer⁹¹

T=primary tumour

TX=primary tumour cannot be assessed
T0=no evidence of primary tumour
Tis=carcinoma in situ: intraepithelial or invasion of lamina propria
T1=tumour invades submucosa
T2=tumour invades muscularis propria
T3=tumour invades through the muscularis propria into subserosa or into non-peritonealised pericolic or perirectal tissues
T4a=tumour penetrates the surface of the visceral peritoneum
T4b=tumour directly invades or is histologically adherent to other organs or structures

N=regional lymph nodes

NX=regional lymph nodes cannot be assessed
N0=no regional lymph node metastasis
N1a=metastasis in one regional lymph node
N1b=metastasis in two to three regional lymph nodes
N2a=metastasis in four to six regional lymph nodes
N2b=metastasis in seven or more regional lymph nodes

M=distant metastasis

MX=distant metastasis cannot be assessed
M0=no distant metastasis
M1a=distant metastasis to one site
M1b=distant metastasis to more than one site
Surveillance, Epidemiology and End Results Program data for 5-year stage-specific relative survival rates in colon cancers:^{91,92}

- Stage I (T1, T2, N0): 97.1%
- Stage IIA (T3, N0): 87.5%
- Stage IIB (T4, N0): 71.5%
- Stage IIIA (T1, T2, N1): 87.7%
- Stage IIIB (T1, T2, N2): 75.0%
- Stage IIIB (T3, N1): 68.7%
- Stage IIIC (T3, N2): 47.3%
- Stage IIIC (T4, N1): 50.5%
- Stage IIIC (T4, N2): 27.1%

5-year stage-specific relative survival rates were similar for rectal cancer as compared with colon cancer.

Figure 9: Colon Cancer TNM Staging: Tumor Size, Invasion and Spread (Cunningham et al., 2010).

Significant efforts into improving performance of such methods have been made with major emphasis on their cost effectiveness. New endoscopic technologies are being developed that incorporate advances such as wider-angle views, enhanced imaging and fluorescent imaging. Aside from technological improvements, the strategies involved in population-based screenings are critical. With successful combination of strategies and technology, it would be possible to reduce the burden of disease, incidence and mortality (Quintero, Hassan, Senore, & Saito, 2012).

An alternative possibly cost reducing and non-invasive or less inconvenient method to endoscopic and faecal screenings is biomarker screening in blood. Blood based methods have the major advantage of minimal invasiveness and required samples. Due to the slow

progression and certain studied genetic, epigenetic, metabolic and protein biomarkers associated with the different stages of disease progression, such biomarkers may serve as prognostic and diagnostic tools. Genetic and epigenetic biomarkers associated with colorectal cancer have been studied extensively, yet there are limited numbers in clinical application. At present there are a few molecular biomarkers in clinical use, namely, faecal hemoglobin in stool, protein CEA and carbohydrate CA 19.9 and methylated gene *SEPT9* (Lofton-Day et al., 2008) in serum or blood. Several other potential biomarkers are reported with promise to improve and build on currently used biomarkers, but require large scale clinical studies to validate their proposed performances (T. Tanaka, Tanaka, Tanaka, & Ishigamori, 2010).

3.2 Tumor-associated antigens

The human immune system protects “self” from infection by recognizing and removing foreign agents. Cells of the immune system undergo maturations to omit those antibodies that are reactive against self-tissues and leaving the tolerant population, however a certain level of autoimmunity is found. In the case of autoimmune diseases these tolerance is broken or the naturally occurring autoimmunity is poorly controlled. Genetic (several genes) and environmental factors contribute in combination to the development of autoimmunity. The main mechanisms, from which autoimmune diseases arise are through exposure to environmental factors that alter the innate immune response, overproduction of proinflammatory cytokines and altering the regulations of inflammation (Fairweather, 2007).

From the decades of research in the human immune system and malignancies, their connection was seen through the presence of antigens against self, or autoantigens and antibodies that are produced against them (autoantibodies). This association and concept was demonstrated prominently by autoimmune diseases such as systemic lupus erythematosus (SLE) and scleroderma, other diseases like type 1 diabetes, gastrointestinal and neurological diseases (Eng M. Tan & Zhang, 2008).

Earlier studies demonstrated several systemic autoimmune diseases having their unique immune response by spontaneously producing autoantibodies reactive to intracellular proteins and nucleic acids. They have contributed to cell and molecular biology studies of such diseases

by giving insights into intracellular mechanisms and processes of the diseases. In the 1970s, two nuclear antigens *Smith antigen* (Sm Ag) and *nuclear ribonucleoprotein* (HNRNP), both with antibodies produced specifically by patients with SLE were identified and demonstrated to be located closely within the nucleus. The combined role of these two to antigens were described later to be associated with splicing of precursor mRNA (E M Tan, 1989).

Alternatively, aberrantly expressed proteins or tumor-associated antigens by tumor cells were found to stimulate immune response through production of autoantibodies (Y. T. Chen et al., 1997; Nelson, 1977; Steiner, Klein, & Klein, 1975; E. M. Tan & SHI, 2003). The underlying mechanisms of how tumor-associated antigens are produced are not fully described, but quite commonly mutations in the gene sequence can lead to mutated antigen product (Somasundaram et al., 2006). Several other mechanisms including alternative splicing, deregulated apoptotic or necrotic processes, post-transcriptional modifications, single nucleotide polymorphisms (SNPs) and overexpression can cause autoantigen production too (Backes et al., 2011)(Utz & Anderson, 1998).

Tumor-associated antigens and corresponding tumor autoantibodies have been described for a number of cancers such as breast, colon, lung, liver, prostate, ovarian, head and neck, leukemia and lymphoma and their cancer specific nature has opened the possibility to serve as prognostic or diagnostic biomarker (Anderson et al., 2008; Bauer et al., 2002; Chatterjee, 2006; Lin et al., 2007; Madrid et al., 2010; Massoner et al., 2012; Orzechowski et al., 2005; Perou et al., 2000; Rapberger et al., 2008; Somers et al., 2002; Steiner et al., 1975; Wang et al., 2005; Wu, Molinaro, Johnson, & Casiano, 2001).

3.3 Protein Microarray Technology

Protein microarrays offer a potent tool to study interactions between proteins and peptides on a large scale. It is the advanced, high-throughput form of immunoassays such as Radioimmunoassay (RIA) described by Yalow *et al.* (Yalow & Berson, 1959) and the enzyme-linked immunosorbant assay (ELISA) by Engvall *et al.* (Engvall, Jonsson, & Perlmann, 1971). Initially, Ekins *et al.* (Ekins, Chu, & Biggart, 1990) proposed that a-few-micron scale solid support antibody assays could yield high sensitivity comparable to that of macroarrays. Advances in studies on DNA microarray technology, such as spotting machinery, support surface chemistries and detection methods and principles, have been successfully applied to protein microarray technology. One of the earliest examples of implementation of DNA microarray to protein arrays is the use of a standard inkjet printer to apply 200µm diameter monoclonal antibody spots onto polystyrene film. Silzel and coworkers were able to reduce reagents, capture antibody, analyte molecule per zone with equal sensitivity and specificity as that of ELISA (Silzel, Cercek, Dodson, Tsay, & Obremski, 1998).

At present protein microarrays utilize recombinant-, fractions- or purified proteins, as well as synthetic peptides, for immobilization onto a microscope slide coated with various surfaces in a planar or 3D platform or are captured by affinity beads in a micro-well plate setting. Planar surface coating chemistries are categorized into 4 main groups based on the binding principle: non-specific- noncovalent and covalent, specific- noncovalent and covalent. The first group includes nitrocellulose and poly(L-lysine), the second, aldehyde and epoxy, the third, certain affinity interaction molecules like streptavidin–biotin, His-tag–nickel-chelates, and the fourth self-assembly monolayers (SAMs) on gold coated surfaces (Tomizaki, Usui, & Mihara, 2005).

Coating chemistries are vital in the overall performance of the protein microarray and should offer low background, maintain proper orientation and reduce the effect on the sensitive 3-D structure of the proteins (Predki, 2007). The most commonly used surface chemistries in antibody microarrays are nitrocellulose, amines, aldehyde or epoxy. However, there have been only a limited number of studies comparing surface chemistries published so

far (Fici et al., 2010; Huang, Huang, Fan, & Lin, 2001; Lee, Kim, Choi, Shin, & Kim, 2011; Nam et al., 2003; Olle et al., 2005; Tomizaki et al., 2005).

Due to the highly complex nature of protein interactions, new technologies to improve performance of protein microarrays have been developed, like the so-called 3D surface setups: agarose, hydrogel and more recently nanoparticles and beads in microwell plate formats. These approaches aim to overcome the difficulty of limited spot density, detecting low abundance proteins, increased signal to noise ratio, decreased non-specific binding and avoiding cross-reactivity (Chandra, Reddy, & Srivastava, 2011; H. Chen et al., 2009; Guillaume et al., 2005; Kang, Trofin, Mota, & Martin, 2005; Lee et al., 2011; H. Tanaka et al., 2009; Tsarfati-Barad, Sauer, Preininger, & Gheber, 2011; Zubtsov et al., 2007).

Spotted protein arrays are processed with cancer patients' or control individuals' serum samples and binding of proteins are detected using methods that avoid cross-reactivity, enable high-resolution signals in high throughput format with good reproducibility and are cost effective (Tomizaki et al., 2005). The same strategies as in immunoassays, like sandwich, antigen capture and direct, do also apply to protein microarrays (Carlsson et al., 2011; Haab, Dunham, & Brown, 2001; Nolen et al., 2008; Pla-Roca et al., 2012; Schweitzer et al., 2002; Sreekumar et al., 2001).

3.4 Protein microarray application for TAA discovery

Different approaches to produce the protein analytes for tumor biomarker discovery and validation exist, ranging from recombinant proteins to tumor tissues fractions. TAA biomarker discovery initially begins with screening on a proteomic scale, subsequently narrowing down to sets of proteins that best distinguish the tumor type under analysis. Several studies have demonstrated the use of different liquid chromatography fractioning of proteins from tumor (Bouwman et al., 2003; Yan et al., 2003) or tumor derived cell lines (Madoz-Gúrpide, Kuick, Wang, Misek, & Hanash, 2008; Nam et al., 2003; Pereira-Faca et al., 2007; Qiu et al., 2004) in large scale. Using such native proteins for functional analysis are advantageous in having accurate post-translational modifications, however, major drawbacks lie with the likelihood of

manifold proteins present in the spots, as well as the dependence on the tissue sample they originate from (Lu, Goodell, & Disis, 2008; Qiu et al., 2004).

Recombinant proteins from cDNA libraries can be expressed via bacterial, insect or mammalian expression systems. With the so-called SEREX technology, high-throughput expression of libraries consisting of thousands of clones is effectively produced by bacterial expression systems. Mammalian cells are the most suitable expression system for human proteins (specifically with respect to post-translational modifications), in regard to correct folding, using many viral vectors such as adenovirus and retrovirus. However, the technique lacks in ability to scale up, providing sufficient yield, being consequently not time and cost effective (Skena, 2005).

Several studies have addressed TAA discovery as a diagnostic biomarker using the protein microarray approach (Table 1). For instance in colon cancer, Nam *et al.* (Nam et al., 2003) identified an antigenic target of colon cancer by using a protein microarray containing 1760 solubilized protein fractions obtained from a human colorectal adenocarcinoma cell line LoVo. Proteins were immobilized on nitrocellulose-coated slides and hybridized with 15 plasma samples each from colon and lung cancers and healthy controls. From the total of 1760 fraction, 39 exhibited higher reactivity to colon cancer samples. One fraction reactive to 9/15 colon cancer sera was identified with mass spectrometry as ubiquitin C-terminal hydrolase isozyme 3 (UCH-L3). Antibody to UCH-L3 could be detected in 19/43 sera from patients with colon cancer, but none of the 54 sera of lung cancer, colon adenoma or healthy subjects.

By using commercial protein microarrays containing 8000 human proteins, Babel *et al.* (Babel et al., 2009) screened sera from colorectal cancer (CRC) patients and healthy subjects and were able to identify 43 proteins that were recognized by tumoral sera but not the control sera. From these 43 proteins, 5 immunoreactive antigens, PIM1, MAPKAPK3, STK4, SRC, and FGFR4 were shown to be highly prevalent in cancer samples. By using an ELISA with PIM1, MAPKAPK3, and ACVR2B they showed specificity and sensitivity values of 73.9 and 83.3% (area under the curve, 0.85), respectively.

From a 37,830 clone recombinant human protein array, Kijanka *et al.* (Kijanka et al., 2010) could identify 22 antigens able to distinguish between colorectal cancer and non cancer patients. 18 antigens are specific for cancer and 4 are for control. Expression of the antigens (p53, high mobility group B1 (HMGB1), TCF3, tripartite motif-containing 28 (TRIM28), longevity assurance gene homologue 5 (LASS5) and zinc finger protein 346 (ZNF346)) using quantitative reverse transcription PCR (Q-PCR) and tissue microarray immunohistochemistry involvement in various cellular processes.

Antigen	Cancer Type	Discovery method	Validation Method	Sample type	Study size	Sensitivity (%)	Specificity (%)	Ages (Mean in years)	Ref.
p53, HMGB1, TRIM28, TCF3, LASS5, ZNF346 and 6 other antigens	CRC	High density protein array	NA	Serum	Training set: 20 patients and 20 controls. Validation set: training samples (40) and additional 23 patient and 20 control samples	83.7	80	Training set: 62 (patients) and 61 (controls)	(Kijanka et al., 2010)
CEA	CRC	Recombinant proteins	Dissociation-Enhanced Lanthanide Fluorescent Immunoassay (DELFI)	Serum	259 (137 men and 122 women)	34	90	67 (Men) and 70 (Women)	(Carpelan-Holmström, Haglund, Lundin, Järvinen, & Roberts, 1996)
5 anti-TAAs: Imp1, p62, Koc, p53 and c-myc	CRC	Recombinant proteins	ELISA	Serum	46	60.9	89.7	56.8 (Patients)	(Liu et al., 2009)
MAPKAPK3 and ACVR2B	CRC	High Density Protein Arrays	ELISA	Serum	52 CRC patients and 42 controls	83.3	73.9	71 (Patients) and 61.7 (Controls)	(Babel et al., 2009)
CCCAP, HDAC5, p53, NMDAR, NY-CO-16	CRC	Recombinant proteins	ELISA	Serum	94 CRC patients and 54 control	77.6	58.5	64 (Patients)	(Chan et al., 2010)

Table 1. Examples of screening and validation studies on TAA biomarker panels using clinical samples of Colorectal Cancer patients and healthy controls. Single biomarker, CEA can be seen to lack sensitivity and conversely, panels of different TAA biomarkers are able to classify disease group from healthy controls with higher sensitivity (Adapted from Syed, 2012).

4 AIMS

Tumors are found to evoke the immune system and stimulate the production of tumor autoantibodies (TAAs), that consequently are a promising serological tool for minimal invasive diagnosis of colon and other cancers (E. M. Tan & SHI, 2003). Protein microarrays are an ideal approach to identify and develop TAA tests (Anderson et al., 2008; Kijanka et al., 2010; Massoner et al., 2012). However, the complex nature of proteins and variability of protocols applied in different research settings make it difficult to produce highly reproducible results and therefore an optimal protein microarray processing protocol is in demand.

The presented thesis was aimed at evaluating the protocol previously applied for protein microarray experiments in the context of TAA candidate marker screening using high-density protein microarrays. Clinical samples from different colon disease stages (prior to and established diagnosis of colon cancer) and healthy controls were used to perform the screening experiments. The shortcomings of the protocol were addressed from an initial screening experiment that was performed using 100 clinical samples belonging to four sample groups: healthy control, low risk polyp, high risk polyps and colon carcinoma, respectively. As the performance of the pilot screening had failed to produce a panel of classifiers with adequate ability to classify disease and healthy groups, several technical issues were addressed that may be important in protein microarray assays. The following goals had been set to improve the protocol and the reproducibility of protein microarray experiments:

- Selecting the ideal slide surface to produce protein microarrays
- Choosing an efficient blocking buffer to reduce background
- Setting an optimal sample incubation time-point to obtain sufficient specific bindings
- Setting an optimal experimental condition to ensure sample distribution over the arrays
- Testing of options to reduce unspecific binding, e.g. eliminate plasma matrix effect

Upon optimizing the protocol, a TAA candidate biomarker screening experiment is to be performed to elucidate a panel of the best classifiers of the four different sample groups. The models for utilizing protein expression profile to predict the classes of future samples are developed based on the *Compound Covariate Predictor*, *Diagonal Linear Discriminant Analysis*, *Nearest Neighbor Classification* and *Support Vector Machines* with linear kernel. Genes differentially expressed at the specified significance level or a defined number of greedy-pairs genes are incorporated into the models. The prediction error of each model will be estimated using leave-one-out cross-validation (LOOCV).

5 MATERIALS

5.1 Chemicals

5.1.1 Sample preparation

- Serum Clot Activator 9ml Vacuette, Greiner Bio-GmbH (Cat# 455092)
- MelonTM Gel Spin Plate Kit for IgG Screening, Thermo Scientific (Cat# 45208)
- MelonTM Gel Purification Buffer (100 ml), Thermo Scientific (Cat# 1859376)
- 96 well Spin Plate (Cat# 1825285)
- Wash Plate (Cat# 1860083)
- Collection Plate (Cat# 1860082)
- NuPAGE[®] MOPS SDS Running Buffer (20x), Invitrogen (Cat# NP0001)
- NuPAGE[®] 4-12 % Bis Tris Gel, 1.0 mm x 17 well, Invitrogen (Cat# NP0329BOX)
- NuPAGE[®] LDS Sample Buffer (4x), Invitrogen (Cat# NP0007)
- NuPAGE[®] Reducing Agent (10X), Invitrogen (Cat# NP0004)
- NuPAGE[®] Antioxidant, Invitrogen (Cat# NP0005)
- Ladder: SeeBlue[®] Plus 2 Prestained Standard (Cat#: LC5925)
- Roti[®]-Blue (5x), ROTH[®] (Cat#: A152.1)

5.1.2 Protein expression and purification

- UniPEx human cDNA library (human fetal brain, T-cells, lung), imaGenes GmbH (IMA-PROT-ARR)
- Glycerol minimum 99% GC, Sigma (Cat# 200-289-5)
- D-(+)-Glucose, Sigma (Cat# G6152)
- A-Lactose-Monohydrate, Sigma (Cat# L2643)
- Na₂HPO₄- di-sodium hydrogen phosphate, anhydrous GR for analysis, MERCK (Cat# 106586)
- KH₂PO₄- potassium dihydrogen phosphate, GR for analysis, MERCK (Cat# 104873)
- (NH₄)₂SO₄- ammonium sulfate GR for analysis, MERCK (Cat# 101217)
- MgSO₄- magnesium sulfate, dried, ROTH (Cat# 0261.1)
- BD BactoTM Tryptone, pancreatic digest of casein, BD Beckton Dickinson (Cat# 211705)

- Yeast extract, Fluka (Cat# 70161)
- Ampicillin sodium salt, Sigma (Cat# A9518)
- Kanamycin sulfate, Applichem (Cat# A1493)
- Tris-HCl Trizma® hydrochloride, Sigma (Cat# T5941)
- rLysozyme, Novagen (Cat# 71110)
- Benzonase nuclease >99% pure, Novagen (Cat# 70664)
- PMSF Phenylmethanesulfonyl fluoride, Sigma (Cat# P7626)
- MgCl_2 hexahydrate, GR for analysis, MERCK (Cat# 105833)
- Urea, PlusOne™ Urea, GE Healthcare (Cat# 171319-01)
- NaH_2PO_4 Sodium dihydrogen phosphate, GR for analysis, MERCK (Cat# 104873)
- Imidazole, Fluka (Cat# 56748)
- SDS solution, Fluka (Cat# 5030)
- NaCl, Sigma (Cat# S3014)
- NaN_3 Sodium azide extra pure, MERCK (Cat# 106688)
- Coomassie (Bradford) Protein Assay Kit, Thermo Scientific (Cat# 23200)

5.1.3 Protein array spotting and probing

- SU8 epoxide coated glass slides, TOP Oberflächen GmbH
- Micro Cleaning solution, Telechem (Cat# MCS)
- MilliQ purified water
- 16K protein microarrays representing 15,770 UniPEx library clones
- 642 arrays
- 642+H arrays
- DIG Easy Hyb (500 ml), Roche (Cat#: 70013-016)
- GIBCO® PBS pH 7.2, Phosphate Buffered Saline (10x), Invitrogen (Cat# 70013-065)
- TWEEN® 20, SIGMA® (Cat# P1379-100ML)
- Triton™ X-100, SIGMA® (Cat# T9284)
- Milk powder: sofortlösliches Magermilch-Pulver, Fixmilch instant
- Goat anti-human IgG (H+L) Alexa Fluor® 647, Invitrogen (Cat# A-21445)
- Penta-His Alexa Fluor® 532 conjugate, Qiagen (Cat# 35330)

- Microscope coverslips
- Humidity chamber

5.1.4 Instruments

- Hard-Shell® Thin-Wall 96-Well Skirted PCR Plates, Bio-Rad (Cat# HSP-9611)
- 96-Well Filter plates, Porvair Sciences (Cat# 360112)
- Pipetboy: BIOHIT, Midi Plus
- Analytical Balance: Pioneer™ PA1
14CM, OHAUS®
- Balance: Pioneer™ PA2102, OHAUS®
- Magnetic stirrer: Variomag poly, Thermo Scientific
- Magnetic stirring hotplate: MR3002 S8, Heidolph
- Centrifuge 5415D, Eppendorf
- Shaker: MS 2 Minishaker, IKA®
- Platform shaker: Titramax 1000, Heidolph
- Arium® p50 UV, Sartorius stedim biotech
- Heraeus Multifuge 3S-R, Thermo Electron Corporation
- Epoch™ Microplate Spectrophotometer, BioTek®
- Take3™ Micro-Volume Plate, BioTek®
- Hybridisation oven: Model, SHEL LAB
- Agilent G2534-60005 Hybridization Gasket Slide - 1 microarray per slide format, Agilent Technologies
- Agilent G2534A Microarray Hybridization Chamber, Agilent Technologies
- DNA Microarray Scanner with “Surescan high-Resolution Technology”, Agilent Technologies
- Zeba™ 96-well Desalt Spin Plate, PIERCE
- GeneAmp® PCR System 2700, Applied Biosystems™
- Freezer: 5812, Thermo Scientific
- OmniGrid 100, GeneMachines
- Centrifuge with plate/slide adapter (e.g. Sorvall Legend RT model)

- XCell SureLock™ Mini Electrophoresis Cell, Invitrogen™
- PowerEase® 500 Power Supply

6 METHODS

6.1 Clinical samples and IgG purification

6.1.1 Clinical samples

Plasma samples were collected within the ongoing “Molecular Epidemiology Colorectal Cancer Study of Austria” (CORSa) conducted at the Institute of Cancer Research, Medical University of Vienna (P. Hofer et al., 2011; P. Hofer et al., 2012). In the CORSa study more than 6,000 Caucasian participants were recruited since May 2002 within a large, province-wide screening project in the province Burgenland, Austria. This CRC screening program uses fecal occult blood tests (FOBT) and all habitants of Burgenland aged between 40 and 80 years are invited annually to take part in the screening. Positively tested participants receive further diagnostic workup including colonoscopy and are asked to take part in the CORSa study. All participants (100%) had a positive FOBT and underwent complete colonoscopy up to the cecum. Together with a whole blood sample, a short questionnaire on demographic and anthropometric factors, dietary and smoking status was obtained. Heparinized plasma was immediately centrifuged at 2000 g for 10 minutes and stored at -80°C. All participants gave written informed consent and the study was approved by the institutional review board “Ethikkommission Burgenland”.

CRC cases were patients with histologically confirmed, previously untreated CRC, newly diagnosed within this screening program. For statistical analysis the polyp group was classified by their histopathology as high-risk and low-risk polyps according to the amount of villous elements. Adenomatous villous, adenomatous tubulovillous, and co-occurrence of adenomatous tubular with tubulovillous polyps were classified as high-risk polyps while hyperplastic polyps and adenomatous tubular polyps were assigned to the low-risk group. The control group comprised individuals with positive FOBT who underwent complete colonoscopy to the cecum and exhibited no pathological findings. Participants with serious medical conditions at baseline, including all malignant diseases were excluded from the study.

For this project, 512 plasma samples of the CORSA cohort were provided, in a test set (n=128) and a validation set (n=384). Test set and validation set comprised of four groups, based on histology (CRC, high-risk polyp, low-risk polyp and control). The four groups comprised 32 samples each (test set) and 96 samples each (validation set), respectively.

For optimization experiments, serum samples from 6 healthy volunteers from the laboratory were collected in Vacuette® serum clot activator containing collection tubes (Greiner Bio-One, Austria). After collection, blood was incubated at room temperature for 30-45 minutes to allow clotting. Serum was obtained by centrifugation of samples for 15 minutes at 2000rcf, aliquoted in cryovials and stored at -80°C until use.

6.1.2 IgG purification

6.1.2.1 Principle

Human immunoglobulin or antibodies are glycoproteins with akin structure and function with around 90% protein and 10% carbohydrate content. The five classes of immunoglobulin, IgG, IgM, IgA, IgD and IgE together represent the humoral immunity in humans. They consist of two identical copies of each “L” light and “H” heavy polypeptide chains that are held together by noncovalent and covalent disulfide bonds forming a bilaterally symmetric structure. The type of heavy chain determines the Ig class and L and H chains each contain a “V” variable domain having highly conserved amino acid composition that comprise the antigen binding sites or epitopes (Figure 10).

The immunoglobulin G or IgG having Gamma heavy chains is the most abundant class of antibody in human serum comprising 75% of total Igs with a concentration of approximately 10-16mg/ml. IgG acts predominantly in the secondary immune response and can cross the placenta, unlike the other classes of Ig in humans, and is responsible for long-term immunity, memory antibodies, neutralization, agglutination, opsonization and complement fixation. The heavy chain of IgG has a molecular weight of 50kDa and light chain of 25kDa.

