Cancer Autoantibody Biomarker

Discovery and Validation

Using Nucleic Acid Programmable Protein Array

by

Jie Wang

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved Apr 2015 by the
Graduate Supervisory Committee:

Joshua LaBaer, Chair
Douglas F Lake
Karen S Anderson
Yung Chang

ARIZONA STATE UNIVERSITY

May 2015

ABSTRACT

Currently in the US, many patients with cancer do not benefit from the population-based screening, due to challenges associated with the existing cancer screening scheme. Blood-based diagnostic assays have the potential to detect diseases in a non-invasive way. Proteins released from small early tumors may only be present intermittently and get diluted to tiny concentrations in the blood, making them difficult to use as biomarkers. However, they can induce autoantibody (AAb) responses, which can amplify the signal and persist in the blood even if the antigen is gone. Circulating autoantibodies is a promising class of molecules that have potential to serve as early detection biomarkers for cancers. This Ph.D thesis aims to screen for autoantibody biomarkers for the early detection of two deadly cancer, basal-like breast cancer and lung adenocarcinoma. First, a method was developed to display proteins in both native and denatured conformation on protein array. This method adopted a novel protein tag technology, called HaloTag, to covalently immobilize proteins on glass slide surface. The covalent attachment allowed these proteins to endure harsh treatment without getting dissociated from slide surface, which enabled the profiling of antibody responses against both conformational and linear epitopes. Next, a plasma screening protocol was optimized to significantly increase signal to noise ratio of protein array based AAb detection. Following this, the AAb responses in basal-like breast cancer were explored using nucleic acid programmable protein arrays (NAPPA) containing 10,000 full-length human proteins in 45 cases and 45 controls. After verification in a large sample set (145 basal-like breast cancer cases / 145 controls / 70 non-basal breast cancer) by ELISA, a 13-AAb classifier was developed to differentiate patients from controls with a sensitivity

of 33% at 98% specificity. Similar approach was also applied to the lung cancer study to identify AAbs that distinguished lung cancer patients from computed-tomography positive benign pulmonary nodules (137 lung cancer cases, 127 smoker controls, 170 benign controls). In this study, two panels of AAbs were discovered that showed promising sensitivity and specificity. Six out of eight AAb targets were also found to have elevated mRNA level in lung adenocarcinoma patients using TCGA data. These projects as a whole provide novel insights on the association between AAbs and cancer, as well as general B cell antigenicity against self-proteins.

To my parents, Hong Zhao (赵宏) and Xiaoming Wang (王晓明)

my wife, Jingying Guo (郭京滢)

and my grandpa, Dianming Zhao (赵殿明)

ACKNOWLEDGEMENTS

I want to recognize all the people without whom my projects would have never succeeded. First and foremost, I would like to recognize my advisor Dr. Joshua LaBaer and Dr. Ji Qiu who mentored me and championed me during the execution of my research. Their advice and encouragement inside and outside the lab made this work possible.

I want to recognize the mentoring and guidance of Dr. Karen S Anderson, Dr. Garrick Wallstrom, and Dr. Jin G Park. Their support was invaluable. I was also very fortunate to depend on the mentorship of the following excellent scientists that guided me and encouraged me during my Ph.D: Dr. Douglas Lake, Dr. Yung Chang, and Dr. Diana Petitti.

I would also recognize my mentors and friends Dr. Xiaobo Yu, Dr. Gokhan Demirkan, Dr. Fernanda Festa, Dr. Laura Gonzales.

I would also express my thanks to Kristi Barker, Dr. Shilpa Shivakumar, Yangyang Tang and Dr. Juan Du for their excellent support in the lung cancer autoantibody project.

I am also very thankful for help and aid of my collaborators of Dr. Jonine D Figueroa from National Cancer Institute, Dr. Williams Rom and Dr. Harvey Pass from New York University.

In addition, I am extremely grateful for the friendships and fellowship of Dr. Carlos Morales Bentazos, Xiaofang Bian, Justin Saul, Brianne Petritis, Dr. Benjamin Katchman, Haoyu Wang, Seron Eaton, and Milen Vitanov, who directly or indirectly worked along with me in my research.

I am grateful for the guidance I received from my program course instructor, Dr. JoAnn C. Williams and coordinator, Maria Hanlin, Laura Hawes and Lauren Dempsey. Thanks to all my classmates in the Biological Design Graduate Program for their camaraderie.

I could not have done anything without the support and love of my wife Jingying Guo, my parents, Xiaoming Wang, Hong Zhao, and my in-laws, Yi Guo, Yinghong Hu.

Finally, I would like to thank the National Cancer Institute Early Detection Research Network (Grant 5U01CA117374-02 to Dr. Joshua LaBaer) for funding this work.

TABLE OF CONTENTS

Page

LIST OF TABLES ... 6

LIST OF FIGURES ... 8

CHAPTER

1 INTRODUCTION ... 10

1.1 Cancer Screening and Early Detection ... 10

1.1.1 Overview ... 10

1.1.2 Basic Statistics for Biomarker ... 12

1.1.2 Challenges with Population Screen for Cancer ... 13

1.2 Blood-based Diagnostic Biomarker for Cancer ... 18

1.2.1 Overview ... 18

1.2.2 Literature Review of Blood-based Cancer Biomarker ... 19

1.2.3 Challenges in Biomarker Study ... 20

1.2.4 A Five-phase Strategy to Develop and Assess a Cancer Screening Test ... 22

1.3 Cancer Autoantibodies ... 26

1.3.1 Cancer Patients can Develop Autoantibodies against Tumor ... 26

1.3.2 Efforts in Searching for Cancer Associated Autoantibodies ... 29

1.4 Emerging Technologies for Profiling Autoantibody Responses ... 33

1.4.1 SEREX ... 34

1.4.2 SERPA ... 35

1.4.3 Protein Microarray ... 36

CHAPTER Page

1.5 Cell-free Protein Microarrays 41

1.5.1 Nucleic Acid Programmable Protein Array 41

1.5.2 Other Cell-free Protein Array Platforms 43

1.6 Project Overview 45

2 DEVELOPMENT OF A VERSATILE PROTEIN MICROARRAY ENABLING ANTIBODY PROFILING AGAINST DENATURED PROTEIN 48

2.1 Abstract 48

2.2 Introduction 49

2.3 Materials and methods 52

2.3.1 Detection of Plasmid DNA and Expressed Proteins on HaloTag Protein Array 52

2.3.1 Blocking of Plasma Samples with *E.coli* Lysates 52

2.3.3 Protein Microarray Denaturation 52

2.4 Results and Discussion 53

2.4.1 Develop a Protein Array Platform to Display Denatured Proteins 53

2.4.2 E.coli Lysate Pre-block Plasma Reduces Non-specific Binding 56

2.4.3 Evaluate Hela IVTT System in Serum Antibody Detection 57

2.5 Conclusion 61

3 PLASMA AUTOANTIBODIES ASSOCIATED WITH BASAL-LIKE BREAST CANCERS 64

3.1 Abstract 64

3.2 Introduction 66

CHAPTER Page

3.3 Materials and methods ... 70

3.3.1 Study Samples ... 70

3.3.2 Protein Array Experiments ... 70

3.3.3 Protein Array Image Analysis and Quantification ... 72

3.3.4 Generating AAb Target Protein Sets for Enrichment Analysis ... 72

3.3.5 Gene Set Enrichment Analysis (GSEA) ... 73

3.3.6 Gene Ontology Cellular Component Analysis ... 74

3.3.7 Antigen Selection for Focused Array ... 74

3.3.8 Power Analysis for the Biomarker Discovery ... 75

3.3.9 Antigen Selection for ELISA Verification ... 76

3.3.10 ELISA Assays ... 76

3.3.11 Statistical Analysis ... 78

3.4 Results ... 80

3.4.1 Sample Tumor Characteristics and Risk Factors ... 80

3.4.2 Characterizing the Global Autoantibody Response in Patients with Basal-like Breast Cancer and Healthy Individuals ... 81

3.4.3 Identification of Candidate AAb Biomarkers for Basal-like Breast Bancer ... 94

3.4.4 Blind Test of AAb Biomarkers for Basal-like Breast Cancer ... 99

3.4.5 An AAb Signature for Diagnosing Basal-like Breast Cancer ... 100

3.4.6 AAb Response in Other Breast Cancer Subtypes ... 103

3.4.7 TP53 AAb Response and TP53 Protein Levels ... 105

CHAPTER Page

3.4.8 TP53 AAb Response and *TP53* Mutation 107

3.4.9 AAbs and Overall Survival in Basal-like Breast Cancer 108

3.5 Discussion 110

3.6 Conclusion 117

4 COMPARATIVE STUDY OF AUTOANTIBODY RESPONSES BETWEEN LUNG ADENOCARCINOMA AND BENIGN PULMONARY NODULE 118

4.1 Abstract 118

4.2 Introduction 120

4.3 Materials and methods 122

4.3.1 Characteristics of Plasma Samples 122

4.3.2 Protein Array Experiments 122

4.3.3 Protein Array Image Analysis and Quantification 122

4.3.4 Candidate Selection 123

4.3.5 Pathway Analysis 124

4.3.6 ELISA assays 124

4.3.7 Statistics and Data Analysis 124

4.4 Results 126

4.4.1 Identification of Candidate AAbs Associated with Lung Adenocarcinoma 126

4.4.2 Validation in Lung Cancer Patients versus Healthy Smoker Controls 128

4.4.3 Classification of Lung Cancer versus Benign Controls 130

CHAPTER Page

4.4.4 Effect of Patient and Disease Characteristics on AAb Positivity 132

4.4.5 Correlation of AAb Targets and Their mRNA Levels 133

4.5 Discussion.. 136

4.6 Conclusion 141

5 SUMMARY AND FUTURE DIRECTIONS.. 142

5.1 Summary 142

5.2 Future Directions 144

REFERENCES 146

APPENDIX.. 169

A FIGURE PERMISSIONS 169

B MANUSCRIPT COPYRIGHT PERSMISSIONS 176

C PUBLICATIONS AND SUBMISSIONS.. 185

LIST OF TABLES

Table Page

1-1. Current Cancer Screening Practiced in the US 15

1-2. Five Phases of Biomarker Development 23

3-1. Established Breast Cancer Risk Factors in the Polish Breast Cancer Study 83

3-2. Descriptive Characteristics of Samples Used to Identify AAbs Associated with BLBC. 86