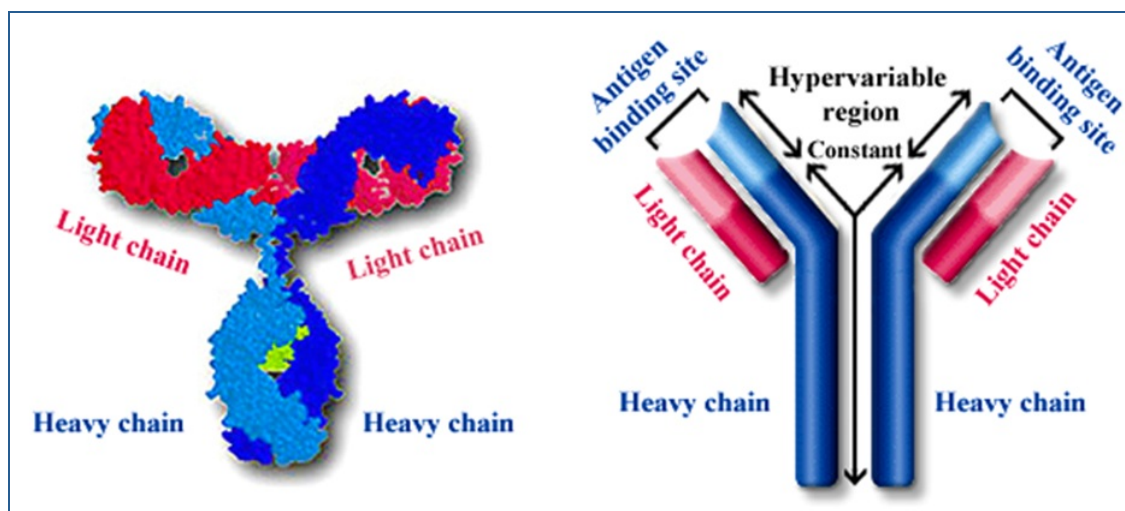


Figure 10. Structure of IgG in 3D (left) and schematic representation (right). Symmetrical structure formation of two heavy and two light chains and terminal ends containing antigen binding site (“Western Blotting,” 2012).

The three main methods for purifying antibodies involve physicochemical fractionation, class-specific affinity and antigen-specific affinity. The Melon™ Gel systems by Thermo Scientific uses physicochemical fractioning with resins that binds to non-IgG proteins present in serum and other sources, resulting in selective IgG fractionation in the flow-through. It also allows removal of interfering proteins such as BSA from the sample (Meulenbroek, 2002; “The human IgG subclasses,” n.d.; Thermo Scientific, 2010a).

6.1.2.2 Procedure

IgG from the samples were purified using the Melon™ Gel Spin Plate Kit for IgG Screening (Cat#45206, Thermo Scientific) according to the manufacturer’s protocol. The factory seals from the Melon™ Gel Spin Plate are removed, the plate is placed on a white wash plate and the entire assembly is centrifuged to remove the storage buffer. Afterwards, the Spin plate is positioned on the blue Collection plate and 100ul of plasma samples diluted 1:10 in the Melon™ Gel Purification Buffer is applied to a well in the Melon™ Gel Spin Plate. To obtain enough IgG for each sample, a total of 300ul of the 1:10 diluted sample was prepared and applied to 3 wells on the spin plate. Samples are incubated in the spin plate for 5 minutes on an orbital plate shaker. The purified IgG is collected in the collection plate by centrifuging the

assembly at 1,000xg for 2 minutes. The IgG solution from each sample is pooled and stored at 4°C until use in protein microarray experiment and at -80°C for longer storage.

6.1.3 Protein expression and purification

6.1.3.1 Principle

The *UniPEX* library (purchased from ImaGenes, Source Bio-Science) applied in this study is a protein expression library consisting of clones in two differently modified pQE (Qiagen) vectors. The library contains 15,456 in-frame ORF clones representing 7,390 human genes that were deduced from 100,000 sequenced clones of human fetal brain, T-cells and lung protein expression libraries. The *UniPEX-1* clones (n=7560) are inserted into pQE30NST vector (Figure 11) that allows auto-inducible expression of His6-tagged fusion proteins and contain Ampicillin and Kanamycin resistances. *UniPEX-2* clones (n=7896) are inserted into pQE80LSN vector (Figure 12) that allows IPTG-inducible expression of His6-tagged fusion proteins and contain Ampicillin resistance gene (Source BioScience imaGenes, 2012). The clones of the *642* and *642+H* arrays were inserted into pQE30NST vector and auto-induction strategy was optimized as in (Stempfer et al., 2010; Syed, 2012).

6xHis tags have affinity to Ni-NTA and when expressed in fusion with the *UniPEX* proteins, this property can be utilized to purify these proteins from the expression host. 6xHis tags are reported to be uncharged, non-interfering to secretion, compartmentalization, folding of the fusion proteins in the cell, structure and function of the purified proteins. Nitrilotriacetic acid (NTA) is a tetradentate chelating absorbent that binds to four ligand-binding sites of the nickel ion that leaves two sites for 6xHis tag to interact as shown in Figure 13. The 6xHis tagged proteins are bound to Ni-NTA coupled to Sepharose® CL-6B resin (Qiagen) that have the capacity to bind 5-10 mg His-tagged proteins per milliliter of agarose resin.

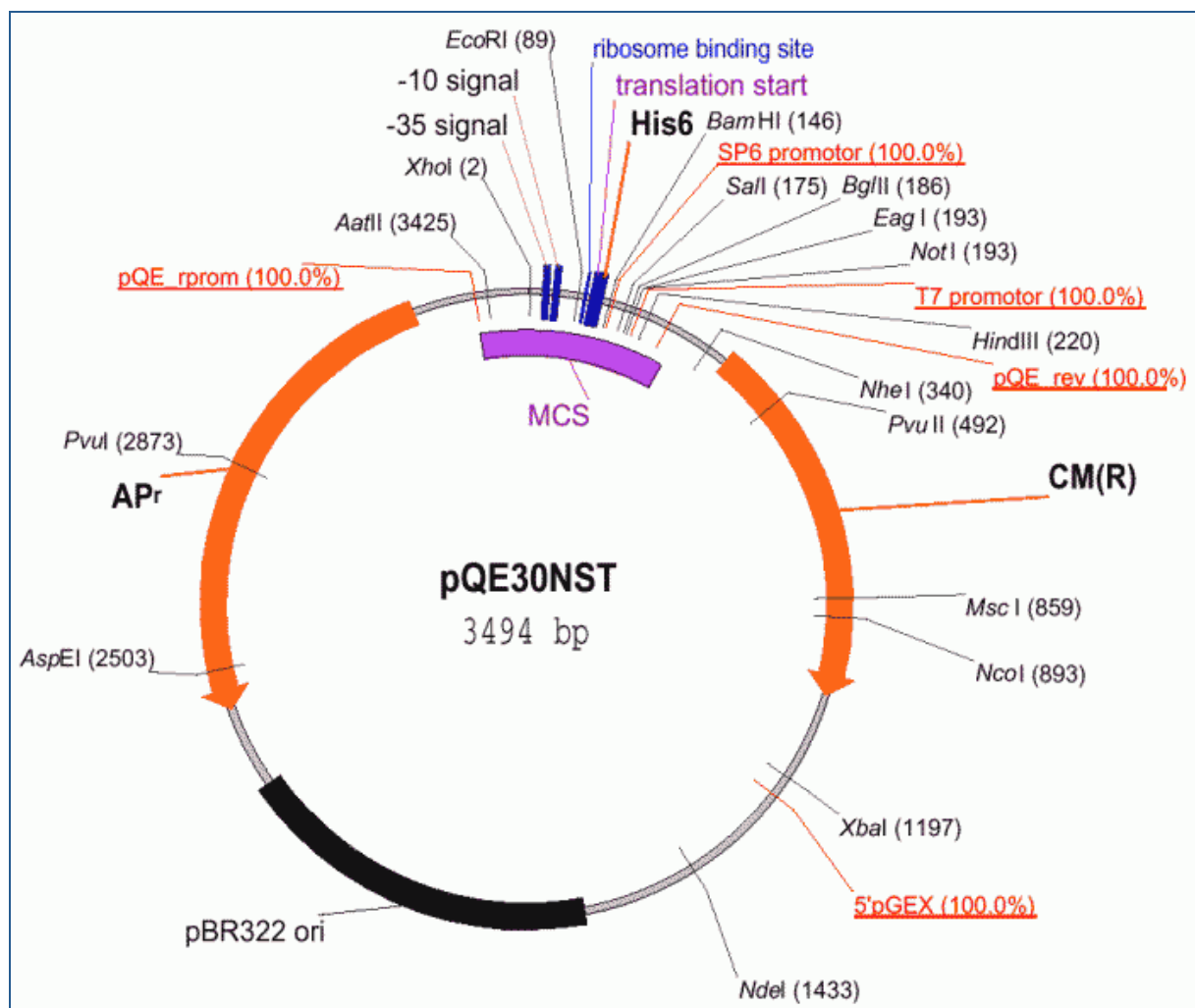


Figure 11. Vector pQE30NST with an N terminus 6xHis tag ("www.imagenes-bio.de," 2012).

Denaturing of the proteins prior to applying onto Ni-NTA would solubilize the proteins and completely reveal the 6xHis tag making binding more efficient. After several rounds of washing, the expression host cell debris and non-specific protein products are excluded, leaving only 6xHis tagged proteins. With low 10-20mM concentrations of imidazole in washing buffers prevents the possibility of low affinity binding of non-tagged endogenous proteins with histidine clusters. A higher concentration of imidazole (>100mM) competes with 6xHis tag binding to nickel ions and is used for protein elution (Qiagen, 2003).

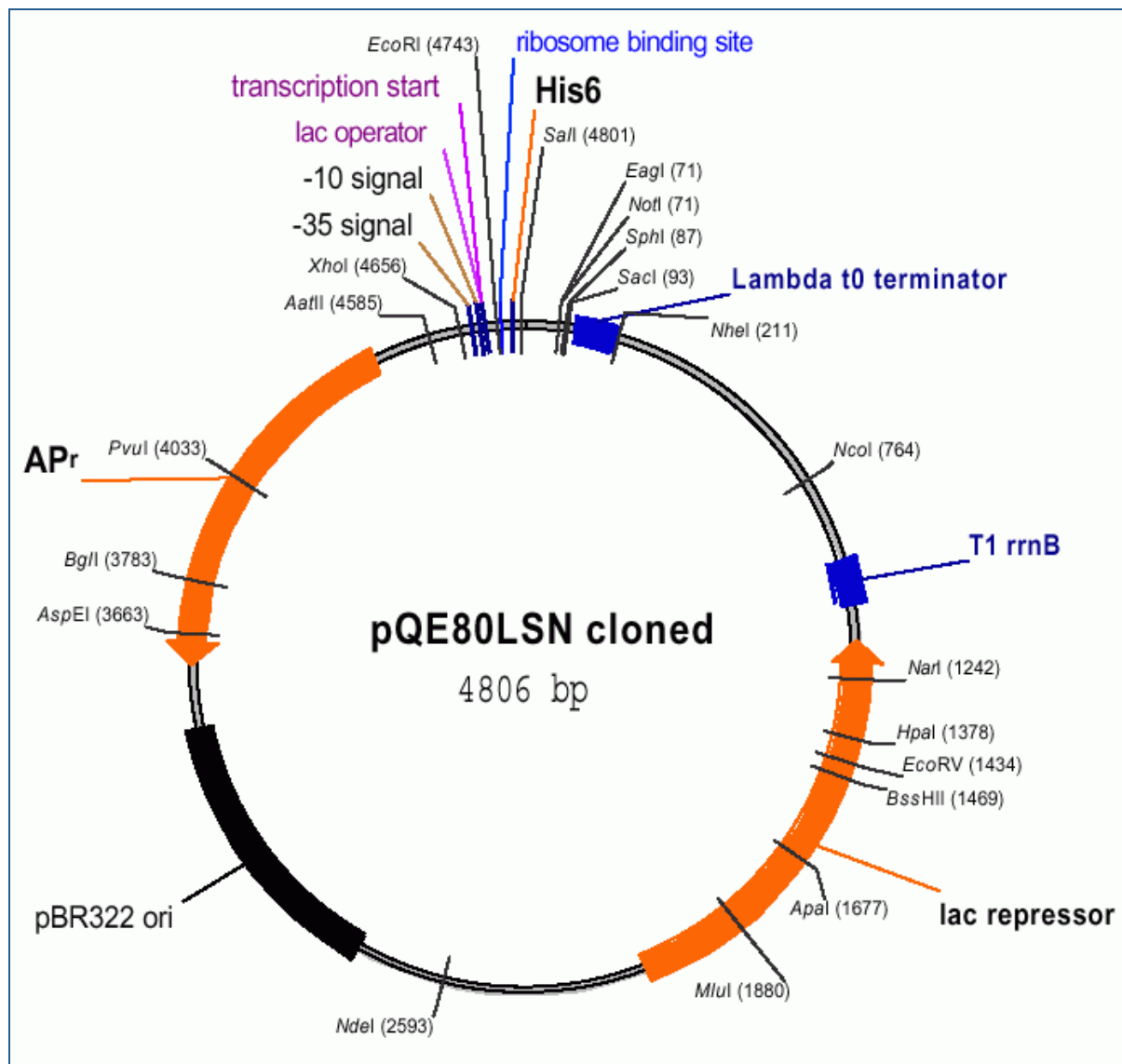


Figure 12. pQE80LSN vector containing an N terminal 6xHis tag and a *lac^q* gene ("www.imagenes-bio.de," 2012).

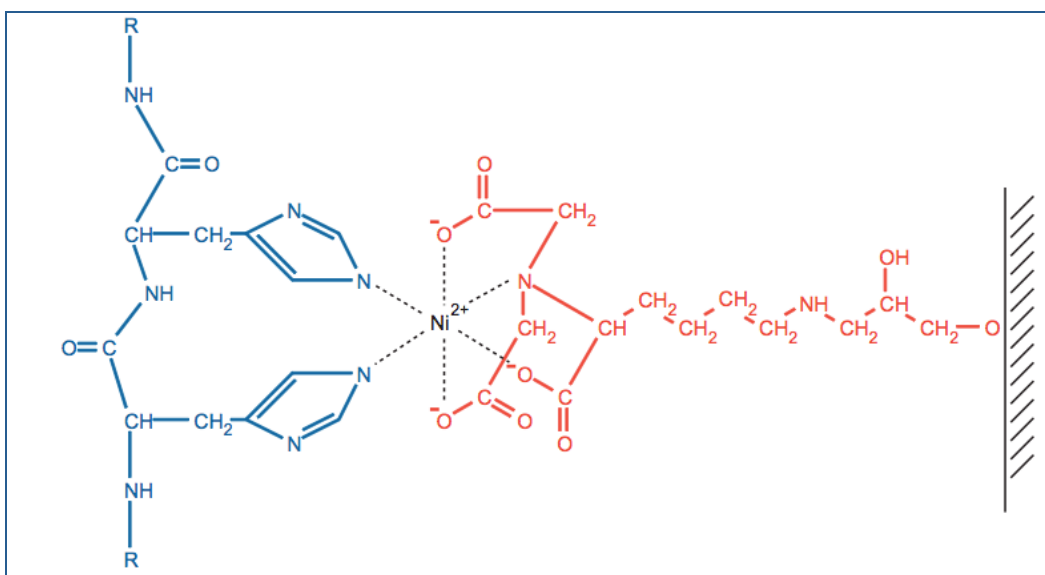


Figure 13. Polyhistidine tag (blue) and Ni-NTA (red) interaction through 2 free sites (2 black dashed lines on the left side of Ni^{2+}) on the Ni^{2+} ion. The Ni-NTA is attached to the agarose beads (black). (Qiagen, 2003).

6.1.3.2 Recombinant protein expression and purification

For use in expressing the *UniPEX-1* and the *642* (Syed, 2012) and *642+H* (Stempfer et al., 2010) clones, the Auto-induction medium was prepared and autoclaved as follows: solution 1: 50X consisting of 25% Glycerol, 2.5% Glucose and 10% α -Lactose, solution 2: 20X with 1M Na_2HPO_4 , 1M KH_2PO_4 and 0.5M $(\text{NH}_4)_2\text{SO}_4$, solution 3: 1000X with 1M MgSO_4 and TB medium with 12g Tryptone, 24g Yeast extract, 4ml Glycerol with pH adjusted to 7.5. All three solutions 1, 2 and 3 and TB medium are autoclaved prior to use. The final medium consisted of 100 $\mu\text{g}/\text{ml}$ Ampicillin and 15 $\mu\text{g}/\text{ml}$ Kanamycin added, after filtration through 0.2 μm syringe filter (Thermo Scientific, Cat# 190-2520), to 20ml solution 1, 50ml solution 2, 1ml solution 3 and 929ml TB medium.

The IPTG-induction medium, for expressing *UniPEX-2* clones contained solution 1: 50X consisting of 25% Glycerol, 2.5% Glucose and, solution 2: 20X with 1M Na_2HPO_4 , 1M KH_2PO_4 and 0.5M $(\text{NH}_4)_2\text{SO}_4$, solution 3: 1000X with 1M MgSO_4 and TB medium with 12g Tryptone, 24g Yeast extract, 4ml Glycerol pH=7.5. All three solutions 1, 2 and 3 and TB medium were autoclaved prior to use. The final medium consisted of 100 $\mu\text{g}/\text{ml}$ Ampicillin after filtration

through 0.2µm syringe filter (Thermo Scientific, Cat# 190-2520) added into to 20ml solution 1, 50ml solution 2, 1ml solution 3 and 929ml TB medium.

1.2ml of induction medium was pipetted into pre-weighed deep 96-well microtiter plates and 3µl of *E. coli* glycerol stock was inoculated. The inoculated plates were sealed with AirPore™ permeable sheets (Qiagen, Cat# 19571) and incubated at 37°C for 8 hours, then for auto-induction 16 hours at 30°C shaking at 1000rpm, for IPTG induction 30°C for 12 hours shaking at 1000rpm, and later were added 90ul of 100mM IPTG solution and incubated 4 more hours at 37°C. The induced cultures were centrifuged at 3500rpm for 10 minutes and the supernatant was discarded. The wet biomasses of the cultures are determined by weighing the plate and subtracting the mass of the empty plate weighed earlier.

For protein purification from cultures, the bacterial pellets are added 90 µL of 50 mM Tris-HCl pH=7.5, with a multi-channel pipette, for lysis and resuspended by vortexing 10 seconds at high speed. The plates were rapidly frozen by putting in liquid nitrogen for 10-15 seconds and thawed in cold water. Then, 10 µL Enzyme-Mix (30KU/µl rLysozyme, 25U/µl Benzonase Nuclease, 1mM PMSF and 2mM MgCl₂) in 50mM Tris-HCl pH=7.5 was added and shaken for 30 min at 37°C 1000rpm. In between the plates were sonicated 3 times for 1 min each with 100% ultrasonic. Plates were rapidly frozen in liquid nitrogen and thawed by sonicating with 100% for 3 minutes; freeze thaw was repeated twice. Afterwards, 20µl 1M NaOH was added in each well and shaken for 10 minutes at 1000rpm, and added 400µl Denaturation buffer, pH=7 (8M Urea, 50mM pH=7.12 Na-P-Buffer, 20mM Imidazole and 0.01% SDS). The solution was mixed by shaking for 30 minutes at 1000rpm. Plates were centrifuged for 45 minutes at 4600rpm at 20°C to separate proteins from cell debris and transfer supernatant containing protein into empty deep 96-well plate. In the meanwhile, Ni-NTA agarose beads were equilibrated by preparing 10ml beads per 50ml Falcon tubes, washing with 3X volume Wash buffer and adding Wash buffer until 25% slurry (10ml light blue colored Ni-NTA beads and 30ml Wash buffer) was obtained. After equilibration, 200µl 25% Ni-NTA-agarose was added into each well, while constantly mixing Ni-NTA agarose in between. Suspension was shaken for 1 hour at 1000rpm then transferred into deep 96-well filter plates and centrifuged for 2 minutes at 500 mg. Plates

were washed three times with 400µl wash buffer (pH=7, 6M Urea, 50mM pH=7.12 Na₂HPO₄, 0.3M NaCl and 20mM Imidazole) and centrifuged each time for 2 minutes at 500 xg. Finally filter plates were placed onto collection plates, 75µl elution buffer, pH=7 (50mM pH=7.12 Na₂HPO₄, 0.3M NaCl, 500mM Imidazole, 0.01% SDS and 0.01% NaN₃) was added to each well and centrifuged for 2 minutes at 500 xg to collect protein eluates. The purified protein eluates were stored at +4°C until determining protein concentration and at -20°C for longer storage. Pipetting during protein purification has been carried out with multi-channel pipettors.

The Na-P-Buffer was prepared by titrating 1M Na₂HPO₄ with 1M NaH₂PO₄ to obtain pH of 7.12. The 0.01% SDS and 0.01% NaN₃ was added to Denaturation buffer or Elution buffer after filtration through 0.2µm syringe filter (Thermo Scientific, Cat# 190-2520). The Denaturation buffer, Wash buffer and Elution buffers were filtered with 0.22µm Nitrocellulose membrane (Millipore, Cat# R0HA46038) before use.

6.1.4 Protein quantification

Concentration of purified IgG's was measured with the Protein A280 protocol of the *Take 3 Session* on the *Epoch Microplate Spectrophotometer*. 2µl of protein elution buffer was read as blank and 2µl of protein sample was read twice. The concentration values obtained from reading were corrected, as the human IgG absorbance relative to BSA standard has a ratio of 0.6 (Thermo Scientific, 2010b, p. 9).

The purified recombinant protein concentrations for protein microarray printing were measured with the Coomassie (Bradford) Protein Assay Kit (Thermo Scientific, Cat# 23200). Bradford Assay standards (Thermo Scientific, Cat# 23209) were serially diluted to 125µg/ml, 250µg/ml, 500µg/ml, 750µg/ml, 1000µg/ml, 1500µg/ml and 2000µg/ml concentrations and measured in triplicates. 5µl of the calibrated standard solutions or 5µl of the purified protein eluates was added into 150µl Coomassie (Bradford) reagent in flat bottom 96-well ELISA microplates (Greiner Bio-one, Cat# 655001) and mixed shortly by vortex at 500rpm. The solutions were incubated at room temperature for 10 minutes and absorbances were read at 595nm with the Protein A280 Microplate protocol on the *Epoch Microplate Spectrophotometer*. A standard curve was prepared by plotting the average blank-corrected 595nm measurement

for each BSA standard against its concentration in $\mu\text{g/ml}$. A linear regression was plotted for the entire set of standards and the linear fit equation was used for calculation of protein concentration from their absorbances.

6.1.5 SDS-PAGE

As quality control for purified IgGs, NuPAGE® Novex 4-12% Bis-Tris 1.0mm, 17 well SDS-PAGE was performed. Samples contained $2\mu\text{g}$ of IgG, $2.5\mu\text{l}$ NuPAGE® LDS buffer (4X) and dH_2O upto $10\mu\text{l}$ total volume. IgG's were denatured at 70°C for 10 minutes and loaded onto gel wells. PAGE was assembled in Vertical Mini-Gel Electrophoresis chamber and filled with 1X NuPAGE® MOPS SDS Running buffer (1X). $10\mu\text{l}$ of SeeBlue® Plus2 pre-stained standard was used as protein size ladder. Electrophoresis was performed for 50 minutes, 125mA per gel with 200V constant. The gels were stained with 1X Roti®-Blue staining solution (1X Methanol and 3X dH_2O) overnight shaking at 500rpm and destained with dH_2O three times washing for 1 hour each while shaking at 500rpm.

To verify recombinant protein purification, NuPAGE® Novex 4-12% Bis-Tris 1.0mm, 17 well SDS-PAGE was performed with additional reagents such as NuPAGE® Reducing agent in sample loading solution and NuPAGE® Antioxidant in gel running buffer. Samples contained $5\mu\text{l}$ of protein eluate, $2.5\mu\text{l}$ NuPAGE® LDS buffer (4X), $1\mu\text{l}$ NuPAGE® Reducing agent (10X) and $1.5\mu\text{l}$ dH_2O . The sample preparations were denatured at 70°C for 10 minutes and loaded onto gel wells. PAGE is assembled in a Vertical Mini-Gel Electrophoresis chamber, filled with 1X NuPAGE® MOPS SDS Running buffer (1X) and added $500\mu\text{l}$ NuPAGE® Antioxidant. Electrophoresis, gel staining and destaining procedures were done according to procedures mentioned above for IgG protein quality control with SDS-PAGE.

6.1.6 Protein microarray production

6.1.6.1 Principle

Microarrays that contain thousands of immobilized proteins offer the possibility to study aspects such as expression, binding, interactions, folding and specificity in a highly parallel, miniaturized and automated manner. As proteins are more complex than DNA by having different structures, polarities, hydrophobicities, sizes and charges, they can interact with the

surface chemistry they are spotted onto, unfold and lose their activity, or lose binding ability to the partner molecules. The type of surface to use for protein microarray experiments should fulfill the following criteria: high binding capacity, retain ability of protein, low variability between slides, high signal-to-noise ratio and long shelf life.

Covalent binding of proteins to functional groups on array surface allows strong, permanent binding and will be persistent during stringent washing conditions. The most common functional groups used for protein microarray immobilization are epoxides, aldehydes and *N*-hydroxy succinimidyl esters (NHS esters) that allow for random covalent binding through reactive groups (Kusnezow & Hoheisel, 2003; Seurnynck-Servoss, Baird, Rodland, & Zangar, 2007). Epoxy-coated slides are derivatized from epoxysilane and proteins bind to the epoxide ring with their amine, hydroxyl and thiol groups (Figure 14) (Arrayit, 2012). Several studies have reported epoxy surfaces to possess lower background and high signals with low spotted proteins and higher sensitivity compared to other covalent binding reactive group surfaces (Kusnezow & Hoheisel, 2003; Li & Reichert, 2002; Seong, 2002). The current study employed *ARChip-Epoxy* (Preininger, Bodrossy, Sauer, Pichler, & Weilharter, 2004), an epoxy based photoresist, that is ideal for microarray surface application due to advantages such as high stability in chemical, thermal and radiation exposures, biocompatibility, rigidity and inexpensiveness (Deepu, Sai, & Mukherji, 2008; Marie et al., 2006).