3-3. List of 14 Antigens Selected from Focused Array Experiments. 97

3-4. List of 15 Antigens Selected from ELISA Verification 99

3-5. Training and Test Statistics for Potential BLBC Autoantibody Biomarkers. 101

3-6. 13-AAb Classifier. 102

3-7. Performance of 13 AAbs in Multiple Breast Cancer Subtypes. 102

3-8. Relationship between TP53 AAb Response and TP53 Tissue IHC Score in Basal-like Breast Cancers. 105

3-9. Relationship between TP53 AAb Response and TP53 Mutation in Basal-like Breast Cancers. 107

3-10. Hazard Ratios for 2 Antigen AAb Responses that are Significantly Associated with Survival amongst 145 Basal-like Breast Cancers. 108

4-1. Sample Information 123

4-2. Discovery and Validation Statistics of Selected AAbs. 129

4-3. Sensitivity and Specificity of Individual AAb from Panel I. 130

4-4. Sensitivity and Specificity of Individual AAb from Panel II. 131

Table Page

4-5. Sensitivity of Individual AAb Using Various Benign Lung Nodules as Control Samples 132

4-6. Positivity of Individual Autoantibody from Panel I by Stage. 133

4-7. Positivity of Individual Autoantibody from Panel II by Stage. 133

LIST OF FIGURES

Figure Page

1-1. Representation of the outcome of a clinical diagnostic test 12

1-2. Dynamic protein concentration range in human plasma sample. 28

1-3. Schematic representation of the development of autoantibody responses against tumor associated antigens (TAA). 29

1-4. Scheme for SEREX and SERPA. 36

1-5. Phage display protein array used to screening autoantibody for prostate cancer..... 38

1-6. Scheme for plasma screening using NAPPA and representative images of array probed with plasma from patient or healthy donors. 41

2-1. HaloTag protein arrays can withstand harsh denaturing conditions......................... 54

2-2. Vector Map of pJFT7-nHalo 55

2-3. Optimization of the plasma screening protocol by blocking plasma samples with E.coli lysate and using HeLa IVTT expression system. 58

2-4. Blocking plasma with E. coli lysate can reduce background signal on HeLa IVTT expressed protein array. 60

2-5. NAPPA using covalent capture chemistry enables the detection of AAb against denatured proteins. 62

2-6. Different AAb response to TPD52 and MYC between native protein array and denatured protein array. 63

3-1. Overview of study design. 81

3-2. Representative array images for “ring” structure surrounding spots. “Ring” was considered as positive response during visual image analysis... 82

Figure Page

3-3. Protein array screening 87

3-4. Array screening reproducibility. 88

3-5. AAb response counts per sample are similar between BLBC and controls. 89

3-7. GSEA analysis of AAb protein targets against their biochemical properties 92

3-8. GO cellular component analysis. 94

3-9. Heatmap of plasma AAb responses against selected proteins based on quantitative analysis in sample set 97

3-10. Heatmap of plasma AAb responses against selected proteins based on visual image analysis in sample set 1 98

3-11. ROC curve of the 13-AAb classifier. 103

3-12. AAb responses in various breast cancer subtypes. 104

3-13. mRNA expression levels of candidate biomarkers in breast invasive carcinoma grouped by subtypes (TCGA). 107

3-14. Overall survival of BLBC patients stratified by AAb responses 109

4-1. Overview of Study Design. 127

4-2. Summary of lung cancer associated antigens discovered from protein array screening. 128

4-3. AAb responses of individual AAb from panel II. 131

4-4. Multivariate analysis of clinical factors and AAb responses. 134

4-5. mRNA expression level of AAb targets from both panels (TCGA). 135

CHAPTER 1

INTRODUCTION

1.1 Cancer screening and early detection

1.1.1 Overview

According to cancer statistics, in 2014, it is estimated that 1,665,540 people in the US were diagnosed with cancer, and 585,720 died of the disease (Siegel et al., 2014). Over 55% lung cancer were diagnosed as metastatic disease with a 5 year survival rate as low as 4%. Similar observation was found for many other types of cancer as well. In ovarian cancer, over 60% cancers were found as distant metastasis, only 15% were detected as localized lesions (Siegel et al., 2014). When diagnosed early, patients will receive treatments when tumors were confined as local diseases, which could drastically increase the chances of being cured. And the 5 year survival will be much higher. Screening for early stage cancer has the potential to reduce the disease morbidity and mortality, to improve the disease prognostication, and to facilitate better disease classification. Currently, in the US, population based cancer screenings are practiced to detect breast, cervical, colon, prostate and lung in defined populations. Generally, they are older people (e.g. >50 year in breast cancer) who might be historically exposed to certain type of environmental carcinogens (e.g. smoking history for lung cancer).

In some cases, implementation of the cancer screenings has been proven to successfully reduce the disease specific mortality. According to the results from several randomized controlled trials, mammography screening was associated with 15-20% breast cancer related mortality (Calonge et al., 2009). Low-dose computed tomography (LD-CT) contributed a 20% reduction in lung cancer specific mortality (Aberle et al.,

2011). These results are highly encouraging for developing screening methods that can benefit broader population.

Multiple factors contribute to these success. First of all, for both lung cancer and breast cancer, currently, effective treatment options are available. When diagnosed early, removing the tumor plus estrogen receptor targeted therapy demonstrated high curing rate and low recurrence in luminal A subtype of breast cancer, which comprises 40% of total breast cancer cases (Parker et al., 2009). In lung cancer, most early stage resectable diseases can be cured by surgery or surgery followed by chemotherapy (Ginsberg and Rubinstein, 1995). Furthermore, according to Surveillance, Epidemiology, and End Results Program (SEER), the life time risk of developing lung cancer is 6.8% among men and women, whereas for breast cancer, it is 12.3% among women. The high prevalence of such diseases justifies the effectiveness of these screening programs, as the risk of having false positives would be relatively lower than other rare cancers.

On the contrary, in prostate cancer, multiple studies showed very limited benefit yet high risk of overdiagnosis and overtreatment as a result of screening using the prostate specific antigen (PSA) test (Andriole et al., 2012; Schroder et al., 2012). It is estimated that over 10 years, approximately 15-20% men will have prostate biopsy triggered by a PSA test (Schroder et al., 2009). About 30% men who have biopsy will have unpleasant experience, such as pain, infection, fever etc (Rosario et al., 2012). The subsequent treatment of diagnosed prostate cancer have also been associated with serious complications. Adequate evidence shows that approximately 0.5% men will die within a month after prostate cancer surgery. And the treatment also leads to long term adverse effects, including urinary inconsistence and erectile dysfunction in 20 to 30% men who

had these treatments (Chou et al., 2011). More importantly, PSA test have also been shown to have limited ability to discern indolent from aggressive disease, resulting in substantial overdiagnosis and overtreatment (Bill-Axelson et al., 2011). Thus US preventative services task force currently recommend against it (Moyer, 2012).

		Disease	
		Present	Absent
Test	Positive	a	b
	Negative	c	d

Figure 1-1. Representation of the outcome of a clinical diagnostic test

1.1.2 Basic statistics for biomarker

To measure the performance of a diagnostic test, sensitivity and specificity are often used. Briefly, sensitivity is a measure of the proportion of positive tests in patients who actually have the diseases, while specificity is the opposite that measures the proportion of negative test in healthy population. In figure 1-1, sensitivity is represented

by a/(a+c); and specificity is represented by d/(b+d). Sensitivity is also often times referred as true positive rate, and specificity is referred as 1 - false positive rate. By varying the test threshold for positive calls, an increase in sensitivity usually leads to a decrease in specificity. To reflect the situation in the actual clinical settings, positive predictive value (PPV) is introduced. It measures the proportion of true positives from all individuals that have positive test, representing the probability of having the disease when an individual is diagnosed positive. In figure 1, PPV is represented by a/(a+b).

1.1.3 Challenges with population screen for cancer

It is important to list out several concerns associated with the current screening scheme in general. First, current cancer screen only focuses on high risk population. The screening tests were only proven to be effective in recommended populations that usually are subjected to very stringent requirements (age, gender, history of smoking, etc.) (Table 1-1). For example, Screening of lung cancer is carried out by LD-CT in population who are between 55 to 80 years old with substantial smoking history (Aberle et al., 2011). Mammography also showed limited ability in detecting cancers with dense breast tissue, which is also associated with younger age. Population not covered by the proposed inclusion criteria may benefit from novel tests that are both efficient and accurate in detecting the disease. As an example, women younger than 50 years old have a higher probability of getting triple negative breast cancers (Foulkes et al., 2010), which are the most aggressive subtype of breast cancer (Dent et al., 2007), but they are not strongly recommended for routine mammography screening (USPSTF, 2009). It is mainly because mammography has low sensitivity in detecting these tumors that are usually associated with dense breast tissues (Dogan et al., 2010; Dogan and Turnbull, 2012). Novel

diagnostic assays targeting such population would have great potential to improve the disease outcomes.

Every screening or diagnostic test has false positive detection. When a diagnostic test does not have a perfect 100% specificity in the designated population, there will be cases of positive diagnosis in healthy individuals. Such proportion will be amplified when the prevalence of the disease is low. The positive predictive value can be calculated using the formula below.

$$ppv = \frac{sens * prev}{sens * prev + (1 - spec) * (1 - prev)}$$

For example, considering a disease with 0.01% prevalence in general population, even a test with 70% sensitivity and 99% specificity will result in only approximately 0.7% positive predictive value, meaning from 1000 people screened positive for the disease, only 7 of them actually has it, the rest are all false positives. As individuals with positive screening test have a greater tendency to be subjected to more invasive tests, usually tissue biopsy, this high proportion of false positives will lead to many unnecessary biopsies. According to the intervention arm of the randomized, controlled Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO), after 14 tests (span 3 years of screening) of various cancer types, the cumulative probability of a subject having at least 1 false-positive test was 60.4% for men and 48.8% for women (Croswell et al., 2009). The cumulative risk of having an invasive procedure for women was 12.3% after 4 tests,

Table 1-1. Current cancer screening practiced in US

Cancer Type	Test	Target Population	Frequency	Proved to reduce cancer associated mortality	USPSTF Grade
Breast	screening mammography	women ages 50 to 74 years old	biennial	Yes	B
Lung	low-dose computed tomography	adults ages 55 to 80 years old; have 30 pack-year smoking history; currently smoke or have quit within the past 15 years	annual	Yes	B
Cervical	Papanicolaou test	women ages 21 to 65 years old	every 3 years	Yes	A
Colorectal	Fecal occult blood test, sigmoidoscopy, or colonoscopy	adults ages 50 to 75 years old	NA	Yes	A
Prostate	Prostate specific antigen test	men ages 55 to 74 years old	annual	No	D

and 22.1% after 14 tests. For men, the probability is 17.2% after 4 tests, and 28.5% after 14 tests (Croswell et al., 2009).

Overdiagnosis is also a concern related to cancer screening, especially for prostate cancer. Most of the time, it refers to the detection of lesions with no malignant potential. Individuals with overdiagnosed cancers may experience all possible adverse effect from the subsequent invasive tests and potential treatments while not receiving any benefit from the screening. Overdiagnosis often leads to unnecessary follow up tests and treatments. Studies also showed the correlation between cancer diagnosis and suicide behavior with relative risk between 1.6 to 2.0 (Baade et al., 2006; Lu et al., 2013). On the other hand, overdiagnosis could also mean the detection of a slow-growing lesion that would not reach lethality before the individual die of a competing cause of death. A special yet frequent case is that a significant fraction of patients with advanced cancer continue to be screened for new cancers. Study estimated that 8.9% women with advanced cancer still receive routine screening mammography. And 15% of men with advanced cancer diagnosis still receive PSA test (Sima et al., 2010).

Except for Pap smear test used for cervical cancer screening, other screening methods highly rely on imaging technology (Table 1-1). Both mammography and low-dose CT depend on X-ray, which put patients under risk of radiation exposure that could potentially cause DNA damage and mutation. This will increase risk of developing cancer. Colonoscopy and sigmoidoscopy are used to screen colon cancer, but themselves are both considered invasive as they could lead to potential complications like perforation and tears in the linings of colon (Imperiale et al., 2000; Lieberman et al., 2000).

In this setting, there is intense effort in the search for non-invasive biomarkers that can detect early disease and monitor for disease progression and recurrence. With the advent of molecularly-targeted therapeutics, biomarkers that can describe biological subtypes of cancer may also be useful for predicting responses to therapeutic interventions and side effect profiles.

1.2 Blood-based diagnostic biomarker for cancer

1.2.1 Overview

An ideal screening test should be performed non-invasively in ready to access samples, such as blood. Traditional biomarker discovery strategy highly relied on knowledge-based, hypothesis-driven testing. By understanding the genetic and molecular alterations underlying the cancer pathogenesis, one can make reasonable projections of detecting the altered molecular events in blood samples. For example, with the knowledge that cancer is a disease caused by genomic mutation, several groups attempted to look for circulating DNA molecules carrying the mutation in serum. Circulating DNA carrying *KRAS* and *BRAF* mutation can be found in blood of patients with colorectal cancer (Bettegowda et al., 2014; Thierry et al., 2014).

The above strategy can only query a small fraction of potential candidates based on the existing knowledge. As a disease with complex etiology, cancer presents great heterogeneity that it is still far from being fully characterized. An alternative strategy relies on the recent development in genomics and proteomics, which enables high throughput profiling of complex biological specimen. This allows us to carry out unbiased systematic screening in order to identify disease-associated biomarkers. Technological advancement in the field of mass spectrometry, protein arrays, next generation sequencing, and cell sorting have facilitated innovative studies on markers that could benefit the early diagnosis of cancer. To achieve this, one needs to compare level of biomarker candidates between patients and disease-free controls usually matched by demographic characteristics to the cases. Differentially presented features can then be selected and further evaluated of their association with the disease.

1.2.2 Literature review of blood based cancer biomarker

Many studies had focused on early detection biomarkers of various types of cancers by applying genomics or proteomics approaches. Biomolecules from various sources have been investigated to search for markers of potential diagnostic and prognostic values. Pietrowska et al. developed an classifier consisted of three spectral components to differentiate patients from healthy controls with 83% sensitivity and 85% specificity by focusing on low molecular weight serum proteins (Pietrowska et al., 2009). Abd Hamid et al. reported a monogalactosylated triantennary structure containing α1,3-linked fucose had a 2 fold increase in breast cancer patients, and a pilot glycoproteome signature of α1-acid glycoprotein, α1-antichymotrypsin and haptpglobin β-chain may mark the onset of metastatic breast cancer (Abd Hamid et al., 2008). Hyung et al. discovered 6 serum protein biomarkers that potentially can predict breast cancer's responses to neo-adjuvant chemotherapy using liquid chromatography-tandem MS method (Hyung et al., 2011). Besides protein and glycan biomarker, circulating miRNAs were also reported to associate with breast cancers in their early stages (Heneghan et al., 2010; Zhao et al., 2010). Huang et al measured plasma level of 12 miRNAs in patients with advanced colorectal neoplasia, and found that miR29a and miR92a are elevated in patients' plasma (Huang et al., 2010). However, the detection of miRNA in blood samples suffers from its instability and the use of appropriate reference gene set (Berger and Reiser, 2013). Circulating tumor cells (CTCs) could also be an indicator of cancer progression (Mostert et al., 2009), but due to the heterogeneity of CTCs and its rareness in blood (1 CTC in 10^6~10^7 leukocytes) (Hong and Zu, 2013; Mostert et al., 2009), enrichment and investigation of CTCs have been extremely difficult.

1.2.3 Challenges in biomarker study

With this great amount of publications in biomarker field, very few of these reported markers are successfully adopted in real-time clinical applications. This fact reflects significant challenges faced by biomarker research and development. A summary of several review articles here will bring up the causes and possible solutions to this issue.

1.2.3.1 Disease heterogeneity

Disease heterogeneity may contribute to the irreproducibility and low sensitivity of reported biomarkers. It is now well accepted that many diseases, including cancer, arising in the same organ, sharing similar symptoms and histological alterations, may comprise multiple subtypes of the disease. Molecular profiling revealed that breast cancer consists of at least 7 subtypes, which differs in their molecular mechanism, aggressiveness, responses to treatment as well as prognosis (Perou et al., 2000). TCGA analysis of non-small cell lung cancer revealed a comprehensive mutation spectra, corresponding to large variations in disease etiology. The prevalence of certain driver mutations could be as low as 1% of the patients. As these subtypes had distinct tumor driving profile, it is reasonable to expect different markers associated with each subtype. Hence, the sensitivity of each individual marker is capped by the prevalence of that subtype it associates with in the studied cohort (Wallstrom et al., 2013). Very few previous biomarker studies in cancer had addressed the issue of the disease heterogeneity. As the subtype distribution may vary from study to study the sensitivity of a subtype specific marker could be affected by the subtype's prevalence in the patient cohort, making it difficult to interpret and repeat the results.

1.2.3.2 Intended use sample

As pointed out by many, groups that perform biomarker discovery experiments usually are lack of access to clinically relevant samples. Many studies were conducted using convenient samples that are not in early stage disease, or having been exposed to treatments, or not having properly matched control subjects. The justification usually is that after the proof of concept experiment, it will be much easier to get funding that will support the access to source with better samples. However, when the initial sample sets used are not intended for disease diagnosis, it is highly possible that 'biomarkers' derived from these studies are irrelevant. To address this fundamental issue, researchers need to use patient/control samples that reflect their 'intended use', that is to say these samples should represent the population that will be focused on when the biomarker assay is implemented.

1.2.3.3 High dimensional data

The new era of large scale molecular profiling of genomics, epigenomics, transcriptomics, proteomics or metabolomics of diseases enables the search for biomarkers in all possible candidate pools. However, most biospecimens are only available in relatively small numbers. Thus, many studies relied on few specimens while screening in thousands of analytes. Such high dimensionality is generally statistically underpowered to conclude the performance of a biomarker candidate. For example, when a single feature is tested significant to differentiate two groups with ($p<0.05$), the probability of the feature to be a false positive is less than 0.05. However, when such comparison was performed in parallel at a large scale, the probability of making incorrect significant conclusion will get inflated. This will dramatically increase the risk of

overfitting. To avoid this problem, we can use statistical adjustment, such as false discovery rate (FDR), Benjamini-Hochberg adjustment, Bonferroni method, to correct for the over-promising p value. Sometimes, it is difficult to estimate the human variation. Another approach proposes to test the selected biomarkers again in an independent sample set in a blind fashion. This is the most stringent way to evaluate the performance of markers selected from a screening study. Markers that cannot survive such procedure may imply that they are chosen by chance due to sample variation.

1.2.4 A five-phase strategy to develop and assess a cancer screening test

To facilitate the development of biomarker-based screening tools for cancer early detection, Pepe et al. proposed a five-phase guideline to achieve clinical utility (Table 1-2) (Pepe et al., 2001).

The first phase is to identify leads for potentially useful biomarkers and prioritize them using assays that are both reliable and reproducible. Sensitivity and specificity of each individual biomarker candidate should be calculated, and Receiver operator characteristic (ROC) curve can be used to assess the marker's general ability to differentiate disease from controls. To determine required sample sizes, they suggested that computer simulations of hypothetical data generated by biologically plausible models could be used. By varying the number of cases and controls, the sample sizes can be determined in order to select a reasonable proportion of biomarkers for further study. In this phase, a list of candidate biomarkers should be identified, and ideally, it is necessary to perform a confirmatory study with the use of an independent set of well-controlled samples.