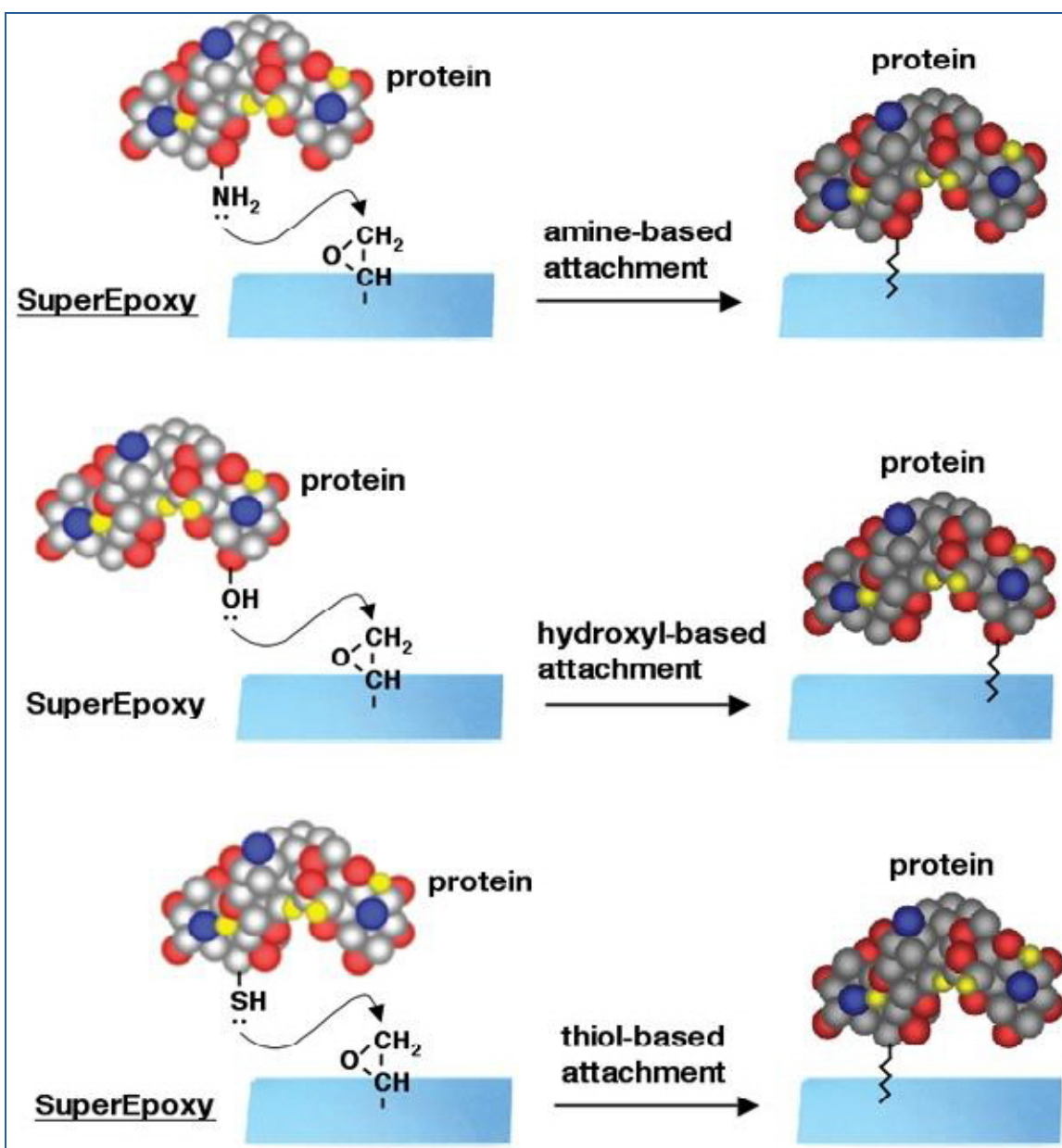


Figure 14. Proteins binding covalently to epoxide groups on the SuperEpoxy surface from ArrayIt® (Arrayit, 2012) through amine- (top), hydroxyl- (middle) and thiol- (bottom) groups. These reactive groups approach the epoxy group as a nucleophile, resulting in covalent coupling of the protein to surface.

Protein spotting technique is another vital factor in the final performance of protein microarrays. Two main approaches exist to deposit the proteins onto arrays: contact printing with pins that touch the surface and Piezo-electric non-contact printing. Among contact printing pins, split pins hold more samples, dispense controlled volumes and are faster

compared to solid tip pins (Figure 15 Hook et al., 2010; Kambhampati, 2004, p. 169). Flat tipped split pins exist that can prevent damage to slide surface by reducing the contact force. Protein spotting is carried out in humidity and temperature controlled, dust-free condition.

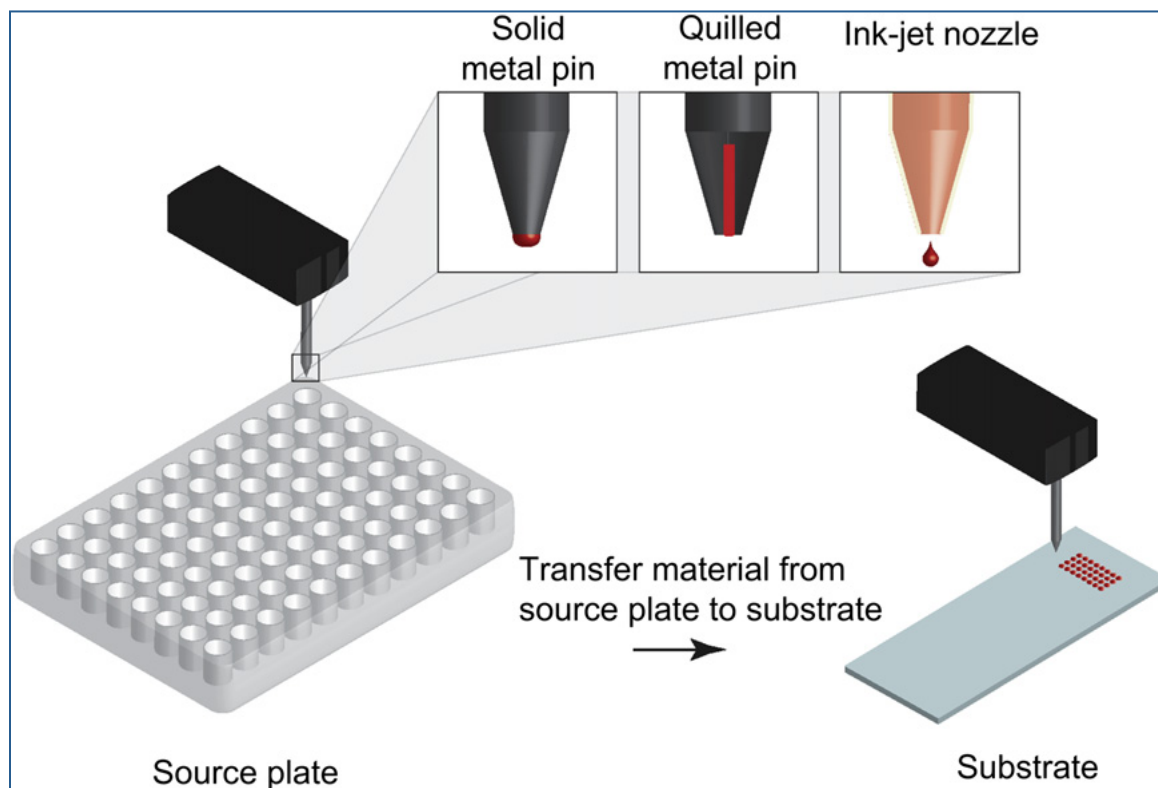


Figure 15. Microarray printing scheme from spotting or source plate via solid, quilled/split pins or ink-jet Piezo-electric tips onto the slide (Hook et al., 2010).

6.1.6.2 Procedure

16K microarrays, containing the expressed proteins from *UniPEX* library 15,770 clones, were printed with 48 946MP2 pins (ArrayIt®) in a contact printer the NanoPrint™ LM210 Platform (Telechem) with relative humidity set at 45%, as provided by Michael Stierschneider at AIT, Tulln, Austria. An array consisted of 48 subarrays, each containing 384 protein spots in source position from 384 well spot-plates (Biozym, Cat# 621521), in 4 by 12 format, each protein being printed in duplicates of 60µm diameter spots. Slides were scanned immediately after printing, using the Tecan laserscanner LS200 with full PMT at 532nm excitation for auto-fluorescence quality control of their usability for protein microarray experiments.

The *642* (Syed, 2012) and *642+H* (Stempfer et al., 2010) designated arrays are printed with 8 SMP3 pins in OmniGrid™ 100, GeneMachines by Ronald Kulovics at AIT, Muthgasse 11, Vienna, Austria. The *642 arrays* consist of 642 sero-reactive clones deduced by candidate marker screening experiment on macroarrays containing 38,016 human proteins, derived from a human fetal brain library (Büssow, Nordhoff, Lübbert, Lehrach, & Walter, 2000), with breast cancer and healthy control sera pools. The *642+H* arrays in addition to the *642* array clones, contain 76 potential tumor associated antigens derived from candidate marker SEREX (serological identification of antigens by recombinant expression cloning) and fetal brain cDNA expression library screenings of brain and lung cancer serum samples. *642* and *642+H* arrays were in spotted 4 identical subarrays, with each spots printed in duplicates. Spot positions on each subarrays are consistent with the protein source position on the 384 well spotting plate (Biozym, Cat# 621521). All arrays were spotted onto *ARChip epoxy* (Preininger et al., 2004) coated slides (T_O_P Oberflaechen GmbH) at 55% relative humidity.

6.1.7 Protein microarray assay

6.1.7.1 Principle

Protein microarray assay procedure follows the same principles as in conventional immunoassays such as ELISA: prior to applying the sample, the arrays are blocked to reduce background and non-specific binding to reactive surfaces around the spots. Several wash steps using buffer solutions to eliminate unbound blocking buffer content, without disturbing the spotted proteins. Protein interaction can be monitored by direct detection or via chemically, radioisotopically or fluorescently labeled proteins. Direct detection methods include mass spectrometry (MS), surface Plasmon resonance (SPR) and grating-coupled surface Plasmon resonance (GC-SPR), atomic force microscopy (AFM), micro-electromechanical systems (MEMS) cantilevers and quartz-crystal microbalance analysis (QCM). Labeled proteins are detected by direct, indirect or in sandwich assay approaches (Figure 16). The most commonly applied approach is sandwich assay, where samples containing target proteins are applied onto arrays allowing for binding with probe on arrays to occur. A secondary labeled detection antibody directed against the captured protein is used to detect this interaction partners. Fluorescence

labeling is frequently applied for protein microarray readout and is now the method of choice and among the different fluorophores Cy3 (green, 532nm excitation maximum) and Cy5 (red, 635 excitation maximum) are popular (Cretich, Damin, Pirri, & Chiari, 2006).

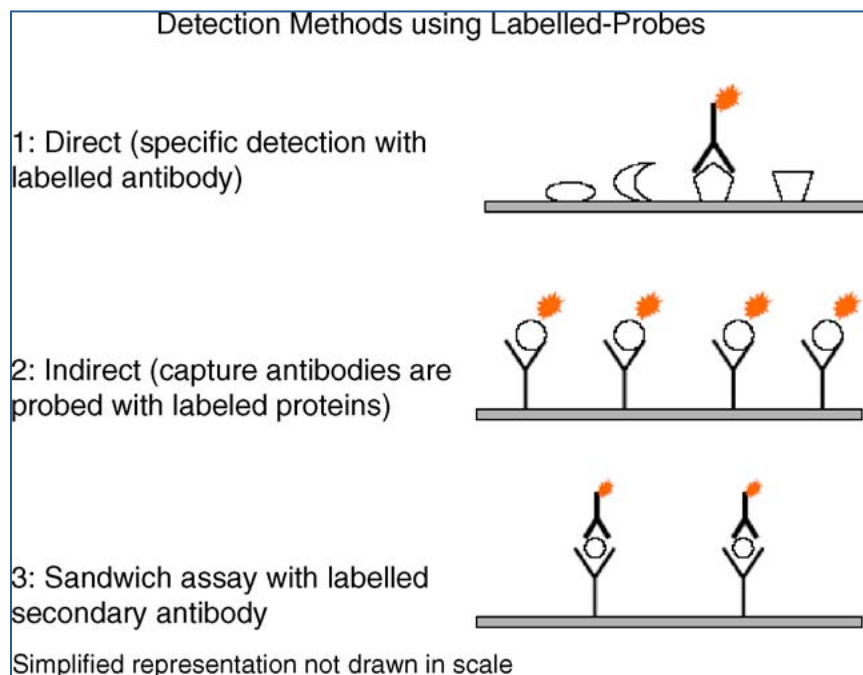


Figure 16. Labeled-probe protein microarray detection approaches (Cretich et al., 2006).

6.1.7.2 Procedure

Protocol 1. Protein microarray assays were performed with a static condition in a humidity chamber with the coverslip approach (the microarray slides were covered with a microscope glass coverslip (Paul Marienfeld, Cat# 0101242) after applying serum sample or detection antibody and put in a humid horizontal chamber). The protein microarrays were blocked using *DIG Easy Hyb* (Roche) blocking solution for 30 minutes, then washed three times for 5 minutes each with *PBSTw* (1X PBS, 0.01% Tween20) and rinsed once with dH_2O . Slides were spin dried for 2 minutes at 900rpm centrifugation. Serum or plasma samples were diluted 1:10 in *PBSTw* and 75 μl is applied onto arrays and covered with microscope slide coverslips. Samples are incubated 2 hours in horizontal humidity chamber and slides then washed three times for 5 minutes each with *PBSTw* and rinsed once with dH_2O . Slides were spin dried for 2 minutes at 900rpm centrifugation. Slides were then incubated with detection antibody *goat anti-human*

IgG Alexa Fluor® 647 (Invitrogen, Cat# A-21445) diluted 1:5000 in *PBSTw* for 1 hour in the humidity chamber. After wash, rinse and drying procedures as mentioned above, the slides were scanned using Agilent high-resolution microarray scanner with Red channel (Helium-Neon laser at 633nm) at 10-micron resolution. Scanned images are saved in TIFF (Tagged Image File Format) image format.

Protocol 2. Protein microarray assays were performed using Agilent Hybridization chamber and gasket slides. Each purified IgG samples from patient plasma were prepared to contain 0.3µg/µl protein concentration in *PBSTx* (1X PBS, 0.01% TritonX-100) with 3% milk powder (*PBSTx+MP3%*) and 500µl (0.3µg/µl IgG) of sample was applied onto gasket slide (Agilent Technologies, Cat# G2534-60005), the protein microarray was placed facing the sample and hybridization chamber (Agilent Technologies, Cat# G2534A) was assembled tightly. Assembly was checked for bubbles and rotated to ensure only a single bubble moves freely along the sample. After 4 hours of sample incubation at 12rpm at 25°C in hybridization oven (Agilent Technologies, Cat# G2545A), the assembly was taken apart, microarray slides were washed with *PBSTx* 3 times for 5 minutes each and rinsed with dH₂O. Microarray slides are then spin dried for 2 minutes at 900rpm centrifugation. *Goat anti-human IgG Alexa Fluor® 647* (Invitrogen, Cat# A-21445) was diluted 1:20,000 in *PBSTx+MP3%* and 500µl was applied to each microarray slides. Fluorescent labeling was done for 1 hour at 12rpm at 25°C and arrays were washed with *PBSTx* 3 times for 5 minutes each and rinsed with dH₂O. Microarray slides were then spin dried for 2 minutes at 900rpm centrifugation and were scanned using *Agilent high-resolution microarray scanner* with Red channel (Helium-Neon laser at 633nm) at 10-micron resolution. Scanned images are saved in TIFF (Tagged Image File Format) image format.

6.1.8 Image processing and data analysis

6.1.8.1 Principle

The scanned slide image obtained from protein microarray experiment is used to obtain intensity values corresponding to each feature on the array. This is achieved by using gridding files in .GAL file format generated during in-house printing or provided from the manufacturer of the array used. Using this grid information the subarrays, feature area and background are

determined followed by subtraction of the background intensity value from the median intensity value or the total intensity value of features.

The final intensity values then undergo transformation by the inverse/reciprocal method that transforms expression values into fold-change or logarithmic transformations that takes logarithm base 2 values. Both methods are appropriate to map up- or down-regulation of proteins, however, the following analyses for identifying differentially abundant proteins are problematic. Subsequently the transformed data are normalized due to variations in the experiments that may arise during microarray experiment procedures (Babu, 2004). Normalization methods range from centering all arrays at the median to *Quantiles* normalization that distributes all intensity values equally. *Quantiles* normalization reduces variability of arrays while retaining reproducibility among same conditions. Final step in the processing of microarray screening experiment data is to perform class classification or prediction analyses according to the class of samples, for example tumor vs non-tumor, and result in a group of features being differentially abundant among classes (Sánchez-Cabo, Rainer, Dopazo, Trajanoski, & Hackl, 2011).

Class comparison analysis is applied for comparing two or more predefined classes by determining the significance of individual proteins with univariate permutation tests. The univariate permutation test computes a t-test (two classes) or F-test (multiple classes) separately for each protein using the normalized log-intensities of the microarrays. The t-test is based on comparing the differences in mean log-intensities between classes relative to the variation expected in the mean differences and the variation is calculated assuming that all the samples are independent (Wright & Simon, 2003).

The aim of class prediction analysis is to develop models for utilizing gene expression profile to predict the class of future samples. The gene expression profile or predictors includes proteins that are differentially expressed between the predefined classes at the univariate parametric significance level less than the specified threshold. The proteins are ranked based on the extent to which they are individually differentially expressed among the classes and selected based on univariate discrimination ability. The proteins can also be selected with the

greedy-pairs method described by Bo and Jonassen (Bo & Jonassen, 2002). This approach ranks proteins based on their individual t-scores on the training set and selects the best ranked protein and finds the one other protein that in combination provides the best discrimination using as a measure the distance between of the two classes with regard to the two proteins when projected to the diagonal linear discriminant axis. The two selected proteins are removed from the protein set and the procedure is repeated on the remaining set until the specified numbers of proteins have been selected. In another approach, namely *Support Vector Machine Recursive Feature Elimination* (SVM RFE) that uses an SVM classifier trained on the data to rank proteins according to their contribution to the prediction performance. It utilizes a weighted linear combination of the protein expressions as a discriminator between the two classes and this is selected to maximize the margin, or the distance between the worst classified samples and the discriminant plane. After removing proteins having low absolute value of weight in the linear combination, a new SVM classifier is developed using the remaining proteins, then a new linear discriminant is found and the process continues iteratively until the desired number of protein is left (R. M. Simon, 2003).

The multivariate predictor for determining to which of the two classes a given sample belongs is achieved by several methods: the *Compound Covariate Predictor*, *Diagonal Linear Discriminant Analysis*, *Nearest Neighbor Predictor*, *Nearest centroid Predictor* and *Support Vector Machine Predictor*. The *Compound Covariate Predictor* is a weighted linear combination of log intensities for proteins that are univariately significant at the specified level and at a stringent significance level, proteins reduced in the multivariate predictor. The univariate t-statistics for comparing the classes are used as weights that are determined by the large values of log-intensities pre-disposing to one class (Radmacher, McShane, & Simon, 2002). The *Diagonal Linear Discriminant Analysis* differs from *Compound Covariate Predictor* as it ignores correlations among the proteins in order to avoid over-fitting the data. The *Nearest Neighbor Predictor* uses *Euclidean* distance as the distance metric and the vector of log-intensities as the expression profile. The nearest neighbor in the training set of the test samples is determined and the class of that nearest neighbor (with a similar expression profile) is taken as the prediction of the class of the test sample. There could be one or three nearest neighbors. For

the *Nearest Centroid Predictor*, the centroid of each training samples belonging to two classes are determined. The centroid is a vector containing mean log-intensities of the training samples. The distance of the expression profile for the test sample to each of the two centroids is measured and the test sample is predicted to belong to the class corresponding to the nearest centroid (R. M. Simon, 2003).

The *Bayesian Compound Covariate Predictor* selects the differentially expressed proteins for distinguishing two classes in a cross-validated training set, the compound covariate is computed which is the weighted average of the log expression values of the selected proteins, with the weights being the t-statistics of the differential expression in that training set. The means of the *Gaussian* distributions of the compound covariate scores of samples in each class in the training set differ among classes but the variances are assumed to be equal and a pooled estimate is used. The cross-validated misclassification rates of the multivariate class predictor are computed using Leave-one-out cross-validation (LOOCV) that omits one sample at a time and with each samples omitted, the entire analysis is carried out from the beginning. A list containing genes that are univariately significant on the reduced training sample is constructed and the accuracy of the prediction is recorded (R. M. Simon, 2003).

For more than two classes, *Binary Tree Prediction* is applied that combines the above mentioned prediction methods as the foundation. At each node of a binary tree the samples are classified into two subsets of classes with the possibility of one or both having multiple subclasses. The splitting of classes is accomplished with the fewest cross-validated misclassification errors and all the possibilities are tested to choose the best one that is accepted as a node of the binary tree. The resulting groups undergo the same procedure again until each group contains only one class (R. M. Simon, 2003).

6.1.8.2 Procedure

Scanned microarray TIFF image were loaded onto *GenePix® Pro Microarray Acquisition & Analysis Software version 6.0*. Images were viewed in red channel at 635nm laser range. The GAL file produced during printing of microarray slides was used to navigate the blocks and features to corresponding fluorescence spots. After all blocks and features were aligned, the

slides were analyzed and the results containing fluorescence intensity values were saved in a GPR file format. Fluorescence intensity-medians after subtraction of local background were obtained from the scanned array images and used for the data analysis.

Statistical analysis of the microarray experiments was performed using the BRB- ArrayTools software 3.8.1 [[http://linus.nci.nih.gov/ BRB-ArrayTools.html](http://linus.nci.nih.gov/BRB-ArrayTools.html)] developed by Dr. R Simon and Amy Peng Lam (R. Simon et al., 2007). Prior to analysis, data was normalized by *Quantiles* normalization and intensity values are Log2 transformed. Classifier models for discriminating patient groups were built using class prediction analysis with seven methods: *Compound Covariate Predictor*, *Diagonal Linear Discriminant Analysis*, *k-Nearest Neighbor* (using k=1 and 3), *Nearest Centroid*, *Support Vector Machines* and *Bayesian Compound Covariate Predictor*. LOOCV receiver operating curve (ROC) analyses from the *Bayesian Compound Covariate Predictor* were conducted and the *Area Under the Curve* (AUC) values were calculated.

7 Results

7.1 Tumor autoantibody candidate marker screening using Protocol 1

The current thesis commenced with tumor autoantibody candidate marker screening by utilizing the 16K protein microarray and the assay was performed according to the *Protocol 1* procedures described in Methods section 6.1.7.2, with 122 samples belonging to four clinical samples groups: Colon carcinoma (n=31), High risk polyps (n=31), Low risk polyps (n=30) and Healthy controls (n=28). The 16K protein microarrays were produced previously from the *UniPEX* library (purchased from ImaGenes, Source Bio-Science) containing 15,456 clones. The recombinant proteins encoded by each clones were expressed and purified according to section 6.1.3.2 and printing of the 16K protein microarray was carried out with settings given in section 6.1.6.2. The plasma samples were diluted in a 1:30 ratio and applied onto the 16K protein microarrays by covering with a glass microscope coverslip to distribute samples. Detection antibody *goat anti-human IgG Alexa Fluor® 647* (Invitrogen, Cat# A-21445) was used to detect protein binding and slide images were obtained by a fluorescent microarray slide scanner. The resulting values from foreground median intensity with background intensity subtracted, obtained after image analysis were used for statistical analyses. Data were normalized over the median of the reference array.

The order of the total 122 samples was randomized and processed in 3 runs on 3 different days, with 48 samples in each of the first two runs and 27 samples within the third run. Samples are displayed according to the experimental run to check for homogeneity using multidimensional scaling with *Euclidean Distance*. Samples are found to be moderately homogeneous as can be seen from figure 17, where overall clustering of samples within the different runs are in close proximity and no significant separation between runs can be observed.

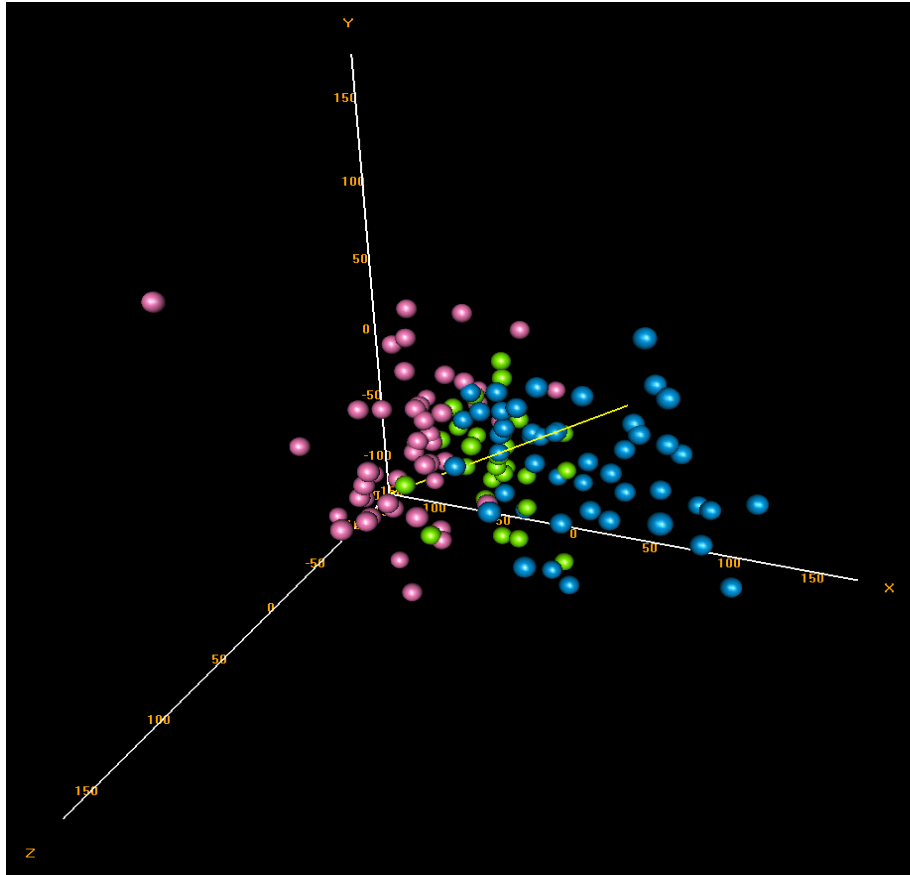


Figure 17. Multidimensional scaling of the three runs of the candidate marker screening experiment using *median over reference array* normalized intensity data. Clustering in 122 samples based on 15053 genes, using the first 3 components (covering 45% of total variation) obtained using the *Euclidean distance* metric. Blue- Run 1, Pink- Run 2 and Green- Run 3.