Table 1-2. Five phases of biomarker development

Phase	**Objectives**
Phase 1 - Preclinical Exploratory	To identify potentially useful biomarkers
Phase 2 - Develop clinical biomarker assay	To estimate sensitivity, specificity
Phase 3 - Retrospective and Longitudinal	To evaluate the biomarker's capacity of detecting preclinical disease
Phase 4 - Prospective screening	To determine the detection rate and the false referral rate
Phase 5 - Cancer control	To estimate the reduction in cancer mortality

The second phase aims to develop the clinical assay that can differentiate patients with cancer from cancer free subjects in a simple and reproducible approach, so that it can be carry out for wide spread screening. And the sensitivity (TPR) and specificity (1-FPR) for the clinical biomarker assay should also be estimated at this stage. To achieve this, ideally, both cases and controls should well represent the target screening population. This means that control subjects should better be recruited through a population based study, because convenient control samples may differ systematically from the intended screening population. They may be referred to the clinic for some health reasons instead of being randomly chosen. In this phase, the effect of demographic factors, such as age, sex, smoking behavior, on biomarkers should also be assessed. Similarly, factors of disease characteristics, such as stage, tumor grade, node status, histology etc. should also be examined for their potential effects on biomarker in patients.

Once the performance of the clinical assay is confirmed in phase 2, it is then necessary to evaluate how early the biomarker will appear. Phase 3 aims to assess the capacity of biomarker assay to detect pre-diagnostic disease using prospectively and

longitudinally collected samples. If the biomarker in cases appears to deviate from controls months or years prior to clinical diagnosis, it shows high potential for screening. Otherwise, if the marker only appears to change close to the time of diagnosis, there is little promise. Once a marker presents promising performance specific criteria for a positive test should be defined for phase 4. Proper algorithms for combinations of biomarkers should also be developed.

Phase 4 is carried out in a prospective setting where the screening test is applied to the target population in order to determine the proportion of screened subjects who test positive and have the disease. In this phase,a positive test will lead to definitive diagnostic procedures and potential treatments. At phase 3, information on cancer stage and characteristics at time of detection cannot be obtained. As a prospective study, such characteristics can be examined in phase 4. However, since only subjects with a positive screening test are referred for further invasive diagnostics, false negatives cannot be identified. Therefore, neither sensitivity nor specificity can be estimated from this phase. Instead, it is possible to evaluate positive predictive value and the proportion who are tested positive but do not have the disease.

The final phase assesses whether the population can benefit from the screening test. It is also possible that even if cancer can be detected early by the screening test, there may still be limited overall benefit for the screened population. In another word, early detection may not contribute to the extension of patients' life span. A related concept is lead-time bias, which refers to the observed longer survival as a result of earlier detection can not reflect the actual benefit of the early detection, the patients may eventually die at the same time with or without being screened. Several factors need to be

taken into consideration. For example, whether the treatment strategy is effective for the screen-detected tumor. False positive results associated with the screening that leads to unnecessary invasive procedures and stress. Whether the screening test will lead to overdiagnosis of un-significant cancer that will eventually self regress. More importantly, whether the cancer can be diagnosed early enough so that effective treatment can be applied. In the case of ovarian cancer early detection using CA125, the biomarker does not appear in the blood early enough to detect the disease at a time when we can successfully intervene. To demonstrate the effectiveness, the study should be conducted ideally as a randomized trial with one arm of population undergone screening, and the other arm not. The ultimate assessment will be the life span of subjects after entering the study.

1.3 Cancer autoantibodies

1.3.1 Cancer patients can develop autoantibodies against tumor

The detection of tumor-shed proteins in serum when tumor is still at early stage as previously described may be a challenging task. Proteomics-based approaches to distinguish cancer-bearing patient sera from healthy control sera have been challenged by the difficulty in identifying small quantities of protein fragments within complex protein mixtures, protein instability, and natural variations in protein content within patient populations. One great challenge facing the search for blood based early detection markers is the extremely low concentration of target molecules/cells shed from the cancer tissues in early stage of cancer development. Sometimes, the process is so slow that it takes years for the tumor specific markers to increase to an amount detectable by a clinical assay. It has been estimated that ovarian cancer could grow occult for 10 years before it can be detected by clinically available CA125 assay (Hori and Gambhir, 2011; Lutz et al., 2008). On the other hand, since the abundance of various proteins in blood sample could be distributed at a scale of 10 orders of magnitude, the few most abundant proteins will potentially be overwhelmingly represented in the spectra of detected proteins, masking the lower abundant peptides. As an example, plasma protein biomarkers detected by human proteome organization (HUPO) project were among the most abundant molecules in the range of μg/ml to mg/ml (Figure 1-1). Thus, it is necessary to consider alternative approaches.

The immune system of our human body mounted a thorough surveillance to recognize nonself invaders from self. It is realized through the adaptive immune responses by generating target specific B cells / T cells. This highly resembles an

intelligent security system, which can detect and destroy a wide variety of infections. Antibodies are produced by B cells to specifically bind to their target molecules. This binding event could either neutralize the pathogenicity of the invader, or trigger antibody dependent cytotoxicity in which such responses will also be memorized when the infection has been removed. Compared to other proteins in the human blood, antibodies have many appealing features as potential biomarkers. First, although tumor proteins may circulate in low concentration potentially due to their low shedding amount, rapid degradation/clearance in blood, their corresponding antibody responses are likely to remain strong. Antibodies are also not subject to common types of proteolysis that affects other proteins, making them highly stable in serum samples. In fact, their half life in blood is over 7 days, so there are limited daily fluctuations, making it easy to collect. Finally, there are many well-developed secondary reagents for their detection. These features make them an ideal repertoire to mining for potential cancer biomarkers.

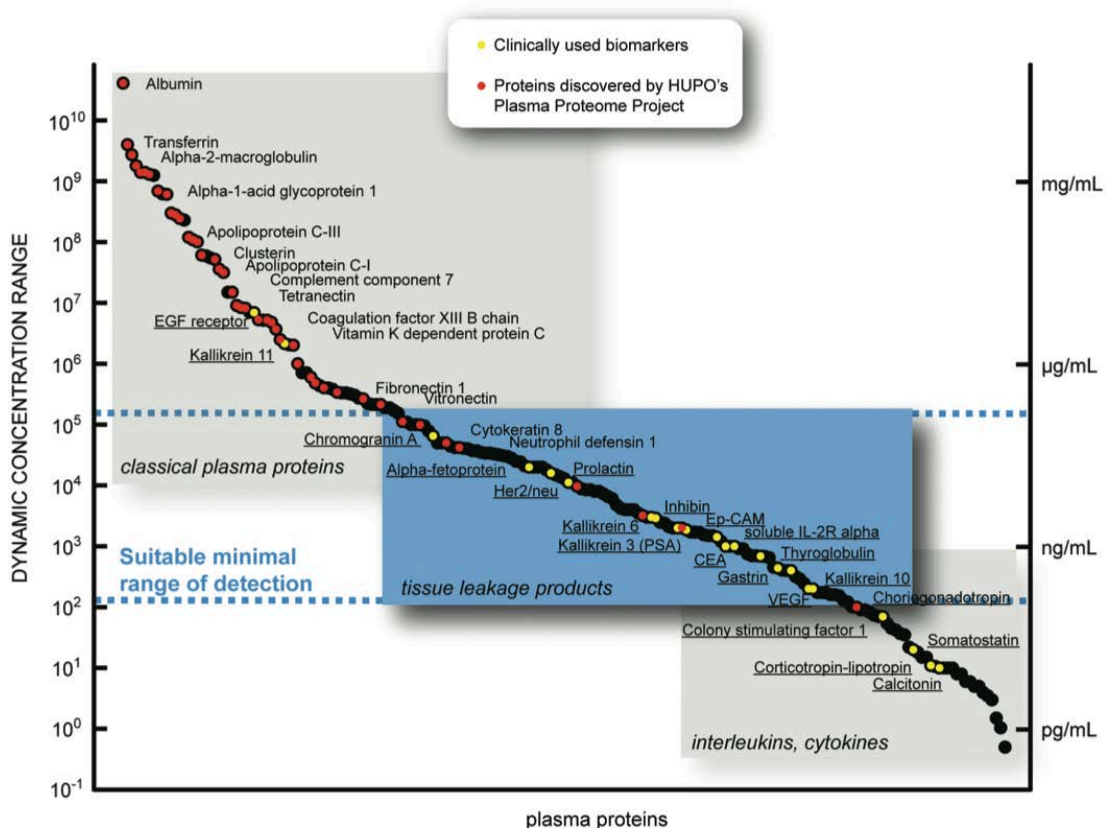


Figure 1-2. Dynamic protein concentration range in human plasma sample. Figure is adapted with permission from Surinova Silvia et al. 2011.

To accomplish the specific recognition, the immune system requires complex mechanisms to distinguish foreign from self. B cells and T cells involved in adaptive immune response will be deleted or undergone anergy, rendering tolerance to molecules that the immune system will encounter in routine circulation. However, this is not always the case in that sometimes the immune system responds to self-derived antigens. It might be due to their location, abundance, altered post-translational modification or mutations. Cancer patients have been observed to produce antibodies to self-proteins that are expressed by their tumors (Figure 1-2). The specific mechanism of which the immunogenicity develops is still unclear, and might differ among cases.

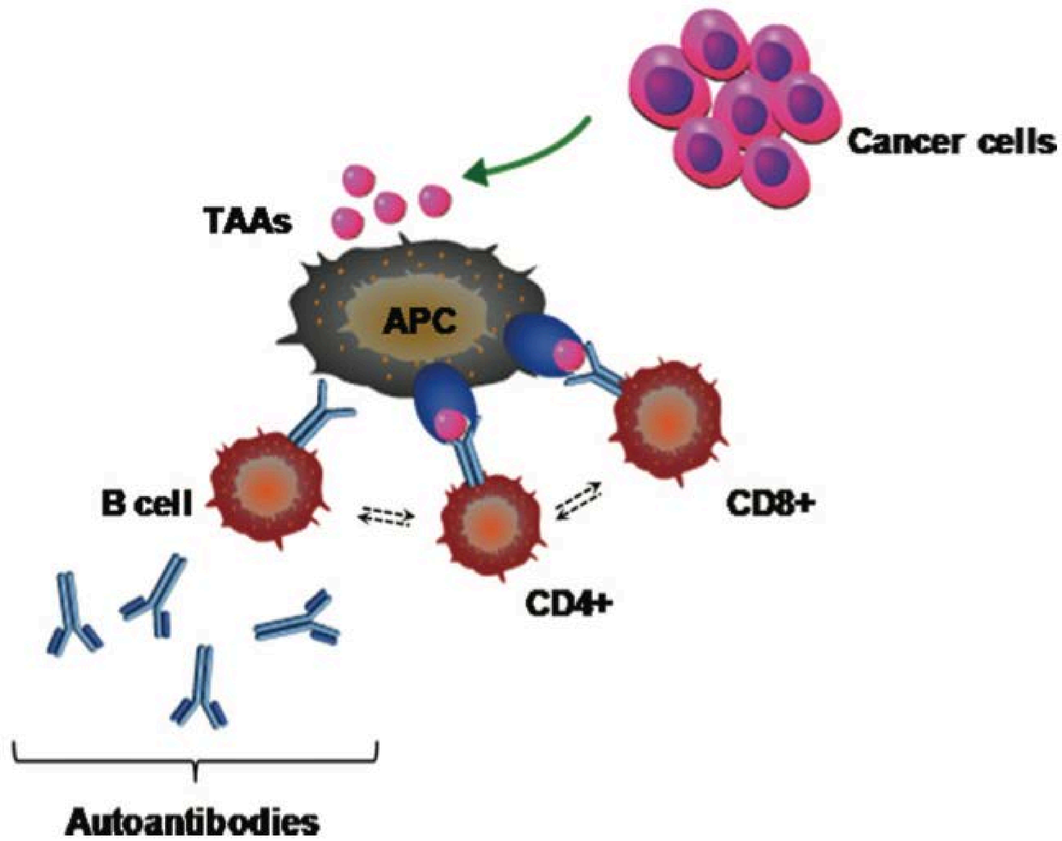


Figure 1-3. Schematic representation of the development of autoantibody responses against tumor associated antigens (TAA). Figure is adapted with permission from Solassol Jerome et. al. 2011

To identify the specific targets of B cells and T cells from tumors may help us identify potential biomarkers for diagnosis, prognosis and classification of cancer. It could also help us understand the immune regulation on cancer progression. These cancer-associated antigens may also serve as potential immunotherapy targets.

1.3.2 Efforts in searching for cancer associated autoantibodies

Over the past two decades, serological analysis of cancer yielded a list of autoantibody biomarkers in several separate studies. In breast cancer, AAb to HER2/neu was first discovered in 1997 by Disis et al. In that study, they analyzed antibody response to HER2 protein in 107 newly diagnosed breast cancer patients and 200 volunteer blood

donors, and detected anti-HER2 antibodies in 11% patients, 0% controls. They also demonstrated that 20% patients with HER2 positive tumors had antibodies (Disis et al., 1997). In one study, using mini protein array, a panel of seven antigens (c-myc, cyclin B1, IMP1, Koc, p53, p62, and survivin) were shown to be able to distinguish different cancers (breast, colorectal, gastric, hepatocellular, lung and prostate) from healthy controls with sensitivities from 0.77-0.92, and specificities from 0.85-0.91 (Koziol et al., 2003), but these markers still have not been independently validated by different groups. Chapman and colleagues tested AAb responses to a panel of six antigens (p53, c-myc, HER2, NY-ESO-1, BRCA2 and MUC1) using 97 patients with primary breast cancer and 94 normal controls. Their results indicated 64% of patients respond to at least one of the six antigens, while the specificity was 85% (Chapman et al., 2007). However, although the specificities of individual protein can reach as high as 91-98%, their sensitivities were relatively low in early stage breast cancer. Desmetz et al. found the combination of five autoantibody biomarkers (FKBP52, PRDX2, PPIA, HSP60 and MUC1) can significantly discriminate primary breast cancer (AUC = 0.73) and carcinoma in situ (AUC = 0.80) from healthy controls among women under the age of 50 (Desmetz et al., 2009). The same group recently reported in a separate study that using five markers (GAL3, PAK2, PHB2, RACK1 and RUVBL1), they discriminated early stage breast cancer from healthy individuals (Lacombe et al., 2013). With a panel of five antigens (RBP-Jk, HMGN1, PSRC1, CIRBP and ECHDC1), Mange et al predicted DCIS to invasive breast carcinoma (IBC) transition (Mange et al., 2012). Our group also discovered a signature of 28 autoantigens that can discriminate malignant breast cancer from benign breast disease (Anderson et al., 2011a). However, none of these studies distinguished subtypes. Given

the distribution of different subtypes, these previously reported biomarkers would presumably perform the best in ER+ breast cancer.

Autoantibodies are also extensively searched in lung cancer. Using 2D gel electrophoresis, western blotting and mass spectrometry identification, Pereira-Faca et al. discovered that AAb to 14-3-3 theta was significantly higher in cancer sera (p=0.0042). The combined performance of 14-3-3 theta with annexin I and protein gene product 9.5 (PGP9.5) gave a sensitivity of 55% at 95% specificity (AUC=0.838) (Pereira-Faca et al., 2007). Qiu et al. conducted a protein microarray study to measure AAb responses in sera from 85 pre-diagnostic lung cancer patients and 85 matched controls. In addition to annexin I and 14-3-3 theta, laminin receptor 1 (LAMR1) were also identified to show high reactivity in lung cancer patients' sera. The panel achieved 51% sensitivity and 82% specificity (Qiu et al., 2008). He et al. detected AAb responses to alpha-enolase in 26 of 94 non small cell lung cancer patients' sera, while only in 1 of 60 healthy controls', and none of the controls subject with small cell lung cancer (n=15), gastrointestinal cancer (n=18), or Mycobacterium avium complex infection of lung (n=9). Similar to above studies, they also reported a panel of three markers (alpha-enolase, carcinoembryonic antigen and cytokeratin 19 fragment) enhanced sensitivity for the diagnosis of NSCLC (He et al., 2007). Recently, Chapman and colleagues developed an autoantibody assay, named EarlyCDT®-lung test, to aid early detection of lung cancer. This test included a panel of seven antigens (p53, NY-ESO-1, CAGE, GBU4-5, SOX2, HuD, and MAGEA4) or a panel of six antigens (p53, NY-ESO-1, CAGE, GBU4-5, Annexin I and SOX2). In one study, they demonstrated the utility of these panels in 776 and 836 individuals with high risk of lung cancer. The seven-AAb panel give a sensitivity of 41% at 91%

specificity, while the six-AAb panel gave a sensitivity of 39% with a specificity of 89% (Chapman et al., 2012). However, currently there is limited number of studies that focused on distinguishing lung cancer patients from benign controls within the LD-CT test positive population.

1.4 Emerging technologies for profiling autoantibody responses

The greatest challenge associated with using AAb as biomarkers is their low prevalence in cancer patients (sensitivity), reflecting the nature of cancer as a heterogeneous disease. Different molecular subtypes of diseases, even when they share the same pathological diagnosis, may produce subtype-specific biomarkers. This gives rise to an apparently low sensitivity for each biomarker, which will only identify those samples that came from individuals who match the molecular diagnosis that produces that biomarker. For example, if a molecular subtype is present in 20% of a disease population, then even when a biomarker for that subtype behaves perfectly, it will only detect that 20% of cases who have that molecular subtype. Its sensitivity for the disease is effectively capped at 20%. Moreover, the interaction between patients' immune system and the tumor may vary from person to person. This in turn may result in different AAb responses even within a subtype. The reported sensitivity of a single antigen is typically 5-20%. To build an effective diagnostic test, it has been demonstrated that by simultaneously analyzing multiple markers and combining them into a panel, higher sensitivity can be achieved (Anderson and LaBaer, 2005). Proteomic scale profiling of AAb responses may provide the potential of identifying panels of antigens that can be utilized in diagnosing and detecting cancers (Surinova et al., 2011).

To identify serum antibody biomarkers, many high-throughput screening assays have been developed. Due to the advances in immune-proteomics in the past decade, many new disease associated antibody markers have been discovered. Here, several widely used technologies for antibody profiling are summarized below.

1.4.1 SEREX

Serological screening of recombinant cDNA library using phage display (SEREX) was first developed 20 years ago, and it had facilitated the identification of hundreds of new AAbs (Scanlan et al., 2002). To perform SEREX, a cancer related cDNA library is first reverse transcribed from RNA of cancer tissues or cell lines, and then constructed into λ-phage vectors. Proteins are expressed recombinantly in *E. coli.* These clones that displays individual tumor peptides are transferred to nitrocellulose membranes and incubated with serum samples from cancer patients and healthy controls (Figure 1-3A). Antigens that are specifically detected in cancer patients are then sub-cloned and sequenced. Using this method, thousands of tumor associated antigens had been discovered, including the cancer testis antigen NY-ESO-1 and SSX2 in esophagus cancer and melanoma (Chen et al., 1997; Sahin et al., 1995), NY-BR-1 to NY-BR-7 in breast cancer (Jager et al., 2001), NY-CO-37 and NY-CO-38 in colon cancer (Scanlan et al., 1998). After the initial discovery, specific phage clones can be produced to develop immunoassays to obtain sensitivity and specificity values, but the identification of the cancer specific clones involves extensive screening and characterization of each serum sample against the library. Although size of the library could be as high as 10^6, the construction of the initial cDNA library retains the biased level of each mRNA. Genes overexpressed could be over-represented. When cancer specific antigens are expressed at low level, they may not even be displayed during the screening. More importantly, the library usually contains substantial amount of reverse transcribed cDNAs from frame shift events and non-coding RNA, the resulted candidate list may not be relevant to the disease. Additionally, antigens expressed by *E. coli* system might not be in their native

conformation, particularly, they may not have proper mammalian post-translational processing. The subsequent verification has to be performed individually, which involves substantial amount of labor-intensive experiments.