<i>Diagonal Linear Discriminant Analysis Classifier</i>					<i>3-Nearest Neighbor-Classifer</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Control	0.2	0.87	0.333	0.769	Control	0.133	0.609	0.1	0.683
High risk	0.29	0.846	0.391	0.778	High risk	0.226	0.484	0.13	0.647
Carcinoma	0.129	0.593	0.098	0.667	Carcinoma	0.129	0.418	0.07	0.585
Low risk	0.267	0.652	0.2	0.732	Low risk	0.133	0.533	0.085	0.653
<i>1-Nearest Neighbor-Classifer</i>					<i>Nearest Centroid Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Control	0.267	0.804	0.308	0.771	Control	0.2	0.848	0.3	0.765
High risk	0.258	0.758	0.267	0.75	High risk	0.387	0.813	0.414	0.796
Carcinoma	0.194	0.637	0.154	0.699	Carcinoma	0.129	0.626	0.105	0.679
Low risk	0.2	0.772	0.222	0.747	Low risk	0.267	0.707	0.229	0.747

Table 2. Performance of classifiers between the Colon cancer test set groups with 4 prediction analysis methods at $p < 0.001$ significance level.

Class prediction analysis was carried out to identify classifiers able to distinguish between the 4 groups of clinical samples, using multiple methods: *Nearest Centroid Predictor*, *Diagonal Linear Discriminant Analysis* and *K nearest neighbor* analysis with $K=1$ and 3. Genes were selected based on a significance level of $p<0.001$. The mean correct classification rate of the 6 classifiers comparing all methods was 22-25% after LOOCV. Sensitivity and specificity values ranged from 13-38% and 48-87%, respectively (Table 2). At $p<0.05$ significance threshold, 639 classifiers were deduced with 22-30% correct classification rate, 6-43% sensitivity and 53-81% specificity.

Class prediction analysis between colon carcinoma and healthy control groups was made with three additional methods, *Compound Covariate Predictor*, *Support Vector Machines* and *Bayesian Compound Covariate Predictor*, a total of 7 methods. Significance levels of classifiers were set at $p<0.001$ and only 1 feature had passed this criterion. With significance level of $p<0.05$, 404 classifiers were elucidated that had correct classification rate of 32-51% among all 7 methods, with 10-60% sensitivity and specificity (Table 3). Low performance was apparent with AUC of 0.295 after cross-validation from *Bayesian Compound Covariate Predictor* analysis, as can be seen from the ROC curve in Figure 18.

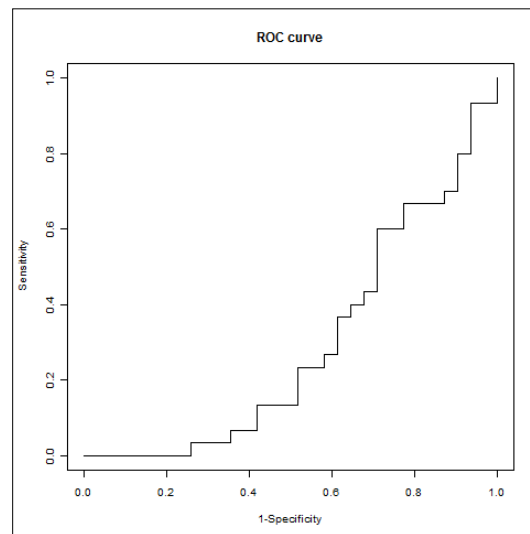


Figure 18. ROC curve of cross validation with *Bayesian Compound Covariate Predictor* analysis for Colon Carcinoma and Healthy Control groups. AUC=0.295.

<i>Compound Covariate Predictor</i>					<i>Diagonal Linear Discriminant Analysis Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Control	0.367	0.387	0.367	0.387	Control	0.367	0.387	0.367	0.387
Carcinoma	0.387	0.367	0.387	0.367	Carcinoma	0.387	0.367	0.387	0.367
<i>1-Nearest Neighbor Classifier</i>					<i>3-Nearest Neighbors Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Control	0.6	0.419	0.5	0.52	Control	0.5	0.355	0.429	0.423
Carcinoma	0.419	0.6	0.52	0.5	Carcinoma	0.355	0.5	0.423	0.429
<i>Nearest Centroid Classifier</i>					<i>Support Vector Machine Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Control	0.267	0.355	0.286	0.333	Control	0.333	0.516	0.4	0.444
Carcinoma	0.355	0.267	0.333	0.286	Carcinoma	0.516	0.333	0.444	0.4
<i>Bayesian Compound Covariate Classifier</i>									
Class	Sensitivity	Specificity	PPV	NPV					
Control	0.133	0.258	0.148	0.235					
Carcinoma	0.258	0.133	0.235	0.148					

Table 3. Performance of classifiers between the Carcinoma and Healthy control groups with 7 prediction analysis methods at $p < 0.05$ significance level.

By taking the Colon carcinoma group as one cancer class and combining all other three groups as non-cancer class, prediction analysis was done with the all 7 methods mentioned above. When feature selection criteria were set to $P < 0.001$, 15 reactive proteins (features) had fulfilled the criteria and their correct classification rate ranged from 58-67%; sensitivity and specificity ranged from 12-83% (Table 4) with an AUC of 0.525 indicating the failure of classification of all the samples (Figure 19).

At $P < 0.05$ significance level, class prediction analyses yielded 752 proteins that correctly classified 66-72% of samples into the two groups, with 3-95% sensitivity and specificity (Table 5) and AUC of 0.513 (Figure 20). The 95% sensitivity was obtained by *3-Nearest Neighbors Classifier* method.

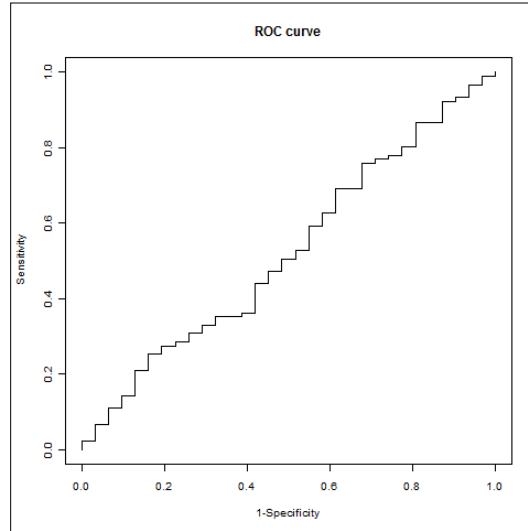


Figure 19. ROC curve of cross validation with *Bayesian Compound Covariate Predictor* analysis for Colon carcinoma group vs Healthy Control, High and Low risk polyps groups at significance level of $p < 0.001$. AUC=0.525.

<i>Compound Covariate Predictor</i>					<i>Diagonal Linear Discriminant Analysis Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
HR LR Control	0.692	0.323	0.75	0.263	HR LR Control	0.681	0.29	0.738	0.237
Carcinoma	0.323	0.692	0.263	0.75	Carcinoma	0.29	0.681	0.237	0.738
<i>1-Nearest Neighbor Classifier</i>					<i>3-Nearest Neighbors Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
HR LR Control	0.725	0.29	0.75	0.265	HR LR Control	0.824	0.226	0.758	0.304
Carcinoma	0.29	0.725	0.265	0.75	Carcinoma	0.226	0.824	0.304	0.758
<i>Nearest Centroid Classifier</i>					<i>Support Vector Machine Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
HR LR Control	0.692	0.323	0.75	0.263	HR LR Control	0.835	0.194	0.752	0.286
Carcinoma	0.323	0.692	0.263	0.75	Carcinoma	0.194	0.835	0.286	0.752
<i>Bayesian Compound Covariate Classifier</i>									
Class	Sensitivity	Specificity	PPV	NPV					
HR LR Control	0.363	0.129	0.55	0.065					
Carcinoma	0.129	0.363	0.065	0.55					

Table 4. Performance of classifiers between the Carcinoma as one class and Healthy control, High and Low risk polyps groups as the other with 7 prediction analysis methods at $p < 0.001$ significance level.

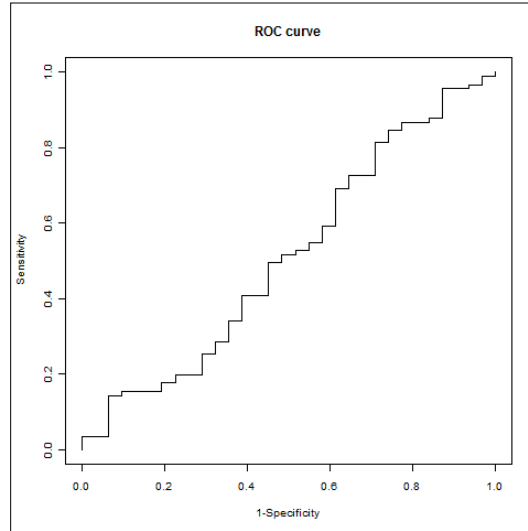


Figure 20. ROC curve of cross validation with *Bayesian Compound Covariate Predictor* analysis for Colon carcinoma group vs Healthy control, High and Low risk polyps groups at significance level of $p < 0.05$. AUC=0.513.

<i>Compound Covariate Predictor</i>					<i>Diagonal Linear Discriminant Analysis Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
HR LR Control	0.835	0.258	0.768	0.348	HR LR Control	0.813	0.258	0.763	0.32
Carcinoma	0.258	0.835	0.348	0.768	Carcinoma	0.258	0.813	0.32	0.763
<i>1-Nearest Neighbor Classifier</i>					<i>3-Nearest Neighbors Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
HR LR Control	0.857	0.097	0.736	0.188	HR LR Control	0.956	0.032	0.744	0.2
Carcinoma	0.097	0.857	0.188	0.736	Carcinoma	0.032	0.956	0.2	0.744
<i>Nearest Centroid Classifier</i>					<i>Support Vector Machine Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
HR LR Control	0.835	0.258	0.768	0.348	HR LR Control	0.824	0.258	0.765	0.333
Carcinoma	0.258	0.835	0.348	0.768	Carcinoma	0.258	0.824	0.333	0.765
<i>Bayesian Compound Covariate Classifier</i>									
Class	Sensitivity	Specificity	PPV	NPV					
HR LR Control	0.549	0.129	0.649	0.089					
Carcinoma	0.129	0.549	0.089	0.649					

Table 5. Performance of classifiers between the Carcinoma as one class and Healthy control, High risk polyps and Low risk polyps groups as the other with 7 prediction analysis methods at $p < 0.05$ significance level.

The above results are significantly poor and these failures were possibly due to the use of total plasma that could have matrix effect or heterophile antibodies interfering in the assay, which may be overcome by using purified IgG instead of total plasma. Several serious flaws and defects in the screening experiment results and array images could be viewed as depicted in Table 6 and Figure 21.

Nº	Type of defect	Possible cause
1	Low classification performance in screening experiment	Interference of plasma matrix or heterophile antibodies
2	Uneven sample distribution and smearing	Protein microarray assay condition during sample incubation
3	Uneven spot morphology Merging of spots	Microarray printing error

Table 6. Type of defects on the processed arrays observed during the candidate marker screening experiment with test set plasma samples and their possible causes.

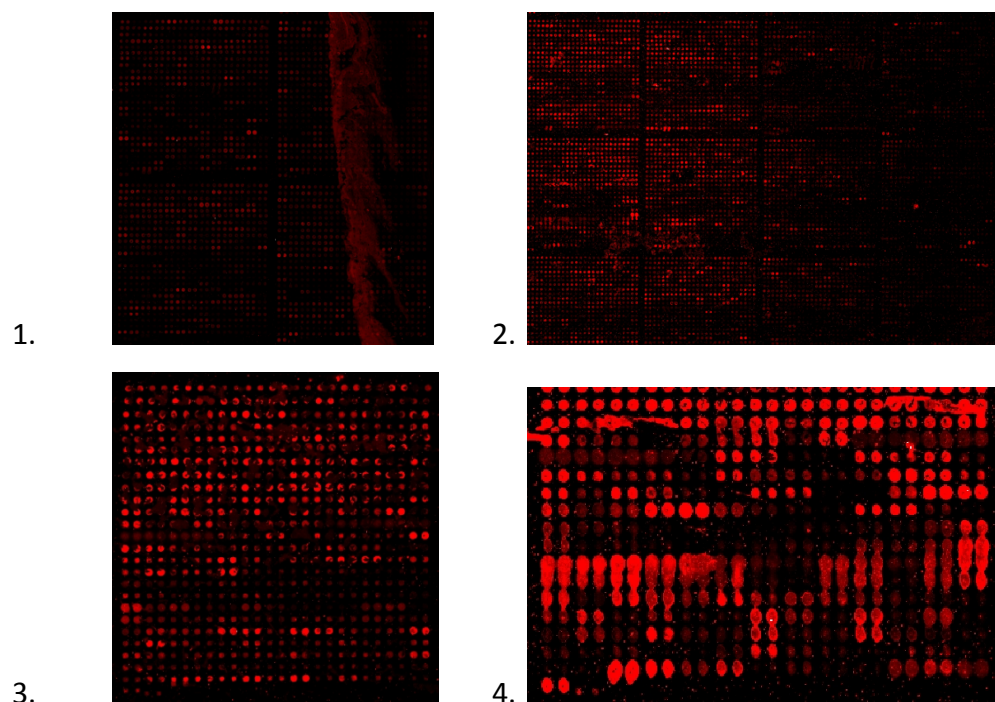


Figure 21. The slide images displaying defects observed from candidate marker screening experiment performed according to the *Protocol 1*. Images 1 and 2 correspond to the defect number 2 listed in table 2 and images 3 and 4 to defect number 3, respectively.

Due to these technical issues at hand, improvements in the Standard Operating Procedure of protein microarray assays with clinical samples needed to be reviewed and optimized for increasing quality and reproducibility of arrays and results. Taking into consideration, the nature of the defects, several optimization experiments were planned aimed to overcome these problems.

7.2 Protein microarray assay optimizations

7.2.1 Surface chemistries and blocking conditions

The foundation for any microarray experiment is the surface chemistry and for a reliable and reproducible result it is vital to choose the correct one for the particular type of probe that is used, the target to be captured and the detection methods applied. This holds true especially in cases of high-throughput experiments dealing with thousands of different probes, where maximal compatibility of samples is intended. To address whether in our experiment, utilization of an alternative surface could improve analyses, we tested several commercially available slides in parallel to in-house developed surface- *SU8 epoxy* along with 4 different blocking conditions (Table 7). Blocking conditions affect the specificity of binding and increase the true signal intensities of the features by minimizing the background (unspecific bindings). With the best combination of surface and blocking, we could significantly improve the performance of our experiments. For each optimization experiments, *642+H* arrays were used. *Protocol 1* was used for this experiment with the above-mentioned adjustments for blocking conditions and instead employed serum sample. The serum samples and detection antibodies are applied onto arrays under coverslips and incubation is carried out in horizontal humidity chamber that lack mixing of samples over arrays.

No	Slides
1	NEXTERION® Slide E
2	Corning® GAPS Amino-silane
3	PolyAn® 3D-Aldehyde
4	PolyAn® 3D-Epoxy
5	Corning® Epoxide
6	NEXTERION® Slide A
7	Anopoli® VEPO
8	ArrayIt® Aldehyde
9	SU8 Epoxy
No	Blocking reagents
1	DIG Easy Hyb
2	3M Urea+ 0.1% SDS
3	PBSTw
4	PBSTw+MP3%

Table 7. List of tested slides and blocking reagents.

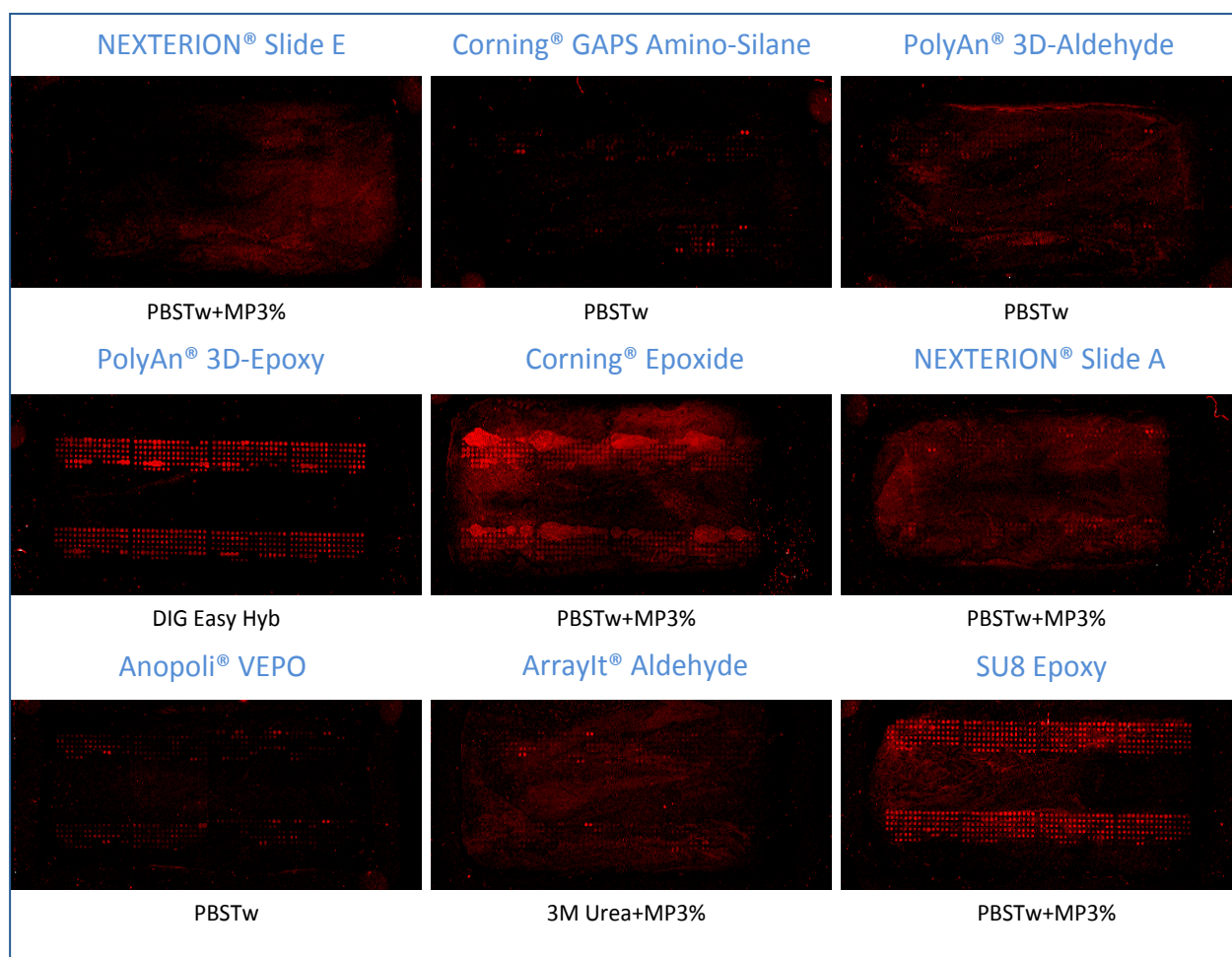


Figure 22. Representative Images of the nine different slides tested. Each subarray on the slides is blocked with 4 different buffers. Aside from slides PolyAn® 3D-Epoxy and SU8 Epoxy, other eight tested slides are prone to smearing or have features with low intensities.

Slides are scanned with 90% PMT (Photomultiplier tube) level and viewed using *GenePix® Pro Microarray Acquisition & Analysis Software version 6.0* with 92% brightness and 95% contrast settings. With the current protocol, two types of epoxy surfaces, namely *SU8 epoxy* and *PolyAn® 3D-Epoxy* performed better compared to the other seven tested slides (Figures 22), with low background and high feature intensities. As with blocking reagents, *DIG Easy Hyb* and *PBSTw+MP3%* (1X PBS, 0.01% SDS, 3% Milk powder) has less smearing compared to 3M Urea with 0.1% SDS or PBSTween20 without milk powder.

Surface *SU8 epoxy* was taken further with the two types of blocking reagents to determine the level of 6xHis-tagged fusion proteins spotted on the 642 array using *Penta-His Alexa Fluor® 532* conjugate. From slide images in figure 23 (left) the two blocking reagents (*DIG Easy Hyb*

and *PBSTw+MP3%*) are quite comparable and from the green (532nm) foreground median (F532 Median) and background (B532) signal intensity values obtained from both blocking conditions, *DIG Easy Hyb* shows less deviation in the foreground signal of features (Figure 23 right). Average F532 median intensity is increased by around 1000 in *PBSTw+MP3%* blocked slides as well as the B532 intensity by around 400 (Figure 23 right).

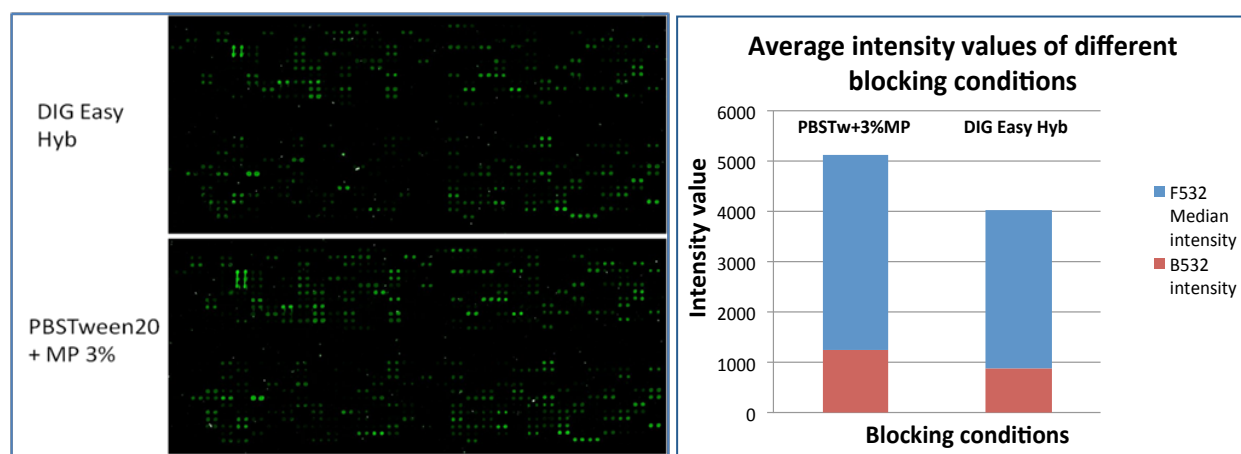


Figure 23. Left. *Penta-His Alexa Fluor® 532* (green) detection of 6xHis tagged fusion proteins on *SU8 642* arrays blocked with *DIG Easy Hyb* and *PBSTw+MP3%*. **Right.** Average values of 532nm foreground median and background intensities of *SU8 642* arrays blocked with *DIG Easy Hyb* and *PBSTw+MP3%*.

It could be concluded from the above results shown in figures 22 and 23 that the utilization of *PolyAn® 3D-Epoxy* or *SU8 epoxy* surfaces for spotting our expressed proteins from the *UniPEX* cDNA library would be feasible. Nonetheless, *SU8 epoxy* slides are preferable since these are cheaper in comparison to the commercial product *PolyAn® 3D-Epoxy* and the conditions defined, could improve the performance of TAA detection.

7.2.2 Assay conditions

A very important factor in protein microarray is sample and reagent homogeneity on the arrays, for reducing uneven fluorescence and smearing effects on the slides. Several systems exist that allow for homogenous mixing, retain humidity and contamination of applied samples during microarray probing procedure, such as *OneArray®* Microarray Hybridization Chambers (Phalanx Biotech), *Corning®* Microarray Hybridization Chambers, and Microarray hybridization

Chamber Kits (Agilent® Technologies). The conventional method of probing in horizontal humidity chamber may possibly be disadvantageous, as the samples cannot dispense over the array, resulting in inhomogeneous distribution. The Microarray hybridization Chamber Kits (microarray hybridization gasket slides and hybridization oven) from Agilent® Technologies in comparison to horizontal humidity chamber were tested for sample homogeneity and binding. Additionally sample incubation durations were tested to observe changes in protein binding on microarrays.

The rotating (Agilent® microarray hybridization oven) and static (horizontal humidity chamber) conditions were compared using test serum samples for durations of 2, 4 and 16 hours of sample incubation. The experiment was carried out according to *Protocol 1* for static condition and the Agilent Hybridization chamber and gasket slides were employed for the rotating condition, using *642+H* arrays. Extending the duration of sample incubation could potentially increase binding. Results suggest that under both static and rotating conditions of sample incubation from 2 to 4 and 16 hours results in significant increase in signal intensities of features (Figure 24). However, under at 4 and 16 hours of incubation under rotating conditions, several features are saturated, with start of saturation taking place between 2 to 4 hours and clearly increasing at 16 hours incubation (Figure 24 Left. white features). This suggests that sufficient binding of sample antibodies with capture antigens already takes place at 4 hours. The rotating condition evidently yields more efficient binding of the sample proteins (consequently detection antibodies) with the antigens on arrays. The scatter plot of log₂ transformed signal intensities of arrays, processed under rotating and static conditions reveals that under rotating conditions, majority of features have increased signal intensities (Figure 25) compared to static conditions. This is apparent by the higher density of the plots towards the x-axis representing the arrays of rotating condition. Therefore, protein microarray binding experiment should be carried out with adequate mixing of samples to ensure movement of target and capture proteins.

To test the reproducibility of the performance of the binding experiment carried under the rotating condition, the correlation coefficient of log₂ transformed intensity values of all

features on the arrays, for each of the 4 replicates of the test samples were computed with the *Pearson* correlation method (Figure 26). The mean correlation coefficient is calculated from individual correlations among the sample replicates within each incubation time-point groups. With increased sample incubation time, correlations are slightly increasing: at 2 hours $r=0.9$, 4 hours $r=0.91$ and 16 hours $r=0.915$. Based on the visualization of arrays (Figure 24), comparison of signal intensities (Figure 25) and reproducibility of experiments (Figure 26), the decision to modify the protein microarray binding experimental condition with rotating in the hybridization oven and sample incubation for 4 hours was made.