1.4.2 SERPA

Serological proteomic analysis (SERPA) was proposed to overcome above shortcomings of SEREX technology (Seliger and Kellner, 2002) (Figure 1-3B). In this approach, in order to identify disease-associated AAbs, thousands of proteins in complex cell lines or tissue specimens are first separated based on their sizes and isoeletric point using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). Then proteins are transferred onto membranes and incubated with patients or healthy controls' serum samples. Spots that only react with patients serum sample can be tracked back to the location on the original protein gel. The corresponding spot will be analyzed by mass spectrometry to determine the protein identity. This method is relatively reproducible, but requires large amount of serum samples. As an antigen discovery platform, it lacks the desired throughput and dynamic range (Anderson and LaBaer, 2005).

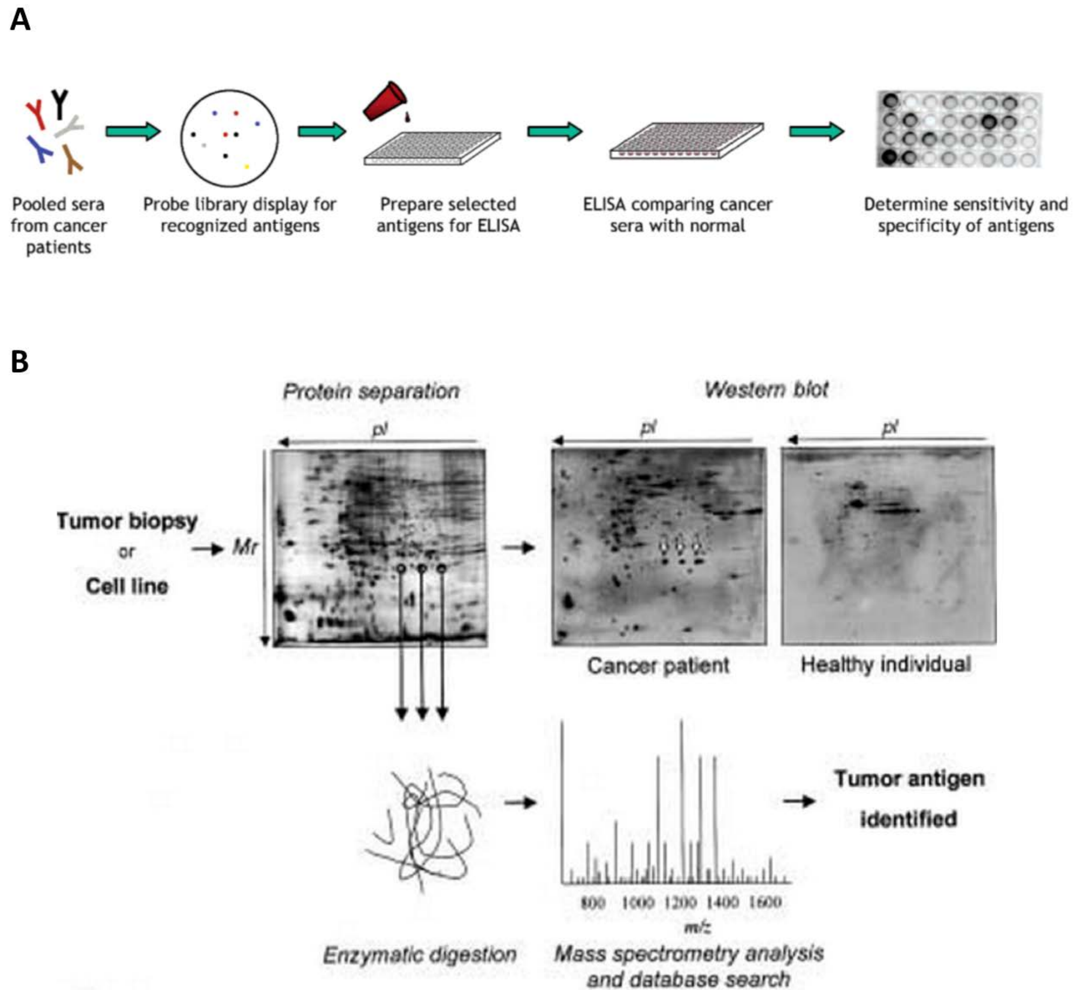


Figure 1-4. Scheme for SEREX and SERPA. **A**. Scheme for SEREX. Figure is adapted with permission from Anderson, et al. 2005 (Anderson and LaBaer, 2005). **B**. Scheme for SERPA. Figure is adapted with permission from Seliger, Barbara et al. 2002 (Seliger and Kellner, 2002).

1.4.3 Protein microarray

Protein array technology was first invented more than a decade ago by MacBeath et al (MacBeath and Schreiber, 2000). Using micro spot arrayer, individual protein samples can be spotted at proteomic scale onto single glass slides. It opens up the possibility of testing antibody responses simultaneously against thousands of protein

targets. Inspired by this idea, many different protein array platforms have been developed.

An early form of protein array expanded the phage display technology by combining a biopanning process to enrich cancer specific antigens and a protein array spotted with the selected phage clones (Wang et al., 2005) (Figure 1-4). To achieve this, a tumor cDNA library is first constructed into T7 phage vectors similar to SEREX. Instead of transferring on to membranes, the phage library is incubated with serum samples from healthy controls to eliminate the non-cancer related antigens from the library. The cleared library is then incubated with patients' sera to enrich for cancer specific antigens. The resulting library can be used to transfect *E. coli.* to reach monoclonality and to propagate. Single phage clones each bearing a single cancer peptide can be spotted on a coated glass slide by a robotic arrayer. These arrays can be used to test autoantibody responses in cases and controls. Clones that react preferentially with cases can be selected and sequenced to confirm the protein identity. The advantage of this technology is that the phage clones printed on array are pre-selected to react with cancer patients' sera while not healthy controls. However, peptides carried by phage are usually short, and due to the nature of cDNA library, many non-coding sequence had also been selected, making it difficult to interpret the result. The pre-selection process could also overlook the possible situation where some AAb that exists in controls' sera may be elevated in cancers. The first negative selection step may be too conservative.

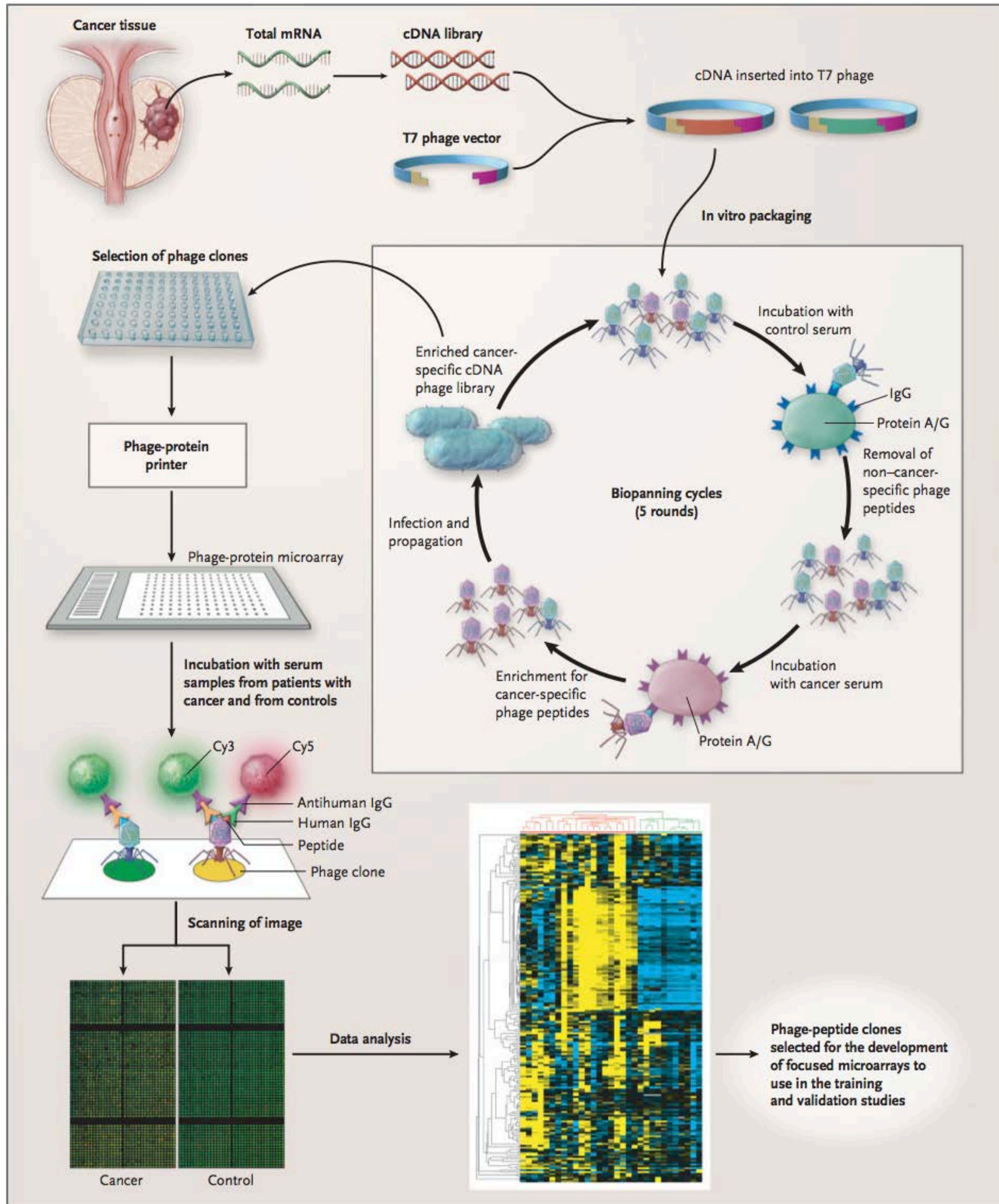


Figure 1-5. Phage display protein array used to screening autoantibody for prostate cancer. Figure is adapted with permission from Wang, Xiaoju et al. 2005 (Wang et al., 2005).