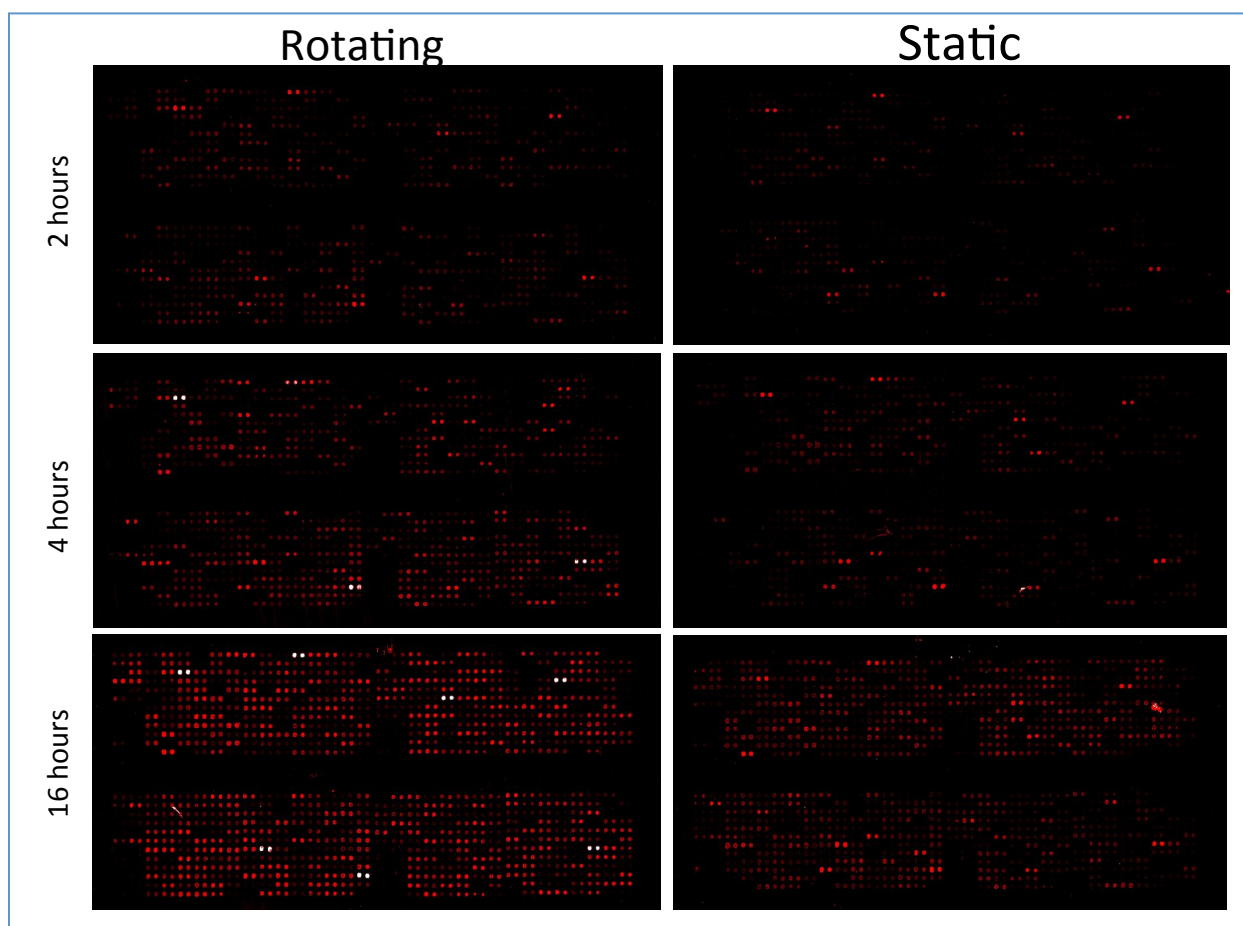


Figure 24. Incubation of serum samples in static and rotating conditions for 2, 4 and 16 hours. The signal intensities of features are increasing as the sample incubation time increases from 2 to 4 and 16 hours in both static and rotating conditions. After 16 hours of sample incubation, several features are seen to be oversaturated (white features) for rotating condition. Saturation level of highly reactive features initially appears at 4 hours of sample incubation in rotating condition.

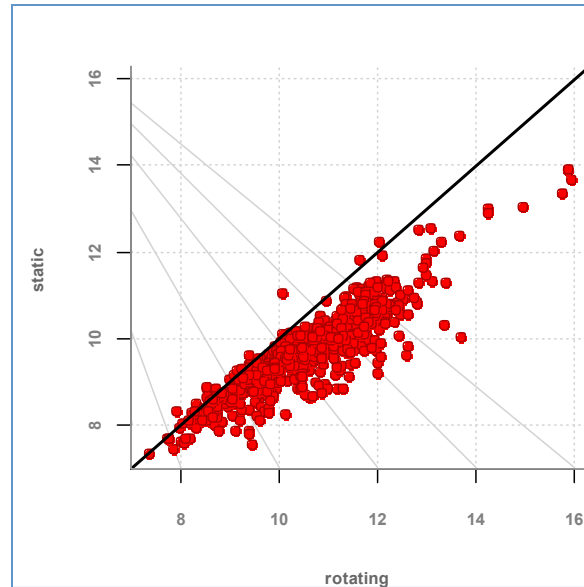


Figure 25. Scatter plot of Log2 transformed intensity values from arrays processed under rotating and static conditions, with all sample incubation time points combined. Overall signal intensities are significantly increased in the rotating condition of microarray experiment.

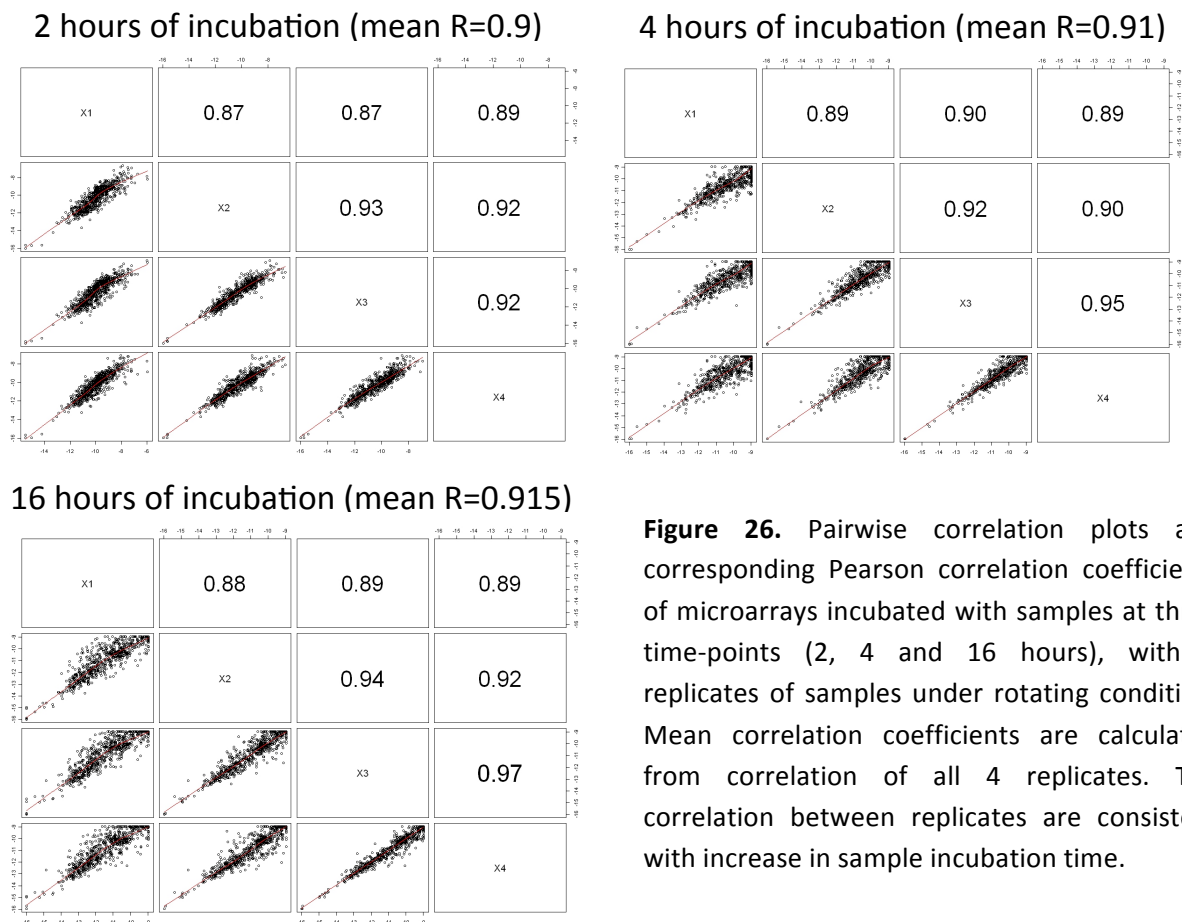


Figure 26. Pairwise correlation plots and corresponding Pearson correlation coefficients of microarrays incubated with samples at three time-points (2, 4 and 16 hours), with 4 replicates of samples under rotating condition. Mean correlation coefficients are calculated from correlation of all 4 replicates. The correlation between replicates are consistent with increase in sample incubation time.

7.2.3 Blocking agent in sample and detection antibody dilutions

In several experiments, it was observed that using *PBSTw* (1X PBS, 0.01% Tween20) buffer as blank or negative control yielded fluorescent features on arrays with *goat anti-human IgG Alexa Fluor® 647* detection. This was taken as an indication that samples or detection antibody dilution conditions are lacking a blocking component, for example the most commonly used BSA or Milk powder, that eliminate high background and unspecific binding by blocking the reactive surface sites. As to compare with two different serum samples, PBST buffer was used as blank or negative control and the probing was carried out as stated in *Protocol 1* under rotating condition, sample incubation 4 hours, with the inexpensive alternative milk powder (3%) added to either sample or detection antibody dilution buffer or both. In addition to detecting captured antibodies with *goat anti-human IgG Alexa Fluor® 647 antibody* (diluted 1:5000), *Penta-His Alexa Fluor® 532* detection antibody (diluted 1:500) was used to detect 6xHis tagged fusion proteins. The dual color (red- human IgG; green- 6xHis-tag) experiment will clarify the performance of blocking effects of milk powder during probing with sample or blank controls.

For serum samples diluted with *PBSTw* of detection antibodies *goat anti-human IgG Alexa Fluor® 647* and *Penta-His Alexa Fluor® 532*, only captured IgG's and not the 6xHis-tag of fusion proteins was detected regardless of the sample dilution with *PBSTw+MP3%* (1X PBS, 0.01% Tween20, 3% Milk powder) (Figure 27 A). Conversely, when detection antibodies were diluted with *PBSTw+MP3%*, both IgG's in samples and 6xHis-tag on arrays were detected, giving a dual-color fluorescence for samples diluted in both *PBSTw* and *PBSTw+MP3%* (Figure 27 B). Result was consistent with buffer blanks, where addition of milk powder in detection antibody dilution buffer was essential (Figure 27 C and D). It was concluded from figure 29 that milk powder is an indispensable component in detection antibody dilution buffer.

As many studies utilized milk powder as a component in their sample dilution buffer (Carlsson et al., 2011; Haab et al., 2001; Ingvarsson et al., 2007; Stempfer et al., 2010) and our results are suggestive of the reduced number of fluorescent features (Figure 27A), it was decided to employ milk powder in both sample dilution and detection antibody dilutions.

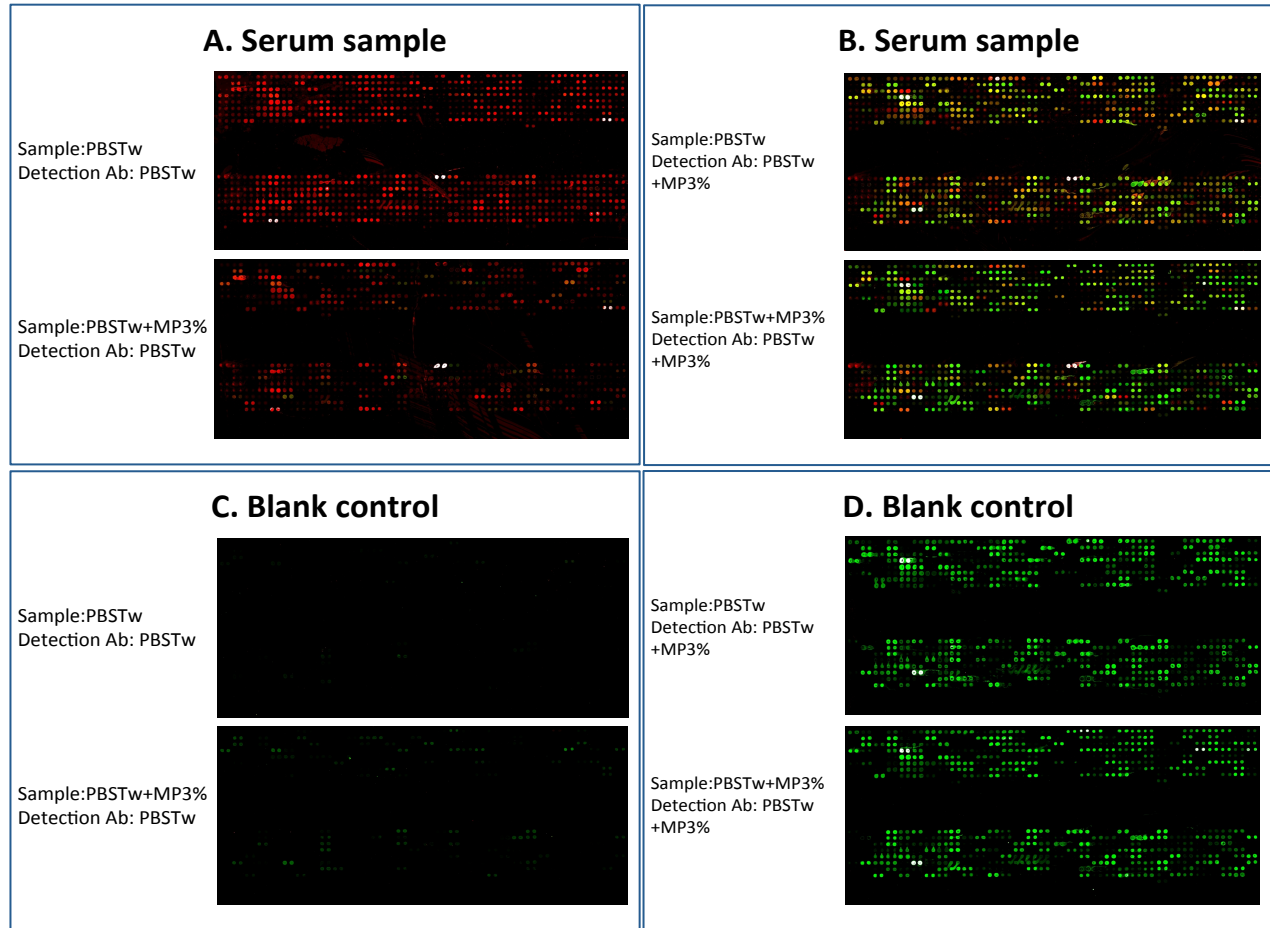


Figure 27. Effect of milk powder in buffer for dilutions of serum sample and detection antibodies *goat anti-human IgG Alexa Fluor® 647* (Red) and *Penta-His Alexa Fluor® 532* (Green), with incubation at rotating conditions. **A.** Serum sample diluted in buffer without (Top. *PBSTw*) and with (Bottom. *PBSTw+MP3%*) milk powder and detection antibody diluted in *PBSTw*. **B.** . Serum sample diluted in buffer without (Top. *PBSTw*) and with (Bottom. *PBSTw+MP3%*) milk powder and detection antibody diluted in buffer *PBSTw+MP3%*. **C.** *PBSTw* (Top) or *PBSTw+MP3%* (Bottom) buffers applied as blank control and detection antibody diluted in *PBSTw*. **D.** *PBSTw* (Top) or *PBSTw+MP3%* (Bottom) buffers applied as blank control and detection antibody diluted in *PBSTw+MP3%*. As seen in A and C, milk powder addition to samples reduces unspecific binding. Comparing A with B and C with D, presence of milk powder in dilution buffer for detection antibody enhances specific binding, making it possible for a dual color detection of both bound IgG's (Red) and the 6XHis-tagged recombinant proteins (Green) on arrays, that is otherwise not possible as indicated in the sections A and C.

7.2.4 Comparison of Buffer Detergents

Another feature within the protein microarray procedures is washing of slides between different steps of incubation and the contents of these washing buffers. Wash buffers for immunoassays or protein microarray experiments generally contain phosphate or Tris buffered saline and detergent, most commonly used are PBS or TBS and nonionic surfactants like 0.01% Tween 20. Tween 20 is a polyoxyethylene surfactant applied for protein immunoassays to reduce surface tension. While Tween 20 is a relatively mild detergent, an alternative Triton X-100, a polyethylene oxide, is also used for protein and peptide microarray washing steps for its ability to enhance penetration of spotted proteins. The *Protocol 1* was used with several modifications: The performances of these two detergents in wash buffer PBS at a concentration of 0.01% were tested and are designated *PBSTx* for Triton X-100 and *PBSTw* for Tween 20 containing buffers. Serum samples and detection antibodies are diluted with buffers with addition of 3% milk powder (*PBSTx+MP3%* and *PBSTw+MP3%*). For each buffer type used (*PBSTx* and *PBSTw*), 2 serum samples are used in 4 repeats. Samples are incubated for 4 hours and the binding procedures are carried out at rotating condition.

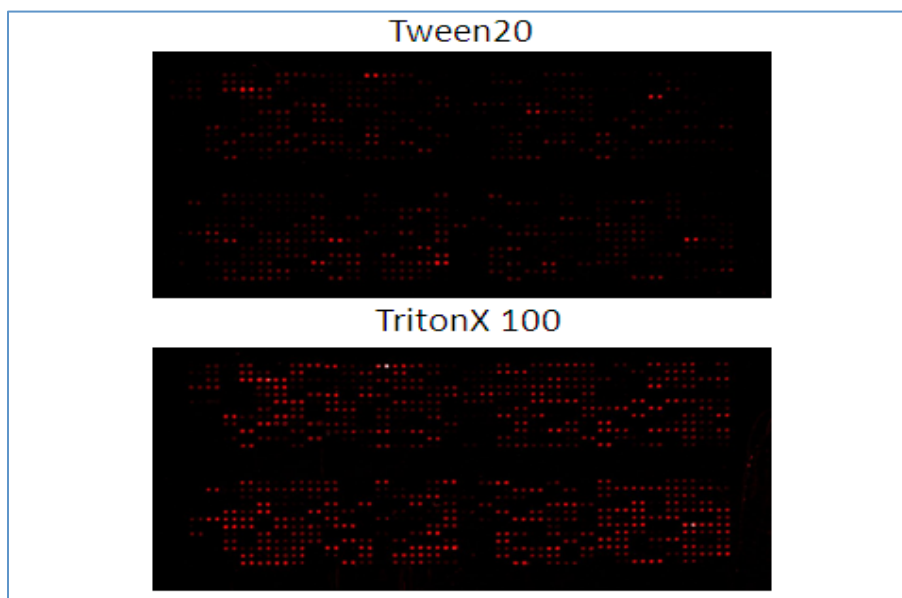


Figure 28. Images of arrays processed with buffers containing detergents Tween20 or TritonX 100. Increase in overall signal intensities are visually apparent for arrays washed with, and incubated with samples and detection antibody diluted in PBS buffer containing TritonX 100 detergent, as compared to Tween 20.

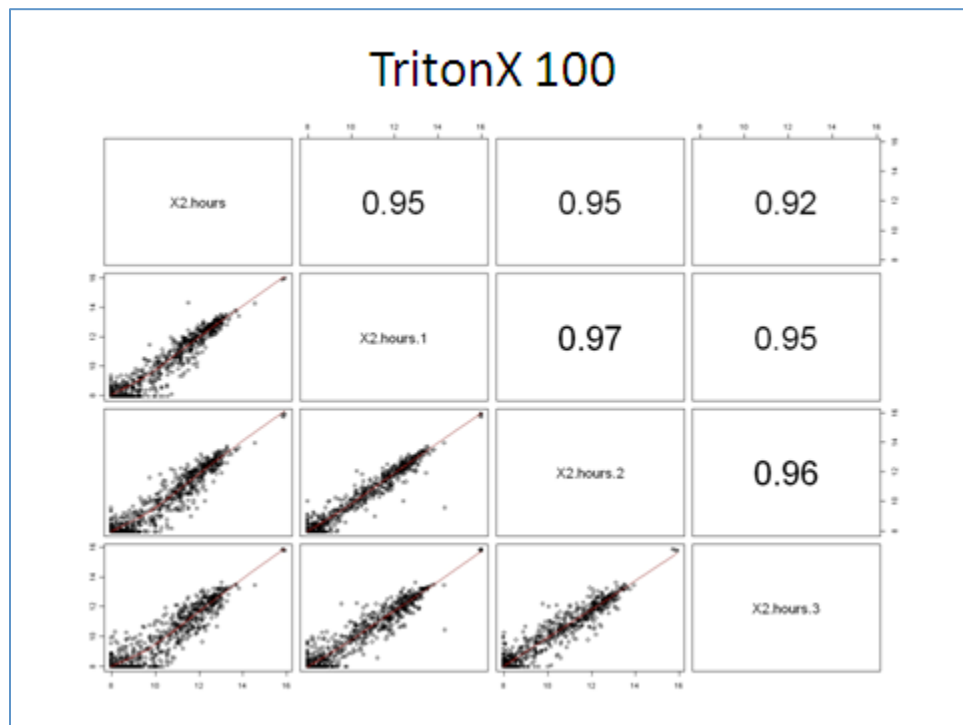
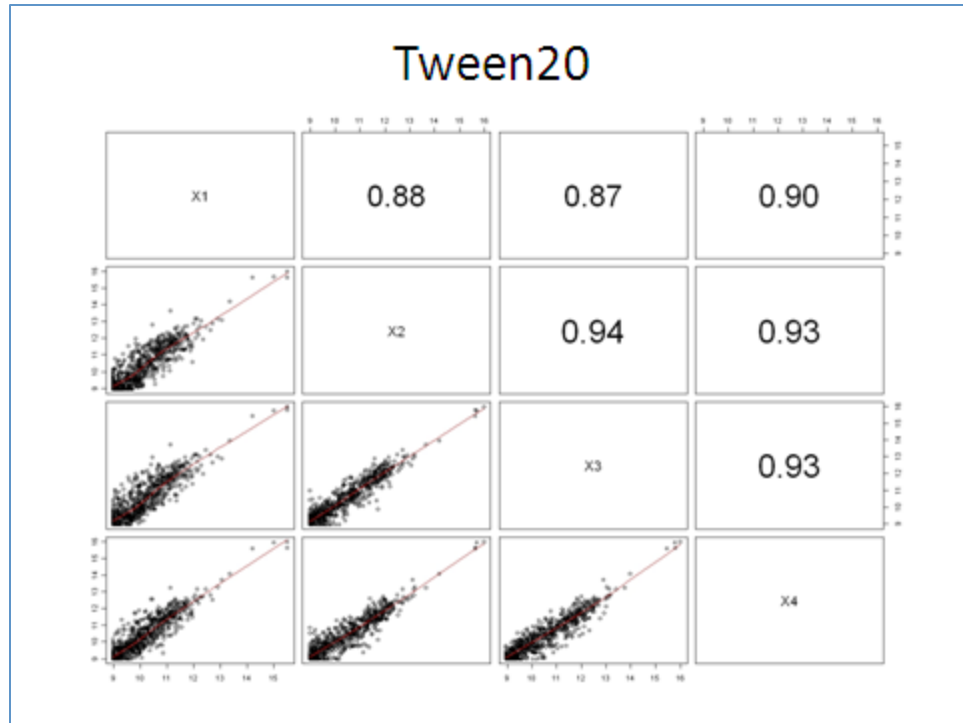


Figure 30. Pairwise correlation plots and corresponding Pearson correlation coefficients of the replicates of each microarray experiment carried out with PBS buffer containing Tween20 or TritonX 100. The mean correlation coefficient of the 4 replicates: Tween20 $R=0.91$ and TritonX 100 $R=0.95$.

The signal intensities of features on arrays are higher for arrays processed with buffer *PBSTx* than *PBSTw* (Figure 28). In order to choose between *TritonX 100* and *Tween20* as a detergent in the PBS buffer used for sample dilution and microarray washing, the correlations between sample replicates within the same experimental setting (*PBSTx* or *PBSTw*) was computed from the log2 transformed intensity values (Figure 29) using the *Pearson* method. A higher correlation coefficient, $r=0.95$, was achieved by samples processed with *PBSTx* buffers, compared to *PBSTw*, $r=0.91$. The results suggest *TritonX 100* detergent can be used for protein microarray experiments yielding higher signal intensities on arrays and reproducibility for sample replicates.

Aside from the results of experiments to observe usability of *Triton X-100* in our protocol, it was recommended by Prof. Franz Steindl (Institut für Angewandte Mikrobiologie, BOKU) for immunoassays and protein microarray application. *Tween 20* usage was replaced with *Triton X-100* as the component in PBS buffer for microarray washing, sample and detection antibody dilutions.

7.2.5 Performance of serum vs plasma for autoantibody screening

As we are screening for potential tumor auto-antibody biomarkers in colon cancer samples and many studies conducted on the theme are primarily employing serum or plasma samples, it is feasible to assume that the non-IgG proteins in these samples could interfere with binding of the antibodies to specific antigens present on protein microarrays. Based on this assumption, the usability of IgG purified from patient plasma for profiling TAA in colon cancer samples was pilot tested with *642+H* array. IgG was considered, as it is the prevalent subtype of Immunoglobulin found in human sera, followed by IgA and IgM that are present in 1:5 ratio to IgG. To compare IgG and serum performance on protein microarrays, IgG was purified from 2 test serum samples with MelonTM Gel IgG Purification kit and diluted with *PBSTw+MP3%* to contain 0.3µg/µL of purified IgG and corresponding serum sample diluted 1:30. Four replicas per sample were incubated on the *642+H* arrays for 4 hours under rotating conditions and for 1 hour with detection antibody *goat anti-human IgG Alexa Fluor® 647* (Red) diluted 1:5000 in *PBST+MP3%*. It is known that adult sera contain around 10-12mg/ml IgG and diluting 1:30 with

PBSTw buffer makes approximate concentration as 0.3-0.4 $\mu\text{g}/\mu\text{L}$ of IgG per sample. Sample IgGs are diluted in *PBSTw+MP3%* buffer, whereas without the blocking effect of milk powder very high number of unspecific binding could be observed.

Apparent from visual inspection of the processed slides in this experiment is that sample serum has significantly higher number of reactive components to the proteins on *642+H* microarray than sample IgGs (Figure 30), suggesting the influence of non-IgG serum proteins.

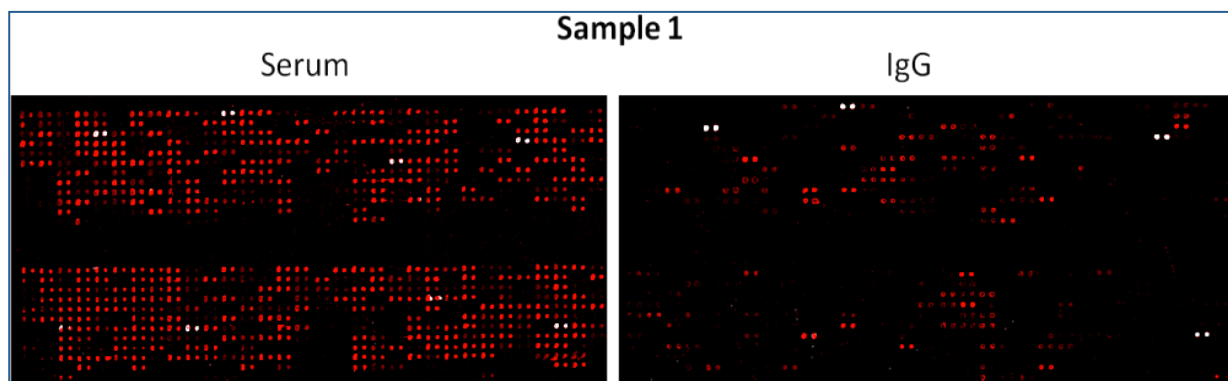


Figure 30. *642+H* array probed with sample 1 serum and IgG. Significant decrease in the number of fluorescent features are observed with sample IgG, however with certain highly fluorescent features consistent with both sample types.

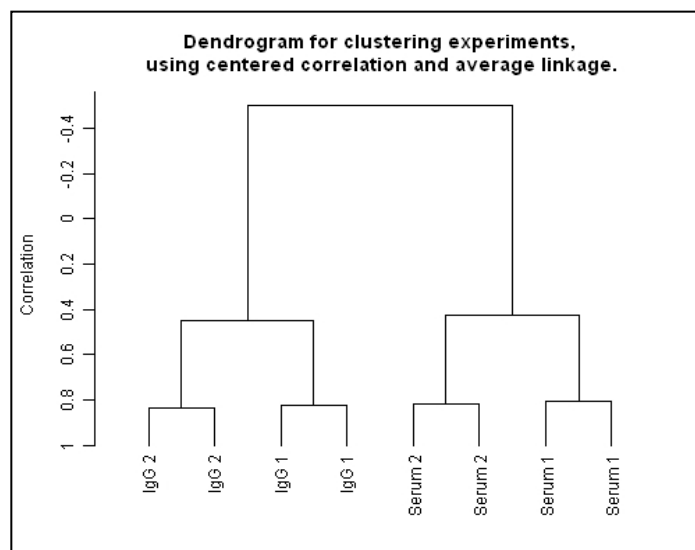


Figure 31. Dendrogram for clustering experiments using centered correlation and average linkage. The correlation of sample replicates, of samples 1 and 2, whether as serum or as purified IgG, are similar at around 0.85. Distinct clusters of IgG and serum samples are seen.

The dendrogram (Figure 31) for clustering experiments using centered correlation and average linkage shows that reproducibility of protein microarray experiments utilizing serum or IgG of samples are comparable with a mean correlation coefficient of 0.85.

It can be viewed from figure 32 that the IgG or serum of the two samples were quite distinct in reactivity over the 642+H arrays. Class comparisons were performed between samples 1 and 2 serum and IgG using features with a significance p value cut-off <0.001. Serum samples class comparison deduced 73 classifiers differentially expressed between the two samples and IgG samples had 49 classifiers, however, when compared to serum sample classifiers there were only 13 classifiers in common.

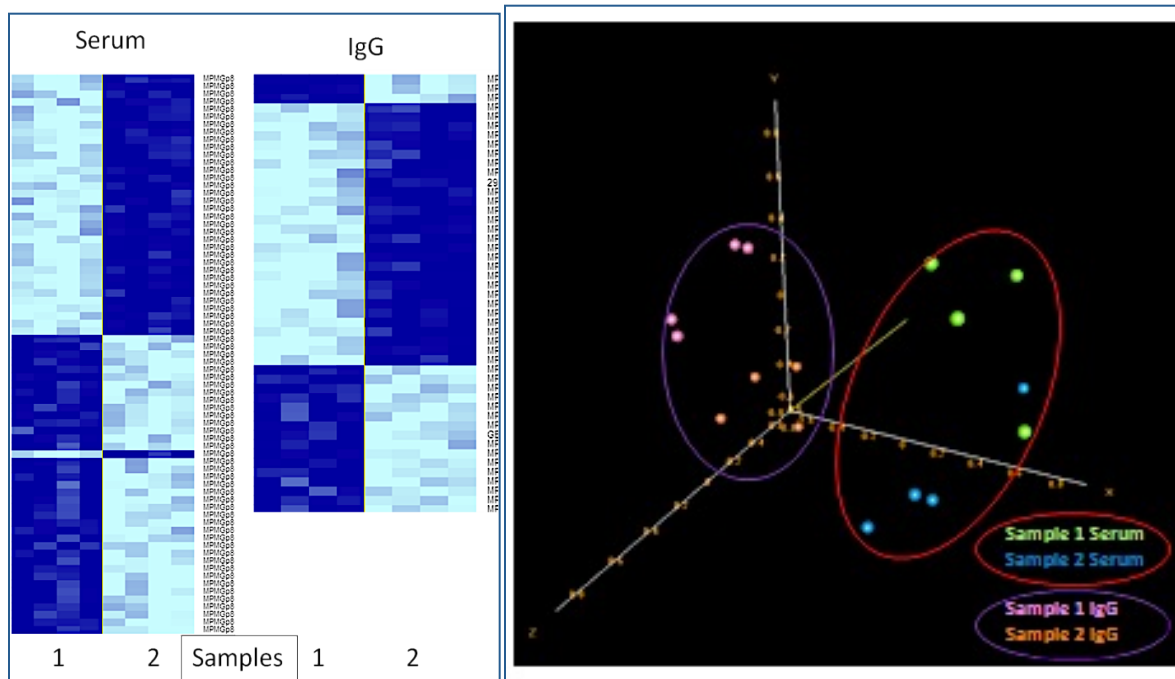


Figure 34. Left. Expression patterns of the first 73 (Serum) and 47 (IgG) reactive proteins with a significance at the nominal 0.001 level of univariate testing among classes (samples 1 and 2).

Right. Visualization of the samples distributions by *Euclidean Distance multidimensional scaling* analysis. Sample serum (Red circled) and IgG (Purple circled) positioned significantly distant in the matrix.

As the IgG and serum of the samples 1 and 2 are producing markedly separate clusters (Figure 31) and the low overlap of classifiers of class comparison between samples 1 and 2, it is feasible to believe that screening for autoantibody markers using purified IgG of samples rather

than the serum would yield smaller number of differentially expressed genes on arrays that have comparable performance of class prediction than those deduced from serum samples. This could be due mainly to the elimination of non-IgG proteins from the serum resulting in possibly an enhanced specificity of autoantibody profile. With the above results in consideration, purification of IgGs from sample serum or plasma was chosen to be incorporated into the protocol to be used for future protein microarray experiments.

7.2.6 Summary of optimizations

Taking into consideration the trial experiments concerning aspects in different steps of protein microarray processing, the following conclusion on the resulting protocol was made (Table 8).

The protocol is designated *Protocol 2* in the Methods section.

Conditions for testing	Options tested	In protocol 1	In protocol 2	Results in Chapter
Slide surface coatings	• SU8 epoxy • PolyAn 3D epoxy • And 8 others	SU8 epoxy	SU8 epoxy	7.2.1
Blocking	• DIG Easy Hyb • PBST+Milk powder 3%	DIG Easy Hyb	DIG Easy Hyb	7.2.1
Effect of Milk powder	• On serum dilution • On purified IgG dilution • On Detection antibody dilution	Serum+PBSTw	Serum+PBSTx IgG+PBSTx+MP3% 2°Ab+PBSTx+MP3%	7.2.3
Probing conditions	• Rotating incubation with Agilent Hybridization Chamber • Static condition in humidity chamber	Static (Coverslip method)	Rotation	7.2.2
Sample incubation duration	• 2 hours • 4 hours	2 hours	4 hours	7.2.2
Wash and dilution buffer	• TritonX 100 • Tween20	Tween 20	TritonX 100	7.2.4
Sample type	• Serum • Purified IgG (MelonGel column)	Serum	Purified IgG	7.2.5

Table 8. Summary of protein microarray protocol optimizations. The tested conditions are listed and the resulting Standard Operating Procedure is named *Protocol 2* (Methods 6.1.7.2).

7.3 Tumor autoantibody candidate marker screening using Protocol 2

The tumor autoantibody candidate marker screening experiment was performed using the newly established procedure designated *Protocol 2* (Methods 6.1.7.2) with the aim to deduce a panel of significantly expressed features between the sample groups with an improved significance and specificity as compared to the previous screening experiment described in Section 7.1. Due to the amount of analysis data and ongoing decisions on patenting results (Intellectual Property Rights), the details on classifiers are not depicted but key findings are given in the results. The Table 9, given below, is a setup that was followed for the screening experiment.

1	Sample processing runs	Total number of samples used	Sample groups and number of samples within each group
	run 1	n=34	Carcinoma n=17, Healthy Control n=17
	run 2	n=30	Carcinoma n=15, Healthy Control n=15
	run 3	n=36	Carcinoma n=9, High risk polyps n=9, Low risk polyps n=9, Healthy Control n=9
	run 4	n=36	Carcinoma n=9, High risk polyps n=9, Low risk polyps n=9, Healthy Control n=9
2	Data transformation	Quantiles Normalization	
		DWD adjustment	
3	Class predictions	1	Healthy Control vs Carcinoma
		2	Healthy Control and Low risk polyps vs High risk polyps and Carcinoma
		3	Healthy Control vs Low risk polyps, High risk polyps and Carcinoma
		4	Healthy Control vs Low risk polyps
		5	Low risk polyps vs High risk polyps
		6	High risk polyps vs Carcinoma.

Table 9. The setup of the candidate marker screening experiment performed according to *Protocol 2* (Methods 6.1.7.2). The first step in the setup shows the distribution of sample groups and number of samples processed among different runs. The second step involves the type of transformation or normalization applied to the background subtracted intensity values after log2 transformation. The final, third, step describes the classes that were defined to perform class prediction analyses for deducing a panel of significantly expressed features.

IgGs were purified from a total of 99 plasma samples and used for 4 runs of 16K protein microarray experiments. Run 1 contained 17 samples each from control and carcinoma groups and run 2 comprised 15 control and 15 carcinoma group samples. The following runs 3 and 4 contained 9 samples from each of the 4 groups: control, high risk, low risk and carcinoma. Due to the total number of test set samples in each group (n=32), 28 samples of the control and carcinoma groups were reanalyzed in the 3rd and 4th runs along with high and low risk samples. The background subtracted median intensity values were log2 transformed (Figure 33) and subsequently were normalized with *Quantiles* normalization (Figure 34) using BRB array tools. For removing “batch effects” due to the different runs, intensity data were adjusted using *Distance Weighted Discrimination* (DWD) method (Figure 35). The *Quantiles* and DWD normalized total intensity values of all arrays shown in figures 34 and 35 are seen to be equally efficient in correcting deviations where medians of each arrays are centered evenly. The same data are viewed by multidimensional scaling by *Principle Component Analysis* (PCA) using Qlucore software, Figure 36, in which *Quantiles* normalized expression data of the four runs exhibited low unification, whereas DWD adjustment to the data resulted in fine clustering of arrays. The features with intensity values having 50th percentile lower than 250 were excluded and 10,172 features remained for the analysis for *Quantiles* normalized data and 10,597 features for DWD adjusted data.

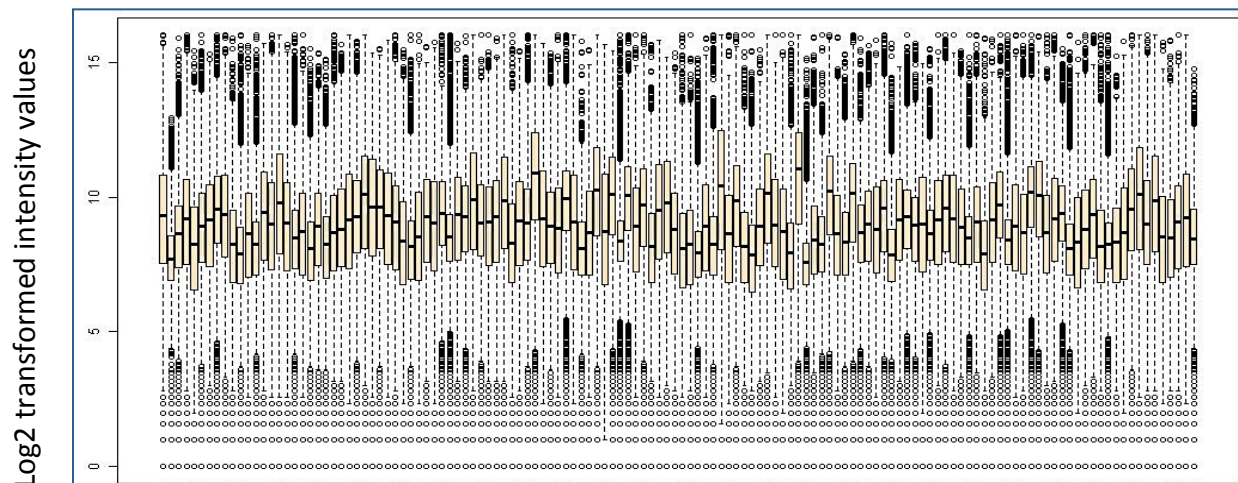


Figure 33. The distribution of log2 transformed intensity values of all arrays prior to normalization.

Log2 transformed intensity values

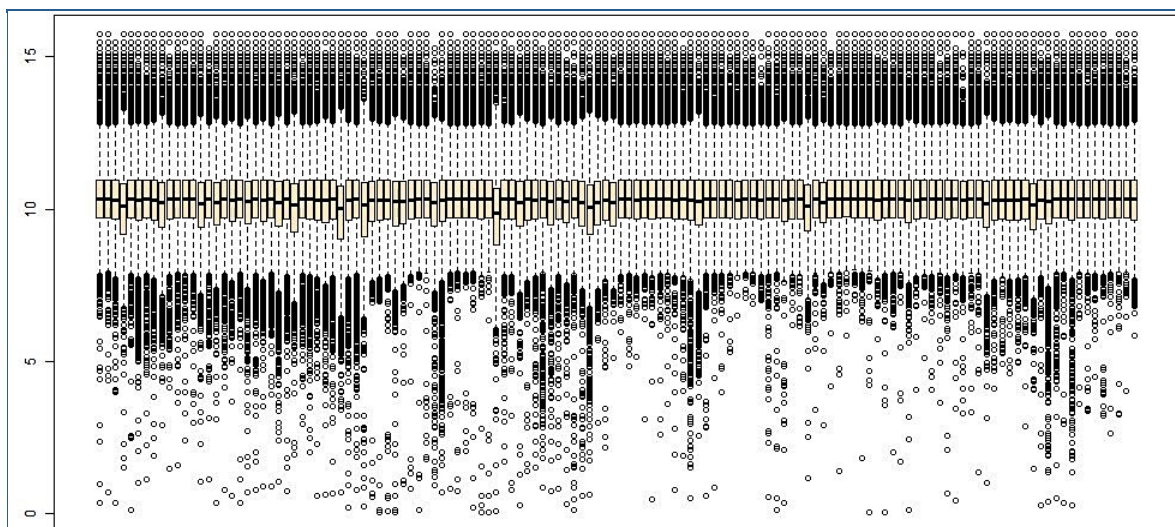


Figure 34. The distribution of log2 transformed intensity values of all arrays after *Quantiles* normalization.

Log2 transformed intensity value

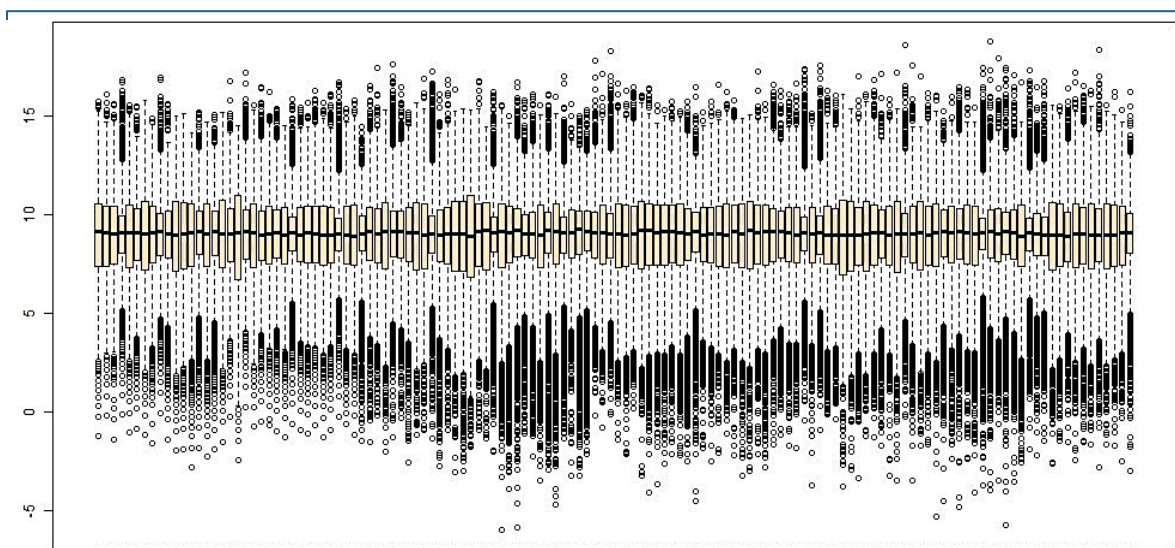


Figure 35. The distribution of log2 transformed intensity values of all arrays after DWD adjustment.

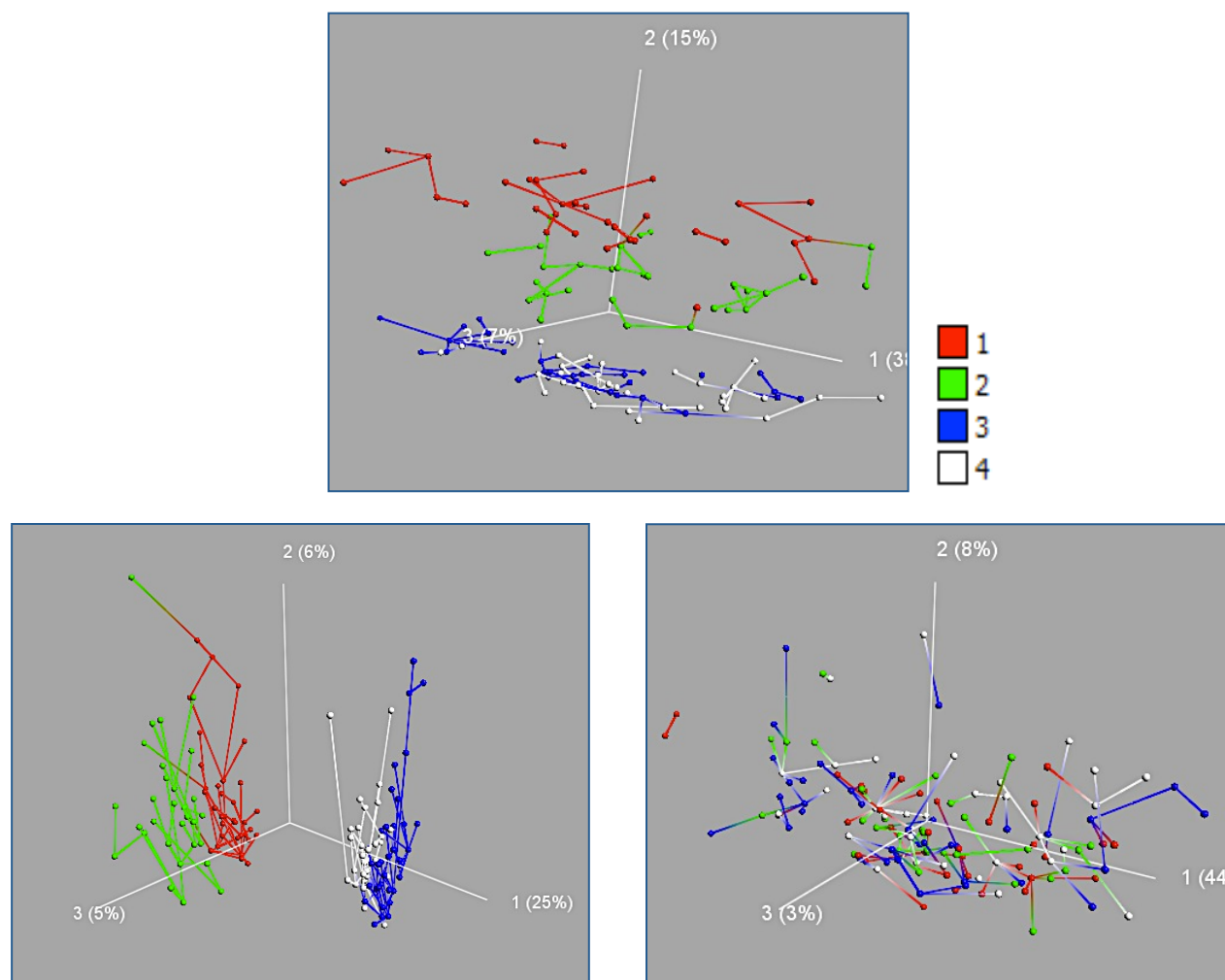


Figure 36. Multidimensional scaling of trends exhibited by expression data according to the 4 processing runs. Visualization carried out by *Principle Components Analysis* using Qlucore software. **Top.** Expression data before normalization. **Bottom left.** Expression data after *Quantiles* normalization. **Bottom right.** Expression data after DWD adjustment. Color legends and numbers represent the runs. Color coding of the screening runs: Red-1, green-2, blue-3, white-4.

For further statistical analyses, both *Quantiles* normalized and DWD adjusted expression data are used. Class prediction analyses were carried out with six different setups as listed in Table 89. In addition, binary tree prediction analysis was made between all 4 groups. The methods applied for class predictions were: *Nearest Centroid Predictor*, *Diagonal Linear Discriminant Analysis*, *K nearest neighbor analysis* with $K=1$ and 3, *Compound Covariate Predictor*, *Support Vector Machines* and *Bayesian Compound Covariate Predictor*. The protein

selection criteria set are: 19-40 greedy-pairs, 40 recursive features elimination, over a grid of significance levels and those with fold-ratio of geometric means between the two classes exceeding 1.25.

7.3.1 Control vs Carcinoma groups

The highest correct classification rate between healthy control (n=49) and carcinoma (n=50) groups for *Quantiles* normalized data was 79% by using *Support Vector Machines* based on 216 features with 0.005 significance threshold value and over 1.25 fold difference among classes. Highest sensitivity and specificity values are achieved with *Compound Covariate Predictor* method with 81% and 72%, respectively (Table 10). The AUC obtained with *Bayesian Compound Covariate Predictor* method was 0.786 (Figure 37).

<i>Compound Covariate Predictor</i>					<i>Diagonal Linear Discriminant Analysis Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.796	0.7	0.722	0.778	Carcinoma	0.735	0.74	0.735	0.74
Control	0.7	0.796	0.778	0.722	Control	0.74	0.735	0.74	0.735
<i>1-Nearest Neighbor Classifier</i>					<i>3-Nearest Neighbors Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.918	0.56	0.672	0.875	Carcinoma	0.939	0.56	0.676	0.903
Control	0.56	0.918	0.875	0.672	Control	0.56	0.939	0.903	0.676
<i>Nearest Centroid Classifier</i>					<i>Support Vector Machine Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.755	0.7	0.712	0.745	Carcinoma	0.878	0.7	0.741	0.854
Control	0.7	0.755	0.745	0.712	Control	0.7	0.878	0.854	0.741
<i>Bayesian Compound Covariate Classifier</i>									
Class	Sensitivity	Specificity	PPV	NPV					
Carcinoma	0.571	0.52	0.538	0.553					
Control	0.52	0.571	0.553	0.538					

Table 10. Performance of classifiers between the Carcinoma (n=49) and Healthy control (n=50) groups with 7 prediction analysis methods at p<0.005 significance level and with 1.25 fold difference among classes.

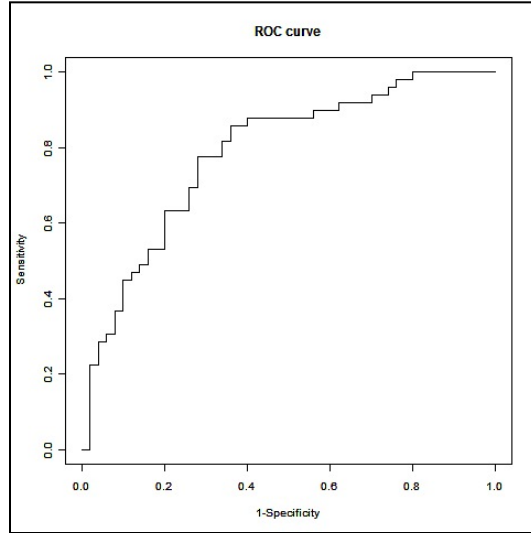


Figure 37. ROC curve of cross validation with *Bayesian Compound Covariate Predictor* analysis for carcinoma (n=49) against control (n=50) groups at significance level of $p < 0.005$ using *Quantiles* normalized data. AUC=0.786.

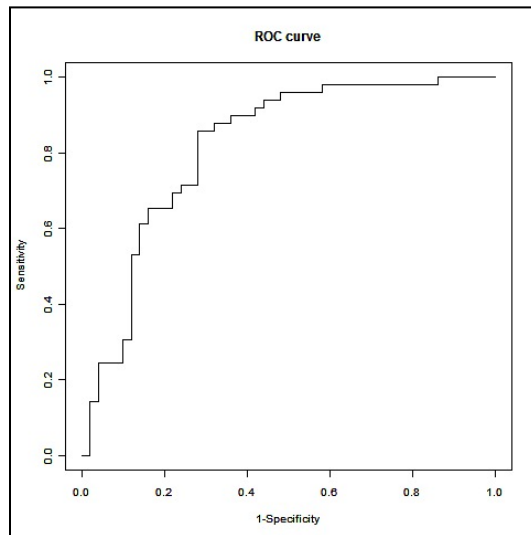


Figure 38. ROC curve of cross validation with *Bayesian Compound Covariate Predictor* analysis for carcinoma against control groups with 25 greedy-pairs genes using *Quantiles* normalized data. AUC=0.818.