The most widely used is the recombinant protein array, which uses purified proteins expressed by bacteria, yeast, or insect cells. This approach has been comprehensively explored to display individual proteins from various species to study

infectious disease, autoimmune disease, as well as cancer. Zhu et. al. reported a protein array containing more than 17,000 purified human proteins expressed by yeast (Fan et al., 2014). A widely used commercially available protein array is Protoarray® manufactured by Thermo Scientific (Yang et al., 2011a). These proteins are produced and purified in a high-throughput way from Sf9 insect cell line, which is the closest to mammalian expression system adopted by groups manufacturing purified protein arrays. Purified proteins were then spotted on a nitrocellulose membrane under 4°C. This approach allows the rapid identification of disease-associated antibodies. Since the protein identity of each spot is known, the corresponding sensitivity and specificity can be directly calculated using the fluorescent signal. Generally, very small amount (microliters) of serum sample is required. However, this approach mainly depends on the success of protein purification. Although the process has been optimized to be conducted in a high-throughput way, it is still not a trivial procedure that involves substantial labor and cost in order to ensure the protein purity and quality (Qiu and LaBaer, 2011). One major challenge for this technology is that printed arrays have limited shelf life that proteins will typically lose their function after a few weeks. This type of protein array has to be stored at low temperature and dry condition (Romanov et al., 2014). In addition, protein yields from the high throughput purification preps could vary widely from spot to spot, and 90% purity is usually considered satisfactory. Both potentially add extra layers of uncertainty during the plasma AAb profiling.

To retain the native features of proteins from tumor, tumor cell lysates can be fractionated by liquid based chromatography to separate various protein species. The fractionated cell lysates can then be printed onto a membrane or microarray, which can

be probed with patients or controls' sera (Qiu et al., 2004). Advantages of this approach include the preservation of protein characteristics from native cancer tissue, and the array potentially displays the entire cancer proteome. However, the protein identity has to be determined in order to develop a clinically validated assay. Generally, this can be achieved by sensitive mass spectrometry. But as each fraction may contain a spectrum of proteins with a large range of concentrations, the immunogenic component may be at very low level, which could be below the detection limit. Another disadvantage is the difficulty to control proteins' orientation on array.

1.5 Cell-free protein microarrays

1.5.1 Nucleic Acid Programmable Protein Array

To extend the protein array shelf life, and to circumvent the tedious and troublesome protein purification process, there were several attempts to develop platforms that take the advantage of cell free expression system by printing DNA that encode the end protein product.

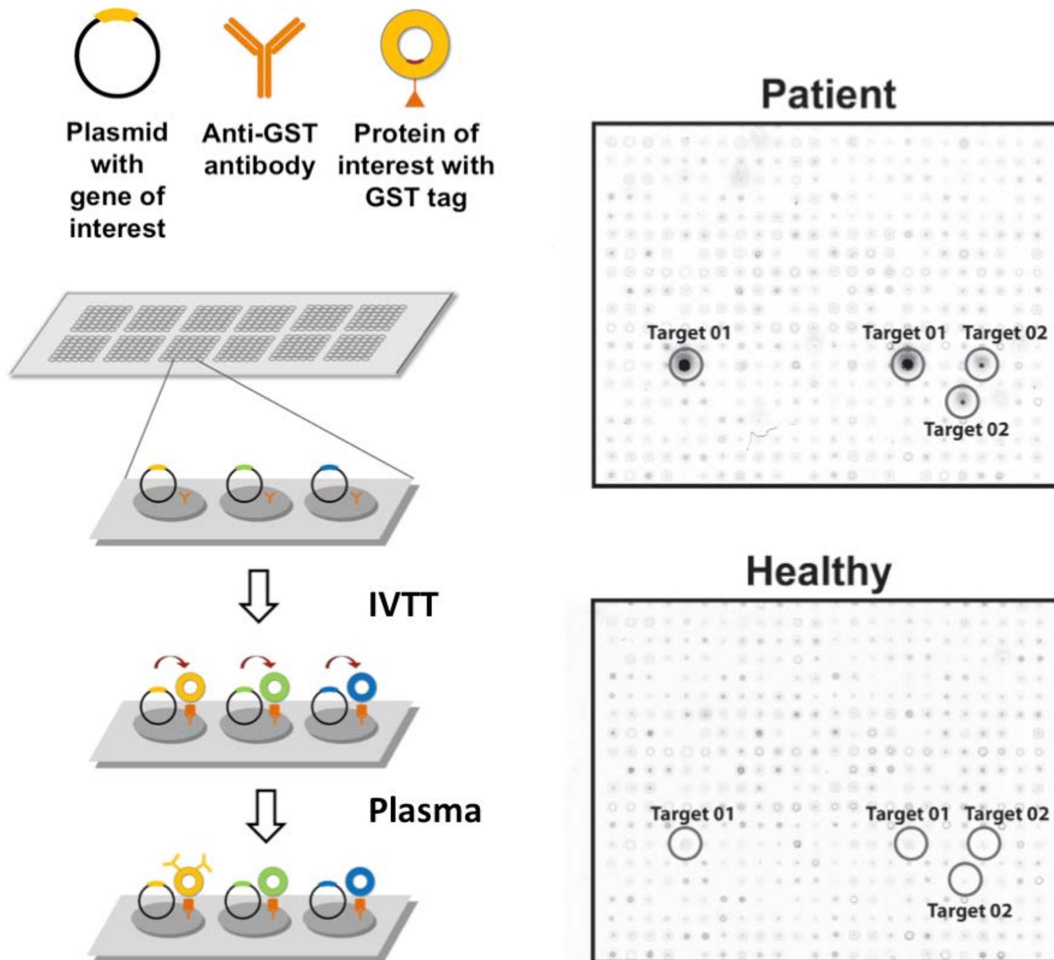


Figure 1-6. Scheme for plasma screening using NAPPA and representative images of array probed with plasma from patient or healthy donors.

Inspired by the high-throughputness of protein array and the cell free expression system, our group developed a platform that coupled DNA plasmid immobilization and

in vitro transcription and translation, called nucleic acid programmable protein array (NAPPA) in 2004 (Ramachandran et al., 2004). Proteins of interest were encoded by the printed plasmids on each spot. A fusion protein product with a n-terminal or c-terminal tag can then be transcribed and translated using a cell free expression system. Affinity reagent that can capture the protein product through interacting with the tag was also co-spotted with the DNA plasmid (Figure 1-5). The current printing chemistry adopted by NAPPA relies on bovine serum albumin (BSA) to drastically increase DNA-binding. Capture anti-GST antibody and BSA are coupled to the amine-coated glass slide through bis(sulfosuccinimidyl)suberate (BS3), an amine-to-amine cross-linker that is homobifunctional, water-soluble, non-cleavable and membrane impermeable.

To manufacture NAPPA, it requires the access to a large collection of cloned and well-annotated cDNA resource. In our case, DNASU (http://dnasu.org) provides us a great collection of cDNAs from multiple species (Seiler et al., 2014). Target proteins of interest were cloned robotically in a high-throughput way using the Gateway cloning technology. And they are stored under -80°C as form of glycerol stocks. Currently, over 200,000 individual cDNA clones are available to the public through DNASU from over 700 species, ranging from virus, prokaryotes to human. Currently, DNASU holds over 74,000 plasmids that contain human genes representing around 13,000 unique proteins.

NAPPA displays proteins that are freshly produced in a human milieu without the extensive processing during protein purification. The *in situ* produced proteins were maintained under physiological condition until assayed within hours and presumably contains proper mammalian post-translational modifications including but not limited to phosphorylation, glycosylation. This enables the high-throughput analysis of protein

functions. In fact, NAPPA had already been demonstrated to display functional protein that can be used for testing protein-protein interaction (Yu et al., 2015), post-translational modification (Yu et al., 2014b), enzymatic activity (Yu et al., 2014b), as well as antibody profiling in complex biological samples (Anderson et al., 2015; Anderson et al., 2011a; Miersch et al., 2013; Montor et al., 2009; Yu et al., 2014a).

1.5.2 Other cell-free protein array platforms

Similar to NAPPA, Protein *in situ* array (PISA) was developed to display proteins by *in situ* transcribing and translating the corresponding immobilized cDNA template. Proteins were then captured through the interaction between a double (His)6-tag fused to the protein and nickel ion-affinity nitrilotriacetic acid ligands on the slide surface (He et al., 2008). Another method adapted mRNA display technology to the protein production in order to cross-link the nascent protein with the puromycin linked DNA (Tao and Zhu, 2006). In this approach, DNA templates were firstly amplified by PCR, and transcribed in situ to mRNA. mRNA products were then hybridized with a single-stranded DNA modified with biotin and puromycin. The hybridized RNA-DNA complexes were printed on a streptavidin-coated slide. When proteins were translated, ribosomes would stall at the RNA-DNA hybrid section, and puromycin would allow proteins cross-link to the DNA oligos which were immobilized onto the slide through the binding of streptavidin and biotin.

A challenge associated with cell free protein array technologies is that during *in vitro* transcription translation, proteins were easy to diffuse to adjacent spots. Recently, our group also developed a newer generation of NAPPA which prints in photolithographically etched discrete silicon nanowells (Takulapalli et al., 2012). Protein

expression in confined wells dramatically reduced protein diffusion and cross-spot binding. Array density increased from 2,500 to 24,000.