When the 25 greedy-pairs criteria was set for the prediction, again 79% correct classification between control and cancer groups could be achieved with *1-Nearest Neighbor* method. AUC was 0.818 for these 50 features with *Bayesian Compound Covariate Predictor* method (Figure

38). With the 50 features Carcinoma and Healthy control groups can be classified with 89% sensitivity and 68% specificity with *1-Nearest Neighbor Classifier* method (Table 11). An exceptionally high, 97% correct classification rate was achieved between healthy control (n=17) and carcinoma (n=17) with *1* and *3-Nearest Neighbor* method for samples in run 1 and an AUC of 0.99 (table 16).

<i>Compound Covariate Predictor</i>					<i>Diagonal Linear Discriminant Analysis Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.816	0.72	0.741	0.8	Carcinoma	0.776	0.66	0.691	0.75
Control	0.72	0.816	0.8	0.741	Control	0.66	0.776	0.75	0.691
<i>1-Nearest Neighbor Classifier</i>					<i>3-Nearest Neighbors Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.898	0.68	0.733	0.872	Carcinoma	0.898	0.54	0.657	0.844
Control	0.68	0.898	0.72	0.733	Control	0.54	0.898	0.844	0.657
<i>Nearest Centroid Classifier</i>					<i>Support Vector Machine Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.837	0.72	0.745	0.818	Carcinoma	0.796	0.66	0.696	0.767
Control	0.72	0.837	0.818	0.745	Control	0.66	0.796	0.767	0.696
<i>Bayesian Compound Covariate Classifier</i>									
Class	Sensitivity	Specificity	PPV	NPV					
Carcinoma	0.653	0.62	0.627	0.646					
Control	0.62	0.653	0.6	0.627					

Table 11. Performance of 50 classifiers (25 greedy-pairs) between the Carcinoma (n=49) and Healthy control (n=50) groups with 7 prediction analysis methods.

For the DWD adjusted expression data, the 25 greedy-pairs genes were able to correctly classify 93% of control and carcinoma samples utilizing *Bayesian Compound Covariate Predictor* method and 89% using a *Compound Covariate Predictor* method with and AUC of 0.927 (Figure 39). The specificity and sensitivities of the methods applied are presented in table 12. The 40 recursive features used to predict the classes control and carcinoma (table 13) had a correct

classification rate of 83% using *Support Vector Machines* method, showing an AUC of 0.88 (Figure 40).

<i>Compound Covariate Predictor</i>					<i>Diagonal Linear Discriminant Analysis Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.837	0.9	0.891	0.849	Carcinoma	0.857	0.84	0.84	0.857
Control	0.9	0.837	0.849	0.891	Control	0.84	0.857	0.857	0.84
<i>1-Nearest Neighbor Classifier</i>					<i>3-Nearest Neighbors Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.898	0.7	0.746	0.875	Carcinoma	0.918	0.7	0.75	0.897
Control	0.7	0.898	0.875	0.746	Control	0.7	0.918	0.897	0.75
<i>Nearest Centroid Classifier</i>					<i>Support Vector Machine Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.857	0.86	0.857	0.86	Carcinoma	0.857	0.78	0.792	0.848
Control	0.86	0.857	0.86	0.857	Control	0.78	0.857	0.848	0.792
<i>Bayesian Compound Covariate Classifier</i>									
Class	Sensitivity	Specificity	PPV	NPV					
Carcinoma	0.796	0.72	0.736	0.783					
Control	0.72	0.796	0.783	0.736					

Table 12. Performance of 50 classifiers (25 greedy-pairs) between the Carcinoma (n=49) and Healthy control (n=50) groups with 7 prediction analysis methods.

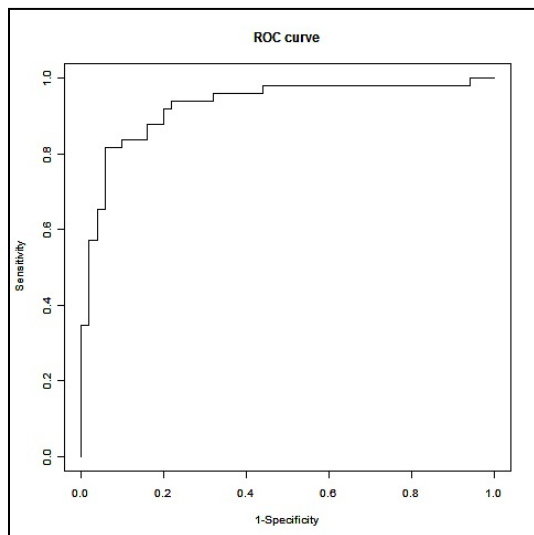


Figure 39. ROC curve of cross validation with *Bayesian Compound Covariate Predictor* analysis for carcinoma against control groups with 25 greedy-pairs proteins using DWD adjusted data. AUC=0.927.

<i>Compound Covariate Predictor</i>					<i>Diagonal Linear Discriminant Analysis Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.816	0.78	0.784	0.812	Carcinoma	0.837	0.74	0.759	0.822
Control	0.78	0.816	0.812	0.784	Control	0.74	0.837	0.822	0.759
<i>1-Nearest Neighbor Classifier</i>					<i>3-Nearest Neighbors Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.837	0.8	0.804	0.833	Carcinoma	0.878	0.74	0.768	0.86
Control	0.8	0.837	0.833	0.804	Control	0.74	0.878	0.86	0.768
<i>Nearest Centroid Classifier</i>					<i>Support Vector Machine Classifier</i>				
Class	Sensitivity	Specificity	PPV	NPV	Class	Sensitivity	Specificity	PPV	NPV
Carcinoma	0.776	0.82	0.809	0.788	Carcinoma	0.816	0.84	0.833	0.824
Control	0.82	0.776	0.788	0.809	Control	0.84	0.816	0.824	0.833
<i>Bayesian Compound Covariate Classifier</i>									
Class	Sensitivity	Specificity	PPV	NPV					
Carcinoma	0.796	0.78	0.78	0.796					
Control	0.78	0.796	0.796	0.78					

Table 13. Performance of 40 classifiers (40 recursive features) between the Carcinoma (n=49) and Healthy control (n=50) groups with 7 prediction analysis methods.

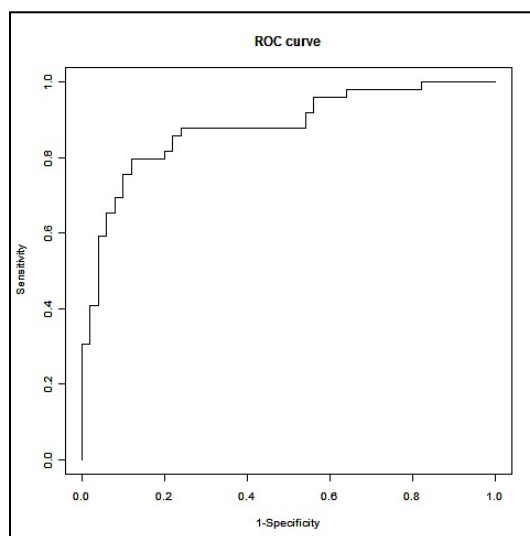


Figure 40. ROC curve of cross validation with *Bayesian Compound Covariate Predictor* analysis for carcinoma against control group with 40 recursive features using DWD adjusted data. AUC=0.88.

7.3.2 Control vs Low risk vs High risk vs Carcinoma

Binary tree classification with *Quantiles* normalized data to classify all four classes, namely control (n=50), low risk (n=18), high risk (n=18) and carcinoma (n=50) were performed and an optimal binary search tree with 3 nodes was obtained (Figure 41). A *Support Vector Machine* was employed using features with a significance p value cut-off <0.001.

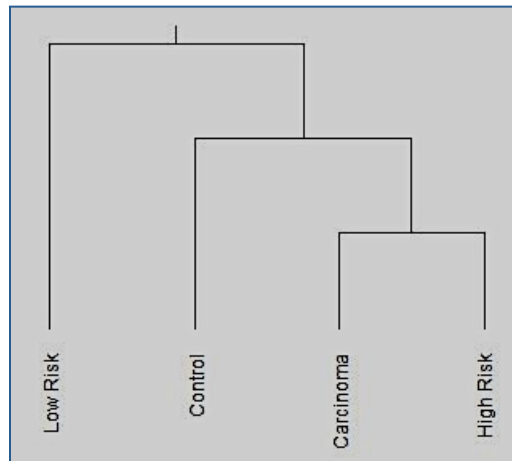


Figure 41. An optimal binary tree prediction of all 4 groups with *Quantiles* normalized data.

Node	Group 1 Classes	Group 2 Classes	Mis-classification rate (%)	Number of classifiers
1	Carcinoma, Control, High Risk	Low Risk	15.7	90
2	Carcinoma, High Risk	Control	13.5	342
3	Carcinoma	High Risk	14.7	16

Table 14. Cross-validation error rates for the optimal binary tree structure shown in Figure 42 and the number of classifiers for classification of each nodes.

Node 1 of the optimal binary tree with 3 nodes consisted of low risk group as one class against the other three groups in second class, predicted with mis-classification rate of 15.7%. Second node contained control group as one class with mis-classification rate of 13.5% and finally, the third node contained carcinoma and high risk groups as separate classes with correct classification rate of 85.3%. The numbers of predictor features for node 1 was 90, node 2 was 342 and for node 3 was 16 (Table 14). There were 8 features overlapping between nodes 1 and 2, 1 feature overlap between nodes 2 and 3 and no overlap between nodes 1 and 3.

For DWD transformed data, an optimal binary search tree having two nodes was obtained at $p < 0.001$ significance level with *Support Vector Machines* prediction method (Figure 42). The first node consisted of carcinoma (n=50), low risk (n=18) and high risk (n=18) groups and the second node of control (n=50) group. Node 1 was predicted with 287 features with mis-classification rate of 7.1%. Node 1 contained carcinoma and high risk groups as one class and low risk as the second, predicted with 150 differentially expressed features with 84.6% correct classification rate (Table 15). There are 12 features overlapping between the classifiers of nodes 1 and 2.

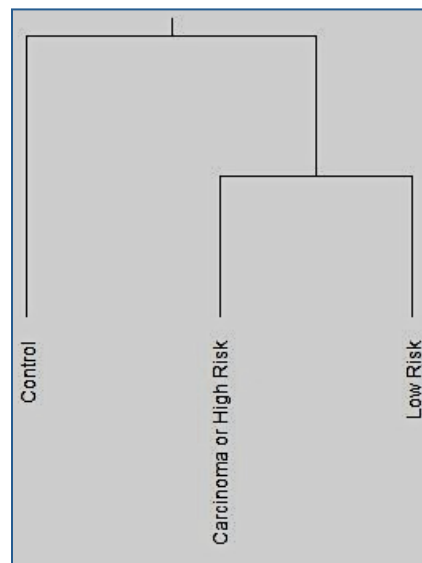


Figure 42. An optimal binary tree prediction of all 4 groups with *DWD adjusted* data.

Node	Group 1 Classes	Group 2 Classes	Mis-classification rate (%)	Number of classifiers
1	Carcinoma, High Risk, Low Risk	Control	7.1	287
2	Carcinoma, High Risk	Low Risk	15.4	150

Table 15. Cross-validation error rates for the optimal binary tree structure shown in Figure 42 and the number of classifiers for classification of each node.

7.3.3 Summary of class prediction analyses and selection of Colon Cancer Candidate array classifiers

Aside from the class prediction analyses of the contrasted control vs carcinoma and between all sample groups, the analyses were made between predefined classes containing different sample groups as described in table 9. The below presented table 16 illustrates the prediction success between the different classes, the algorithm used and the gene selection criteria.

Quantiles normalization		Class prediction success					
		Class prediction method	Gene selection criteria	Sensitivity	Specificity	%, Correct classification	AUC
Classes specified for Class prediction	Healthy Control vs Carcinoma	1-Nearest	25 greedy-pairs	0.941	1	97	0.99
	Healthy Control and Low risk polyps vs High risk polyps and Carcinoma	Support Vector Machines	40 recursive features	0.806	0.836	82	0.852
	Healthy Control vs Low risk polyps, High risk polyps and Carcinoma	1-Nearest Neighbor	0.001 significance level	0.917	0.64	81	0.777
	Healthy Control vs Low risk polyps	3-Nearest Neighbor	40 recursive features	0.889	0.611	75	0.864
	Low risk polyps vs High risk polyps	1-Nearest Neighbor	19 greedy-pairs	0.882	0.778	83	0.778
	High risk polyps vs Carcinoma.	Support Vector Machines	20 greedy-pairs	0.882	0.824	85	0.863
DWD adjustment		Class prediction success					
		Class prediction method	Gene selection criteria	Sensitivity	Specificity	%, Correct classification	AUC
Classes specified for Class prediction	Healthy Control vs Carcinoma	Support Vector Machines	25 greedy-pairs	0.941	0.941	94	0.969
	Healthy Control and Low risk polyps vs High risk polyps and Carcinoma	Diagonal Linier Discriminant Analysis	25 greedy-pairs	0.791	0.836	81	0.849
	Healthy Control vs Low risk polyps, High risk polyps and Carcinoma	1-Nearest Neighbor	0.005 significance level	0.976	0.56	82	0.839
	Healthy Control vs Low risk polyps	3-Nearest Neighbor	25 greedy-pairs	0.833	0.944	89	0.948
	Low risk polyps vs High risk polyps	Support Vector Machines	25 greedy-pairs	0.882	0.833	86	0.807
	High risk polyps vs Carcinoma.	1-Nearest Neighbor	25 greedy-pairs	0.647	0.706	68	0.626

Table 16. Summary of class prediction analysis performed between the specified classes containing different settings of sample groups. The algorithm, gene selection criteria and prediction successes (sensitivity, specificity, percentage of correct classification and cross-validated AUC from the *Bayesian Compound Covariate Predictor*) are presented.

A total of 632 features differentially expressed between the 4 samples groups were chosen for the *Colon Cancer Candidate* array after stringent prediction analyses. The prediction classes, number of classifiers obtained and the corresponding criteria set for the analysis are summarized in table 17. The 632 selected features correspond to clones that represent 593 genes.

Classes		Number clones selected from			Genes	Criteria for selection	
		Quantil	DWD	Quantil+DWD		Quantil	DWD
1	Controls vs Carcinomas	307	119	371	314	25 greedy-pairs, grid of alpha, 1.25 fold	15 and 25 greedy-pairs, 40 recursive features
2	Control, Low risk vs Carcinoma, High risk	82	50	110	104	25 greedy pairs, 40 recursive features	25 greedy pairs
3	Carcinoma vs High risk vs Low risk vs Control- "BinTree" nodes	125		125	116	Node 1 Car,HR,LR vs Con	Node 1 Car,HR, LR vs Con
4	Control vs Carcinoma, High risk, Low risk	50	80	106	96	25 greedy pairs	25 greedy pairs, 40 greedy pairs
5	Control vs Low risk	40	50	85	80	40 recursive features	25 greedy pairs
6	Low risk vs High risk	50	50	78	67	19 and 25 greedy pairs	25 greedy pairs
7	High risk vs Carcinoma	75	86	127	116	20 greedy pairs, 40 recursive features	25 greedy pairs, 40 recursive features
		729	435	632	593		

Table 17. Summary of classifier clones selected for the *Colon Candidate Array* production.

In addition to the above 632 selected clones, a review of research articles reporting TAA biomarker screening and discovery in colon cancer using protein microarray was made and a total of 105 genes have been selected to be incorporated into the *Colon Cancer Candidate* array. Of those 105 genes, 5 were also found in the 632 selected clones from *UniPEX* library 16K array candidate marker screening experiments using a test set of colon cancer plasma samples. As a result, 732 clones deduced, that will comprise the *Colon Cancer Candidate* array which will be further validated using a larger set of Colon cancer plasma samples.

8 DISCUSSION

An evident need for improvements of technical aspects and reagents suited for the type of protein microarray experiments performed for the current candidate cancer marker screening experiment arose when the commonly used protocol for protein microarrays (Protocol 1) was utilized for the pilot screening experiment. The limited performance of the protocol in distinguishing between the 4 groups of clinical samples was a clear indication. Therefore a series of protein microarray protocol optimization experiments were carried out addressing issues such as: surface chemistry, slide surface blocking reagents, experimental conditions with respect to motion and duration, blocking components for reduction of non-specific binding, buffer detergents to reduce tension and increase penetration of the spotted proteins and captured antibodies, and finally reduction of sample complexity by applying purified IgG from serum samples. For the current *UniPEX* cDNA library derived proteins used for production of *16K* arrays, epoxy treated surfaces clearly performed to our advantage with less technical flaws and higher signal-to-noise ratio. An efficient blocking of the surface *SU8* epoxy could be accomplished by the blocking solution *DIG Easy Hyb*.

The drawback of uneven distribution of samples or detection antibodies over arrays with static conditions was overcome by carrying out probing with gasket rimmed cover slides that give enough space for samples to move over the arrays while positioned in rotating hybridization ovens with controlled temperature. This allows a closed environment during target binding, with constant mixing, which is vital for protein microarrays. Additionally, extending sample incubation time point to 4 hours increases the antibody capturing by the proteins present on arrays. With both rotating conditions and longer incubations, the overall signal intensities increased and due to the saturation of some features at 16 hours of incubation, this option was excluded. As the 4 hours of sample incubation yielded higher intensities than 2 hours of incubation and had comparable sample replicate correlations, the option to extend to 4 hours was chosen.

In protein immunoassay procedures, the presence of a blocking component during sample and detection antibody probing is demonstrated to be essential. This also held true for the

current optimizations, where buffers containing 3% milk powder used to dilute samples and detection antibodies distinctly increased specific binding. Dual color experiments clearly demonstrated that appropriate dual fluorescence could be obtained with two detection antibodies directed against human IgG and 6xHis-tag of spotted proteins when buffer containing milk powder is used for their dilutions. Dual fluorescence could not be observed in probing without milk powder, suggesting that binding of the second detection antibody applied onto arrays are significantly enhanced with the blocking agent. It may also be due to binding of detection antibodies instead onto the planar surface of the coverslips. Furthermore, buffer detergents Tween20 and Triton X-100 were tested for their compatibility with the current slide surface chemistry, arrayed proteins and samples. Triton X-100 is observed as better suited, by presenting higher signal intensities and enhanced reproducibility with regards to the correlations between sample replicates.

Finally, the complexity of samples for autoantibody screening was addressed and a possibility to eliminate predominant, competitive proteins in serum or plasma was pursued. With the aim of screening for autoantibodies, it is reasonable to consider usage of IgG fractions purified from serum or plasma in tumor autoantibody screening. Thus, IgG was purified from serum samples were tested and results showed the anticipated outcome of lower numbers of classifiers. Interestingly, very low number of classifiers could be seen to overlap between the deduced classifiers of serum and IgG samples when class comparison analysis was done between the two samples used. As the IgG and serum of samples have comparable performance in sample discrimination and reproducibility, the application of purified IgG of samples was speculated to have increased specificity for TAA profiling on protein microarrays.

An “amended” Standard Operating Procedure was established from the optimization experiments and designated Protocol 2, which was used for candidate marker screening experiments. The second screening experiment using the same test set samples used for the first screening attempt showed significantly improved performance in class predictions. The healthy control and carcinoma groups were predicted with 95% correct classification with an AUC=0.99. After several class prediction analyses between different combinations of the

samples groups into classes, a total of 632 reactive proteins were selected. To possibly improve the performance of the candidate marker panel, similar studies performed for tumor autoantibody discovery in colorectal cancer, 100 genes were obtained from already published data and included in the list of clones for *Colon Cancer Candidate* array production. This array will be further used to validate the 732 clones in tumor autoantibody discovery experiments using a validation sample set already available, comprising of 384 clinical samples belonging to the four groups.

Although several issues concerning improvements in protein microarray procedures were dealt with here, due to the highly intricate nature of proteins, the problem with reproducibility of experiments still persist. Aside from setting an optimized protocol for protein microarrays, it is also important to develop bioinformatics tools that overcome the biases introduced by factors such as processing runs, operators, stored samples, protein microarray printing batches and many more batch effects. Here we successfully applied *Distance Weighted Discrimination* (DWD) that specifically removes systematic bias from microarray data ("Distance Weighted Discrimination (DWD)").

In comparison to published data on TAA classification of healthy control and colon cancer samples using protein microarray screening methods, the 50 proteins deduced by *1-Nearest Neighbor* has a significantly improved performance of 97% correct classification, 94-100% sensitivity and specificity with and AUC of 0.99 (samples control n=17 and carcinoma n=17). For example, Kijanka et.al could identify 22 antigens able to distinguish between colorectal cancer (training set n=20, training set n=43) and non cancer (training set n=20, training set n=40) patients using a high-density protein microarray containing 37,830 recombinant human proteins. The deduced 22 antigens classified sample groups with 83.7% sensitivity and 87% specificity (Kijanka et al., 2010). Further validation with a greater number of samples should be carried out to confirm our findings.

9 CONCLUSIONS

The following conclusions are drawn from the experiments carried out within the frame of this thesis: epoxy surface is best suited for printing of *UniPEX* library derived proteins. Rotating condition applied to protein microarray experiments plays a critical role in enhanced specific binding of target and capture proteins. Upon testing different sample incubation time points, 2 hours of sample incubation is not sufficient to reach binding plateau of antibodies to proteins on *16K* array, rather 4 hours is a better alternative. It was concluded from the experiments within the frame of the current thesis that Triton X-100 detergent increases binding of antibodies on arrays compared to Tween20. Purified IgG could be used for TAA screening with improved specificity than serum samples, resulting in improved performance, as indicated by the TAA candidate marker screening experiment according to the *Protocol 2*. Results of the screening experiment with the new SOP, prediction success could be enhanced to 95% correct classification between control and colon carcinoma sample groups. Although the current improvement of the protocol regarding the protein microarray experiment using the recombinant proteins utilized within the frame of this thesis gives a promising result, other technical aspect could still be improved. This includes optimal expression and purification of the proteins used or even label-free detection of binding experiments performed on protein microarrays. The results presented herein are a very promising basis for further validation using the discovered TAAs as biomarkers for colon cancer non-invasive detection and studies that follow with regards to elucidating their specific functions in colon carcinogenesis.

10 BIBLIOGRAPHY

- Anderson, K. S., Ramachandran, N., Wong, J., Raphael, J. V., Hainsworth, E., Demirkan, G., Cramer, D., et al. (2008). Application of protein microarrays for multiplexed detection of antibodies to tumor antigens in breast cancer. *Journal of proteome research*, 7(4), 1490–1499.
- Arrayit SuperEpoxy 1 and 2 Protein Microarray Substrate Slides - Epoxy Epoxysilane Epoxide Groups Slide E Covalent Protein Coupling. (n.d.). Retrieved July 4, 2012, from http://www.arrayit.com/Products/Microarray_Slides/Epoxy_Slides/epoxy_slides.html
- Babel, I., Barderas, R., Díaz-Uriarte, R., Martínez-Torrecedrada, J. L., Sánchez-Carbayo, M., & Casal, J. I. (2009). Identification of tumor-associated autoantigens for the diagnosis of colorectal cancer in serum using high density protein microarrays. *Molecular & Cellular Proteomics: MCP*, 8(10), 2382–2395. doi:10.1074/mcp.M800596-MCP200
- Babu, M. M. (2004). Introduction to microarray data analysis. *Computational Genomics: Theory and Application*, 225–249.
- Backes, C., Ludwig, N., Leidinger, P., Harz, C., Hoffmann, J., Keller, A., Meese, E., et al. (2011). Immunogenicity of autoantigens. *BMC genomics*, 12(1), 340.
- Bauer, C., Brass, N., Diesinger, I., Kayser, K., Grässer, F. A., & Meese, E. (2002). Overexpression of the eukaryotic translation initiation factor 4G (eIF4G-1) in squamous cell lung carcinoma. *International journal of cancer*, 98(2), 181–185.
- Bo, T., & Jonassen, I. (2002). New feature subset selection procedures for classification of expression profiles. *Genome biology*, 3(4), 1–0017.
- Bouwman, K., Qiu, J., Zhou, H., Schotanus, M., Mangold, L. A., Vogt, R., Erlandson, E., et al. (2003). Microarrays of tumor cell derived proteins uncover a distinct pattern of prostate cancer serum immunoreactivity. *Proteomics*, 3(11), 2200–2207. doi:10.1002/pmic.200300611
- Büssow, K., Nordhoff, E., Lübbert, C., Lehrach, H., & Walter, G. (2000). A Human cDNA Library for High-Throughput Protein Expression Screening. *Genomics*, 65(1), 1–8. doi:10.1006/geno.2000.6141
- Carlsson, A., Wingren, C., Kristensson, M., Rose, C., Fernö, M., Olsson, H., Jernström, H., et al. (2011). Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34), 14252–14257. doi:10.1073/pnas.1103125108
- Carpelan-Holmström, M., Haglund, C., Lundin, J., Järvinen, H., & Roberts, P. (1996). Pre-operative serum levels of CA 242 and CEA predict outcome in colorectal cancer. *European Journal of Cancer*, 32(7), 1156–1161.
- CDC - Colorectal Cancer Risk Factors. (2012, June 20). Retrieved June 29, 2012, from http://www.cdc.gov/cancer/colorectal/basic_info/risk_factors.htm
- Chan, C. C., Fan, C. W., Kuo, Y. B., Chen, Y. H., Chang, P. Y., Chen, K. T., Hung, R. P., et al. (2010). Multiple serological biomarkers for colorectal cancer detection. *International Journal of Cancer*, 126(7), 1683–1690.
- Chandra, H., Reddy, P. J., & Srivastava, S. (2011). Protein microarrays and novel detection platforms. *Expert Review of Proteomics*, 8(1), 61–79. doi:10.1586/epr.10.99

- Chatterjee, M. (2006). Diagnostic Markers of Ovarian Cancer by High-Throughput Antigen Cloning and Detection on Arrays. *Cancer Research*, 66(2), 1181–1190. doi:10.1158/0008-5472.CAN-04-2962
- Chen, H., Jiang, C., Yu, C., Zhang, S., Liu, B., & Kong, J. (2009). Protein chips and nanomaterials for application in tumor marker immunoassays. *Biosensors & Bioelectronics*, 24(12), 3399–3411. doi:10.1016/j.bios.2009.03.020
- Chen, Y. T., Scanlan, M. J., Sahin, U., Türeci, Ö., Gure, A. O., Tsang, S., Williamson, B., et al. (1997). A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. *Proceedings of the National Academy of Sciences*, 94(5), 1914.
- Cretich, M., Damin, F., Pirri, G., & Chiari, M. (2006). Protein and peptide arrays: Recent trends and new directions. *Biomolecular Engineering*, 23(2–3), 77–88. doi:10.1016/j.bioeng.2006.02.001
- Croce, C. M. (2008). Oncogenes and cancer. *New England journal of medicine*, 358(5), 502–511.
- Cunningham, D., Atkin, W., Lenz, H.-J., Lynch, H. T., Minsky, B., Nordlinger, B., & Starling, N. (2010). Colorectal cancer. *The Lancet*, 375(9719), 1030–1047. doi:10.1016/S0140-6736(10)60353-4
- Deepu, A., Sai, V. V. R., & Mukherji, S. (2008). Simple surface modification techniques for immobilization of biomolecules on SU-8. *Journal of Materials Science: Materials in Medicine*, 20(S1), 25–28. doi:10.1007/s10856-008-3471-9
- Distance Weighted Discrimination (DWD) —. (n.d.). Retrieved July 7, 2012, from <https://cabig.nci.nih.gov/community/tools/DWD/>
- Ekins, R., Chu, F., & Biggart, E. (1990). Multispot, multianalyte, immunoassay. *Annales De Biologie Clinique*, 48(9), 655–666.
- Engvall, E., Jonsson, K., & Perlmann, P. (1971). Enzyme-linked immunosorbent assay. II. Quantitative assay of protein antigen, immunoglobulin G, by means of enzyme-labelled antigen and antibody-coated tubes. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 251(3), 427–434.
- Fairweather, D. (2007). Autoimmune Disease: Mechanisms. In John Wiley & Sons, Ltd (Ed.), *Encyclopedia of Life Sciences*. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from <http://doi.wiley.com/10.1002/9780470015902.a0020193>
- Ferlay J. (2010). Shin HR, Bray F, Forman D, Mathers C and Parkin DM.GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer; 2010. Available from: <http://globocan.iarc.fr>, accessed on 22/05/2012.
- Fici, D. A., McCormick, W., Brown, D. W., Herrmann, J. E., Kumar, V., & Awdeh, Z. L. (2010). A protein multiplex microarray substrate with high sensitivity and specificity. *Journal of Immunological Methods*, 363(1), 60–66. doi:10.1016/j.jim.2010.10.005
- Guilleaume, B., Bunes, A., Schmidt, C., Klimek, F., Moldenhauer, G., Huber, W., Arlt, D., et al. (2005). Systematic comparison of surface coatings for protein microarrays. *Proteomics*, 5(18), 4705–4712. doi:10.1002/pmic.200401324
- Haab, B. B., Dunham, M. J., & Brown, P. O. (2001). Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biology*, 2(2), RESEARCH0004.

- Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, 100(1), 57–70.
- Henry C. Pitot. (1993). The molecular biology of carcinogenesis. *Cancer*, 72(Supplement S3), 962–970. doi:10.1002/1097-0142(19930801)72:3+<962::AID-CNCR2820721303>3.0.CO;2-H
- Hofer, P., Baierl, A., Feik, E., Führlinger, G., Leeb, G., Mach, K., Holzmann, K., et al. (2011). MNS16A tandem repeats minisatellite of human telomerase gene: a risk factor for colorectal cancer. *Carcinogenesis*, 32(6), 866–871.
- Hofer, Philipp, Baierl, A., Bernhart, K., Leeb, G., Mach, K., Micksche, M., & Gsur, A. (2012). Association of genetic variants of human telomerase with colorectal polyps and colorectal cancer risk. *Molecular Carcinogenesis*. doi:10.1002/mc.21911
- Hook, A. L., Anderson, D. G., Langer, R., Williams, P., Davies, M. C., & Alexander, M. R. (2010). High throughput methods applied in biomaterial development and discovery. *Biomaterials*, 31(2), 187–198. doi:10.1016/j.biomaterials.2009.09.037
- Huang, R. P., Huang, R., Fan, Y., & Lin, Y. (2001). Simultaneous detection of multiple cytokines from conditioned media and patient's sera by an antibody-based protein array system. *Analytical Biochemistry*, 294(1), 55–62. doi:10.1006/abio.2001.5156
- Humphries, A., & Wright, N. A. (2008). Colonic crypt organization and tumorigenesis. *Nature Reviews Cancer*, 8(6), 415–424. doi:10.1038/nrc2392
- Ingvarsson, J., Larsson, A., Sjöholm, A. G., Truedsson, L., Jansson, B., Borrebaeck, C. A. ., & Wingren, C. (2007). Design of recombinant antibody microarrays for serum protein profiling: targeting of complement proteins. *Journal of proteome research*, 6(9), 3527–3536.
- Kambhampati, D. (2004). *Protein Microarray Technology*. Wiley-VCH.
- Kang, M., Trofin, L., Mota, M. O., & Martin, C. R. (2005). Protein capture in silica nanotube membrane 3-D microwell arrays. *Analytical Chemistry*, 77(19), 6243–6249. doi:10.1021/ac0508907
- Kijanka, G., Hector, S., Kay, E. W., Murray, F., Cummins, R., Murphy, D., MacCraith, B. D., et al. (2010). Human IgG antibody profiles differentiate between symptomatic patients with and without colorectal cancer. *Gut*, 59(1), 69–78. doi:10.1136/gut.2009.178574
- Kusnezow, W., & Hoheisel, J. D. (2003). Solid supports for microarray immunoassays. *Journal of Molecular Recognition*, 16(4), 165–176.
- Lee, D.-W., Kim, H. J., Choi, C.-H., Shin, J.-H., & Kim, E.-K. (2011). Development of a protein chip to measure PKC β activity. *Applied Biochemistry and Biotechnology*, 163(6), 803–812. doi:10.1007/s12010-010-9084-z
- Li, Y., & Reichert, W. M. (2002). Adapting cDNA Microarray Format to Cytokine Detection Protein Arrays†. *Langmuir*, 19(5), 1557–1566. doi:10.1021/la026322t
- Lin, H. S., Talwar, H. S., Tarca, A. L., Ionan, A., Chatterjee, M., Ye, B., Wojciechowski, J., et al. (2007). Autoantibody approach for serum-based detection of head and neck cancer. *Cancer Epidemiology Biomarkers & Prevention*, 16(11), 2396–2405.
- Liu, W., Wang, P., Li, Z., Xu, W., Dai, L., Wang, K., & Zhang, J. (2009). Evaluation of Tumour-Associated Antigen (TAA) Miniarray in Immunodiagnosis of Colon Cancer. *Scandinavian journal of immunology*, 69(1), 57–63.

- Lofton-Day, C., Model, F., DeVos, T., Tetzner, R., Distler, J., Schuster, M., Song, X., et al. (2008). DNA methylation biomarkers for blood-based colorectal cancer screening. *Clinical chemistry*, 54(2), 414–423.
- Lu, H., Goodell, V., & Disis, M. L. (2008). Humoral Immunity Directed against Tumor-Associated Antigens As Potential Biomarkers for the Early Diagnosis of Cancer. *Journal of Proteome Research*, 7(5), 2168–2168. doi:10.1021/pr800228r
- Luebeck, E. G., & Moolgavkar, S. H. (2002). Multistage carcinogenesis and the incidence of colorectal cancer. *Proceedings of the National Academy of Sciences*, 99(23), 15095.
- Madoz-Gúrpide, J., Kuick, R., Wang, H., Misek, D. E., & Hanash, S. M. (2008). Integral protein microarrays for the identification of lung cancer antigens in sera that induce a humoral immune response. *Molecular & Cellular Proteomics: MCP*, 7(2), 268–281. doi:10.1074/mcp.M700366-MCP200
- Madrid, F. F., Chen, W., Tang, N., Zhang, X., Dong, J., Sagar, H., & Abrams, J. (2010). Autoantibodies in Breast Cancer Identify Proteins Involved in Self-Renewal and Epigenetic Chromatin Remodeling. *Open Biomarkers Journal*, 3, 13–20.
- Marie, R., Schmid, S., Johansson, A., Ejsing, L., Nordström, M., Häfliger, D., Christensen, C. B., et al. (2006). Immobilisation of DNA to polymerised SU-8 photoresist. *Biosensors & Bioelectronics*, 21(7), 1327–1332. doi:10.1016/j.bios.2005.03.004
- Markowitz, S. D., & Bertagnolli, M. M. (2009). Molecular Basis of Colorectal Cancer. *New England Journal of Medicine*, 361(25), 2449–2460. doi:10.1056/NEJMra0804588
- Massoner, P., Lueking, A., Goehler, H., Höpfner, A., Kowald, A., Kugler, K. G., Amersdorfer, P., et al. (2012). Serum-autoantibodies for discovery of prostate cancer specific biomarkers. *The Prostate*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/pros.21444/full>
- Meulenbroek, A. . (2002). *Human IgG subclasses: useful diagnostic markers for immunocompetence*. Amsterdam: CLB.
- Nam, M. J., Madoz-Gurpide, J., Wang, H., Lescure, P., Schmalbach, C. E., Zhao, R., Misek, D. E., et al. (2003). Molecular profiling of the immune response in colon cancer using protein microarrays: occurrence of autoantibodies to ubiquitin C-terminal hydrolase L3. *Proteomics*, 3(11), 2108–2115. doi:10.1002/pmic.200300594
- Nelson, D. S. (1977). Autoantibodies in cancer patients. *Pathology*, 9(2), 155–160.
- Nolen, B. M., Marks, J. R., Ta'san, S., Rand, A., Luong, T. M., Wang, Y., Blackwell, K., et al. (2008). Serum biomarker profiles and response to neoadjuvant chemotherapy for locally advanced breast cancer. *Breast Cancer Research: BCR*, 10(3), R45. doi:10.1186/bcr2096
- Olle, E. W., Messamore, J., Deogracias, M. P., McClintock, S. D., Anderson, T. D., & Johnson, K. J. (2005). Comparison of antibody array substrates and the use of glycerol to normalize spot morphology. *Experimental and Molecular Pathology*, 79(3), 206–209. doi:10.1016/j.yexmp.2005.09.003
- Orchekowski, R., Hamelinck, D., Li, L., Gliwa, E., VanBrocklin, M., Marrero, J. A., Vande Woude, G. F., et al. (2005). Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer. *Cancer research*, 65(23), 11193.
- Pereira-Faca, S. R., Kuick, R., Puravs, E., Zhang, Q., Krasnoselsky, A. L., Phanstiel, D., Qiu, J., et al. (2007). Identification of 14-3-3 theta as an antigen that induces a humoral response in

- lung cancer. *Cancer Research*, 67(24), 12000–12006. doi:10.1158/0008-5472.CAN-07-2913
- Perou, C. M., Sørli, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., et al. (2000). Molecular portraits of human breast tumours. *Nature*, 406(6797), 747–752.
- Pla-Roca, M., Leulmi, R. F., Tourekhanova, S., Bergeron, S., Laforte, V., Moreau, E., Gosline, S. J. C., et al. (2012). Antibody colocalization microarray: a scalable technology for multiplex protein analysis in complex samples. *Molecular & Cellular Proteomics: MCP*, 11(4), M111.011460. doi:10.1074/mcp.M111.011460
- Predki, P. F. (2007). *Functional protein microarrays in drug discovery* (Vol. 8). CRC.
- Preininger, C., Bodrossy, L., Sauer, U., Pichler, R., & Weilharter, A. (2004). ARChip Epoxy and ARChip UV for covalent on-chip immobilization of pmoA gene-specific oligonucleotides. *Analytical Biochemistry*, 330(1), 29–36. doi:10.1016/j.ab.2003.12.037
- Qiagen. (2003, June). Qiaexpressionist 06/2003. Retrieved from <http://www.qiagen.com/literature/default.aspx?Term=Qiaexpressionist&Language=EN&LiteratureType=1+2+3+4+53&ProductCategory=0>
- Qiu, J., Madoz-Gurpide, J., Misek, D. E., Kuick, R., Brenner, D. E., Michailidis, G., Haab, B. B., et al. (2004). Development of natural protein microarrays for diagnosing cancer based on an antibody response to tumor antigens. *Journal of Proteome Research*, 3(2), 261–267.
- Quintero, E., Hassan, C., Senore, C., & Saito, Y. (2012). Progress and Challenges in Colorectal Cancer Screening. *Gastroenterology Research and Practice*, 2012, 1–8. doi:10.1155/2012/846985
- Radmacher, M. D., McShane, L. M., & Simon, R. (2002). A paradigm for class prediction using gene expression profiles. *Journal of Computational Biology*, 9(3), 505–511.
- Rapberger, R., Perco, P., Sax, C., Pangerl, T., Siehs, C., Pils, D., Bernthaler, A., et al. (2008). Linking the ovarian cancer transcriptome and immunome. *BMC systems biology*, 2(1), 2.
- Sánchez-Cabo, F., Rainer, J., Dopazo, A., Trajanoski, Z., & Hackl, H. (2011). Insights into global mechanisms and disease by gene expression profiling. *Methods in Molecular Biology (Clifton, N.J.)*, 719, 269–298. doi:10.1007/978-1-61779-027-0_13
- Schena, M. (2005). *Protein microarrays*. Jones & Bartlett Learning. Retrieved from <http://books.google.com/books?hl=en&lr=&id=2-q0EaVqjbAC&oi=fnd&pg=PR2&dq=Schena+M.+Protein+microarrays.+Sudbury,+Mass:+Jones+and+Bartlett%3B+2005.&ots=liylYxP6-f&sig=UGISGzW6YZOkvu43J48LVEUIme8>
- Schweitzer, B., Roberts, S., Grimwade, B., Shao, W., Wang, M., Fu, Q., Shu, Q., et al. (2002). Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nature Biotechnology*, 20(4), 359–365. doi:10.1038/nbt0402-359
- Seong, S.-Y. (2002). Microimmunoassay Using a Protein Chip: Optimizing Conditions for Protein Immobilization. *Clinical and Diagnostic Laboratory Immunology*, 9(4), 927–930. doi:10.1128/CDLI.9.4.927-930.2002
- Seurynck-Servoss, S. L., Baird, C. L., Rodland, K. D., & Zangar, R. C. (2007). Surface chemistries for antibody microarrays. *Front Biosci*, 12(7), 3956–3964.
- Sherr, C. J. (2004). Principles of tumor suppression. *Cell*, 116(2), 235–246.
- Silzel, J. W., Cercek, B., Dodson, C., Tsay, T., & Obremski, R. J. (1998). Mass-sensing, multianalyte microarray immunoassay with imaging detection. *Clinical Chemistry*, 44(9), 2036–2043.

- Simon, R., Lam, A., Li, M. C., Ngan, M., Menenzes, S., & Zhao, Y. (2007). Analysis of gene expression data using BRB-Array Tools. *Cancer informatics*, 3, 11.
- Simon, R. M. (2003). *Design and Analysis of DNA Microarray Investigations*. Springer.
- Somasundaram, R., Swoboda, R., Caputo, L., Otvos, L., Weber, B., Volpe, P., van Belle, P., et al. (2006). Human Leukocyte Antigen-A2–Restricted CTL Responses to Mutated BRAF Peptides in Melanoma Patients. *Cancer research*, 66(6), 3287–3293.
- Somers, V. A., Brandwijk, R. J., Joosten, B., Moerkerk, P. T., Arends, J. W., Menheere, P., Pieterse, W. O., et al. (2002). A panel of candidate tumor antigens in colorectal cancer revealed by the serological selection of a phage displayed cDNA expression library. *The Journal of Immunology*, 169(5), 2772–2780.
- Source BioScience imaGenes. (n.d.). Protein Macroarrays Manual. Retrieved from http://www.lifesciences.sourcebioscience.com/media/290406/sbs_ig_manual_proteinarray_v1.pdf
- Sreekumar, A., Nyati, M. K., Varambally, S., Barrette, T. R., Ghosh, D., Lawrence, T. S., & Chinnaiyan, A. M. (2001). Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Research*, 61(20), 7585–7593.
- Statistik Austria, Cancer incidence and mortality in Austria. (2010). KREBSINZIDENZ UND KREBSMORTALITÄT IN ÖSTERREICH.
- Statistik Austria, Health statistics yearbook. (2010). Jahrbuch der Gesundheitsstatistik, Österreichisches Krebsregister.
- Steiner, M., Klein, E., & Klein, G. (1975). Antinuclear reactivity of sera in patients with leukemia and other neoplastic diseases. *Clinical Immunology and Immunopathology*, 4(3), 374–381.
- Stempfer, R., Syed, P., Vierlinger, K., Pichler, R., Meese, E., Leidinger, P., Ludwig, N., et al. (2010). Tumour auto-antibody screening: performance of protein microarrays using SEREX derived antigens. *BMC Cancer*, 10, 627. doi:10.1186/1471-2407-10-627
- Syed, P. (2012). Evaluation of auto-antibody serum biomarkers for breast cancer screening and in silico analysis of sero-reactive proteins. *Journal of Molecular Biochemistry*, 1(2). Retrieved from <http://www.jmolbiochem.com/index.php/JmolBiochem/article/view/53>
- Tan, E M. (1989). Interactions between autoimmunity and molecular and cell biology. Bridges between clinical and basic sciences. *Journal of Clinical Investigation*, 84(1), 1–6. doi:10.1172/JCI114127
- Tan, E. M., & SHI, F. D. (2003). Relative paradigms between autoantibodies in lupus and autoantibodies in cancer. *Clinical & Experimental Immunology*, 134(2), 169–177.
- Tan, Eng M., & Zhang, J. (2008). Autoantibodies to tumor-associated antigens: reporters from the immune system. *Immunological reviews*, 222, 328–340. doi:10.1111/j.1600-065X.2008.00611.x
- Tanaka, H., Hanasaki, M., Isojima, T., Takeuchi, H., Shiroya, T., & Kawaguchi, H. (2009). Enhancement of sensitivity of SPR protein microarray using a novel 3D protein immobilization. *Colloids and Surfaces. B, Biointerfaces*, 70(2), 259–265. doi:10.1016/j.colsurfb.2008.12.037
- Tanaka, T., Tanaka, M., Tanaka, T., & Ishigamori, R. (2010). Biomarkers for Colorectal Cancer. *International Journal of Molecular Sciences*, 11(9), 3209–3225. doi:10.3390/ijms11093209

- The human IgG subclasses. (n.d.). Retrieved July 3, 2012, from http://wolfson.huji.ac.il/purification/PDF/affinity/CALBIOCHEM_HumanIgG_Subclasses.pdf
- Thermo Scientific. (2010a). Thermo Scientific Pierce Antibody Production and Purification Technical Handbook. Retrieved from http://www.piercenet.com/files/1601975_AntibodyHB_Version2_INTL.pdf
- Thermo Scientific. (2010b). *Thermo Scientific Pierce Protein Assay Technical Handbook*. Retrieved from http://www.piercenet.com/files/1602063_PAssayHB_122910.pdf
- Tomizaki, K., Usui, K., & Mihara, H. (2005). Protein-detecting microarrays: current accomplishments and requirements. *Chembiochem: A European Journal of Chemical Biology*, 6(5), 782–799. doi:10.1002/cbic.200400232
- Tsarfaty-Barad, I., Sauer, U., Preininger, C., & Gheber, L. A. (2011). Miniaturized protein arrays: Model and experiment. *Biosensors & Bioelectronics*. doi:10.1016/j.bios.2011.02.030
- Utz, P. J., & Anderson, P. (1998). Posttranslational protein modifications, apoptosis, and the bypass of tolerance to autoantigens. *apoptosis*, 14, 19.
- Wang, X., Yu, J., Sreekumar, A., Varambally, S., Shen, R., Giacherio, D., Mehra, R., et al. (2005). Autoantibody signatures in prostate cancer. *New England Journal of Medicine*, 353(12), 1224–1235.
- Weinberg, R. A. (2007). *The biology of cancer*. New York: Garland Science.
- Western Blotting. (n.d.). Retrieved July 3, 2012, from http://www.cyanagen.it/index.php?option=com_content&view=article&id=2&Itemid=144
- Wright, G. W., & Simon, R. M. (2003). A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics*, 19(18), 2448–2455.
- Wu, X., Molinaro, C., Johnson, N., & Casiano, C. A. (2001). Secondary necrosis is a source of proteolytically modified forms of specific intracellular autoantigens: implications for systemic autoimmunity. *Arthritis & Rheumatism*, 44(11), 2642–2652.
- www.imagenes-bio.de. (n.d.-a). Retrieved July 3, 2012, from http://www.imagenes-bio.org/info/vectors/pQE30NST_pic.shtml
- www.imagenes-bio.de. (n.d.-b). Retrieved July 3, 2012, from http://www.imagenes-bio.org/info/vectors/pQE80LSN_cloned_pic.shtml
- Yalow, R. S., & Berson, S. A. (1959). Assay of plasma insulin in human subjects by immunological methods. Retrieved from <http://www.nature.com/nature/journal/v184/n4699/abs/1841648b0.html>
- Yan, F., Sreekumar, A., Laxman, B., Chinnaiyan, A. M., Lubman, D. M., & Barder, T. J. (2003). Protein microarrays using liquid phase fractionation of cell lysates. *Proteomics*, 3(7), 1228–1235. doi:10.1002/pmic.200300443
- Zubtsov, D. A., Savvateeva, E. N., Rubina, A. Y., Pan'kov, S. V., Konovalova, E. V., Moiseeva, O. V., Chechetkin, V. R., et al. (2007). Comparison of surface and hydrogel-based protein microchips. *Analytical Biochemistry*, 368(2), 205–213. doi:10.1016/j.ab.2007.04.040

CURRICULUM VITAE

Surname	Sergelen	Date of Birth:	July 27, 1986
Given Name	Khulan	Sex:	Female
Phone:	(43-1) 9221637	Marital Status:	Married
Cell Phone:	(43) 0650-3814231	Nationality:	Mongolian
Email:	hulanser@gmail.com	Khulan.Sergelen.fl@ait.ac.at	

Education

October 2009- July 2012

MSc, Molecular Medicine, Master of Molecular Biology, University of Vienna, Vienna, Austria

Thesis topic: *"Evaluation of protein microarray technology for tumor autoantibody screening in colon cancer"*

September 1, 2003 - June 8, 2007

BSc in Genetics and Molecular Biology National University of Mongolia, Mongolia

Thesis topic: *"Preparing and studying the activity of FMDV 3ABC antigen and its positive serum"*

Achievements

2011: LSK funded aid for student research

2010: FWF funded aid for student research

2007: 1st Place, Medical Branch, "MASTER-2007" Student Science Conference

2005: "UMAP Taiwan Scholarship" for 1 year International Exchange Program

2002-2003: Gold medals for National English Olympiad (team and individual achievement)

Professional Experience

May 2011- July 2012: Master's student "Protein-chip-technology for improving Colon and Prostate Cancer diagnostic testing", Health and Environment Department, Austrian Institute of Technology, Vienna, Austria

March-December 2010 : Research student "Molecular analysis of mba gene ON/OFF switch in *Ureaplasma parvum*", IBMH, University of Veterinary Medicine, Vienna, Austria

July 2007: Molecular Biologist, Bacteriology Laboratory, "Gyals" Medical Center, Ulaanbaatar, Mongolia

July 2006: Laboratory of Respiratory Virology, National Center for Communicable Diseases. Cell culturing and virus subtyping.

Certificates, training courses

Dec 2008: Molecular Biology diagnostic methods, Bioneer, South Korea

Sep 2008: Lecturer, "Current Advances in Laboratory Diagnostics Workshop", Ulaanbaatar, Mongolia

Nov-Dec 2007: Bacteriology Samples to result, Automated methods for ID and AST testing, BioMerieux, France

Technical Skills

Biochemistry: Spectrophotometric analysis

Immunology: Antisera generation and polyclonal antibody isolation, ELISA

Cell Culture: Adherent cell culture

Microbiology: Bacterial identification with classical and modern methods, antibiotic resistance testing;

Mycoplasma cultivation; Viral subtyping

Molecular Biology: Real-Time PCR; cloning; SDS-PAGE; agarose gel electrophoresis; Western, Southern; restriction mapping;

Animal model: laboratory rabbits, drosophila

Chip technology: Microarray spotting, hybridization, bioinformatics analyses

Language Proficiency

Mongolian (native), English (fluent) and German (good)

TOEFL score: 610

TWE score: 5

Referees**Ass.Prof. Battsetseg Gonchigoo, DVM, PhD, ScD**

Head, Laboratory of Ectoparasitology and protozoology

Mongolian Institute of Veterinary Medicine

Zaisan, 210153

Ulaanbaatar, Mongolia

Phone: 976-91119027

Fax: 976-11-341553

Email: gotsetseg@magicnet.mn

Prof. Nymadawa Pagbajabyn, MD, PhD, DSc(Med)

President, Mongolian Academy of Medical Sciences,

Professor, Health Science University of Mongolia,

Central Post, P.O.Box-596

Ulaanbaatar, Mongolia

Phone: 976-99112306

Fax: 976-11-265916

Email: nymadawa@gmail.com

Dr. Carl-Ulrich Zimmermann, PhD

Senior Scientist

Institute for Bacteriology, Mycology and Hygiene

Veterinary Medical University of Vienna

Veterinärplatz 1

A-1210 Vienna, Austria

Phone: +43-1-250772112

Email: Carl-Ulrich.Zimmerman@vetmeduni.ac.at

Dr.Andreas Weinhäusel, Dipl.-Ing. Priv.-Doz. Dr.

Project Leader

Molecular Medicine

Health and Environment

AIT Austrian Institute of Technology GmbH

Muthgasse 11

A-1190 Vienna, Austria

Phone: +43-50550 4401

Email: Andreas.Weinhäusel@ait.ac.at

Publications

- P.Syed, K.Vierlinger, A.Kriegner, **K.Sergelen**, J.Luna-Coronell, C.Rappaport-Fürhauser, C.Nöhammer, C.F.Singer, A.Weinhäusel. Evaluation of auto-antibody serum biomarkers for breast cancer screening and in silico analysis of sero-reactive proteins. Journal of Molecular Biochemistry. 2012: 1(2).
- P.Syed, J.A.Luna-Coronell, **K.Sergelen**, I.Gyurján, A.Weinhäusel. The current status of cancer biomarker research using tumour-associated antigens for minimal invasive and early cancer diagnostics. Journal of Proteomics. Submitted for publication.