1.6 Project Overview

The goal of this doctoral dissertation is to study the autoantibody profiles of basal-like breast cancer and lung cancer using immune-proteomics approach, and to assess possible clinical applications of the identified AAbs to the detection of the diseases. To achieve this goal, specific objectives were set as follows:

i) Develop and optimize current plasma screening strategy using NAPPA platform.

   a. Optimize traditional NAPPA to increase signal to noise ratio when probing complex plasma samples.

   b. Adapt new surface protein capture chemistry that uses HaloTag and HaloTag ligand covalent interaction.

   c. Apply the platform to detect antibody responses against both conformational and linear epitopes in plasma samples.

ii) Conduct the most comprehensive proteomics-scale screen of autoantibody responses to 10,000 human proteins in basal-like breast cancer

   a. High throughput screening of AAb responses in basal-like breast cancer patients and controls plasma sample to identify disease associated AAb candidates.

   b. Verify the association of these candidate AAbs to BLBC in larger sample set using clinically relevant ELISA assay.

   c. Investigate the association between AAb responses and protein expression, mutation, patients survival.

iii) Profile autoantibody responses to 10,000 human proteins in non-small cell lung cancer

a. High throughput screening of AAb responses in plasma samples from lung adenocarcinoma patients and controls with matched smoking history to identify disease associated AAb candidates.

b. Verify the responses of these candidate AAbs in larger sample set using clinically relevant ELISA assay.

c. Compare the responses of these candidate AAbs between cancer patients and patients with CT positive benign pulmonary nodules using clinically relevant ELISA assay.

Based on the results of this dissertation project, all objectives have been successfully achieved. The scientific contributions of the research are the following:

i) A method was developed to increase the signal to noise ratio of profiling complex plasma samples on NAPPA array. Specifically, *E. coli* lysate was used to pre-block plasma samples to reduce background binding of circulating antibody. And Hela cell lysate in vitro protein expression system were applied to display higher level of proteins on array, increasing the signal of antibody detection. In addition, a method to display denatured proteins on array was developed, allowing the profiling of antibodies against hidden linear epitopes.

ii) In the study of AAb responses in basal-like breast cancer, AAbs to 26 proteins were initially identified to associate with basal-like breast cancer using NAPPA displaying over 10,000 human proteins. 13 of them were confirmed in a larger sample set with a combined sensitivity of 33% at 98% specificity.

AAb to MN1 and TP53 proteins were also found to relate to worse prognosis. This panel has potential to improve the detection of BLBC after further validation in larger independent cohorts.

iii) The profiling of AAb responses in 40 patients with lung adenocarcinoma and 40 controls revealed tumor-associated antigens involved in embryonic morphogenesis process and organ development. Further evaluation of 19 antigens by ELISA confirmed two panels of AAbs that can differentiate lung cancer patients from smoker controls as well as CT positive benign controls respectively.

CHAPTER 2

DEVELOPMENT OF A VERSATILE PROTEIN MICROARRAY ENABLING ANTIBODY PROFILING AGAINST DENATURED PROTEIN

2.1 Abstract

Purpose: We aim to develop a protein microarray platform capable of presenting both natural and denatured forms of proteins for antibody biomarker discovery. We will further optimize plasma screening protocols to improve detection.

Experimental design: We developed a new covalent capture protein microarray chemistry using HaloTag fusion proteins and ligand. To enhance protein yield, we used HeLa cell lysate as an *in vitro* transcription translation system (IVTT). *E. coli* lysates were added to the plasma blocking buffer to reduce non-specific background. These protein microarrays were probed with plasma samples and autoantibody responses were quantified and compared with or without denaturing buffer treatment.

Results: We demonstrated that protein microarrays using the covalent attachment chemistry endured denaturing conditions. Blocking with *E. coli* lysates greatly reduced the background signals and expression with IVTT based on HeLa cell lysates significantly improved the antibody signals on protein microarrays probed with plasma samples. Plasma samples probed on denatured protein arrays produced autoantibody profiles distinct from those probed on natively displayed proteins.

Conclusions and clinical relevance: This versatile protein microarray platform allows the display of both natural and denatured proteins, offers a new dimension to search for disease-specific antibodies, broadens the repertoire of potential biomarkers, and will potentially yield clinical diagnostics with greater performance.

2.2 Introduction

Antibodies are the products of the human adaptive immune response, which usually targets antigens from foreign pathogens in order to neutralize or destroy them. Disease-specific antibodies have great clinical utilities. The diagnosis and clinical evaluation of numerous infectious and autoimmune diseases rely on profiling serum antibodies against specific antigens. One special class of disease-specific antibodies is called autoantibody because it targets self-proteins. In autoimmune diseases, AAbs play critical roles during disease pathogenesis. The discovery of AAbs in cancers against tumor associated antigens has generated great interest because of their potential use as early detection biomarkers (Piura and Piura, 2011). The detailed mechanisms of how certain proteins become TAAs and trigger the immune response are still unclear, but the development of AAbs may relate to tumor antigen overexpression, mutation, or altered post-translational modification which will be further discussed in chapter 3 and 4. As AAbs are the responses of our body to the aberrant nature of cancer, they may serve as indicators of tumor initiation, treatment response, or disease prognosis (Piura and Piura, 2011). Circulating AAbs are easy to access, very stable, and highly specific, making them particularly useful as potential serum cancer biomarkers.

The human immune system generates antibodies against both conformational and linear epitopes. Conformational epitopes are recognized by antibodies when the epitope domain is properly folded and may include amino acids from distant parts of the linear polypeptide brought together by secondary and tertiary folding. In contrast, linear epitopes comprise a single linear peptide, usually 7 amino acids long (Larman et al., 2011). Linear epitopes may sometimes be buried inside a folded protein, preventing them

from detection by antibodies. AAbs have been linked to both conformational and linear epitopes (Kim et al., 1994; Schwartz et al., 1999).

Protein microarrays displaying full-length human proteins enable screening humoral immune responses against thousands of antigens in parallel to search for disease-specific AAb biomarkers (Kijanka and Murphy, 2009; Ramachandran et al., 2008c), including cancer (Anderson et al., 2011a; Nam et al., 2003; Qiu et al., 2004) and autoimmune diseases (Hueber et al., 2002). Conventional protein microarrays present proteins produced in and purified from bacteria, insect cells or yeast, which are then displayed on glass slides through various attachment chemistries including: amine reactive chemistry, anti-tag antibodies; and hydrophobic interactions (nitrocellulose coated slides) (Hurst et al., 2009; MacBeath and Schreiber, 2000; Ramachandran et al., 2004; Ramachandran et al., 2008b; Zhu et al., 2001). Low signals and high backgrounds have plagued detection sensitivity and specificity on protein microarrays probed with serum. This is particularly true in the case of AAbs where the responses to self-proteins are often weaker than those to foreign invaders. The backgrounds may be associated with the methods for protein production and/or the printing conditions.

Signals of AAb responses are usually related to the amount of immobilized proteins and the mechanism of epitope presentation. In this study, we aim to improve the detection of AAbs on microarrays by reducing background and improving specific signals. Protein microarrays generally display folded or semi-folded proteins, which could obscure some linear epitopes buried within the proteins. These protein microarrays may fail to identify AAbs that recognize such epitopes. There is a need to develop a platform that displays antigens in both native and denatured states.

To this end, we developed a versatile protein microarray platform capable of presenting proteins for antibody and AAb screening in both native and denatured states, as needed, that is based on our Nucleic Acid Programmable Protein Array (NAPPA) platform. NAPPA in its standard format expresses epitope-tagged proteins just-in-time from printed cDNA and captures them to the surface with an anti-tag antibody. Proteins on NAPPA are expected to be properly folded by virtue of the chaperone proteins used during expression and have been demonstrated to display appropriate protein-protein interactions and enzymatic activities (Ramachandran et al., 2004; Ramachandran et al., 2008b).

2.3 Materials and methods

2.3.1 Detection of plasmid DNA and expressed proteins on HaloTag protein array

Printed plasmid DNA was detected by staining the arrays with PicoGreen. Protein display was detected essentially as described using rabbit anti-HaloTag antibody (Promega) followed by Alexa Fluor 647 labeled goat anti-rabbit IgG (Invitrogen) (Ramachandran et al., 2008b; Ramachandran et al., 2008c).

2.3.1 Blocking of plasma samples with *E.coli* lysates

To make *E. coli* lysates, overnight DH5α culture in LB media was pelleted at 5000g for 15min, and re-suspended in PBS with 0.2% Tween 20. Sonication was performed at 20 kHz, power intensity level 4 for 600 cycles of 1 sec on and 1 sec pulse. The lysates were then boiled at 99˚C for 5min and centrifuged for 10min at 13000g. The supernatant was collected as the blocking buffer.

For plasma blocking experiments, plasma samples were blocked with 5% milk in *E.coli* lysate blocking buffer for 2.5 h before applied on to the expressed protein arrays.

2.3.3 Protein microarray denaturation

Expressed slides were treated with the denaturing buffer (125mM Tris-HCl, 2% SDS, 100mM β-mercaptoethanol, pH 6.8) at 37˚C for 30min with gentle rocking. Control slides were incubated with 125mM Tris-HCl, pH 6.8.

2.4 Results and Discussion

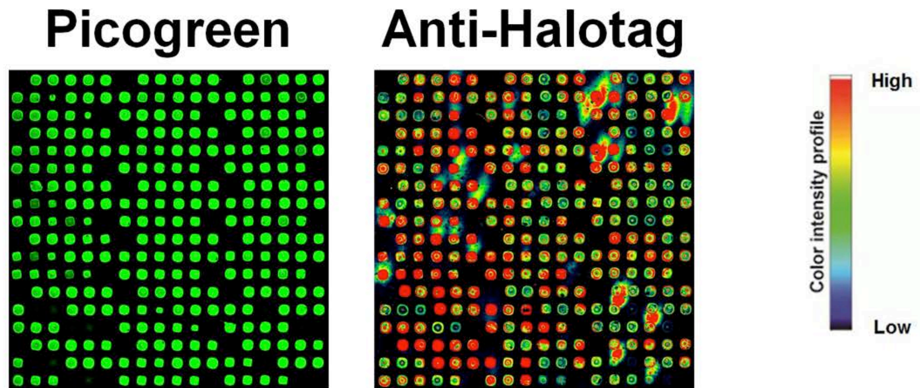
2.4.1 Develop of protein array platform to display denatured proteins

Standard NAPPA is a useful platform for displaying conformational epitopes; however, it was not clear if it could withstand denaturation. Indeed, after treatment of a standard NAPPA array with 125 mM Tris-Cl, 2% SDS, 100mM β-mercaptoethanol at 37℃ for 30 min with mild agitation, we were unable to detect the protein using a protein specific antibody that recognizes a linear epitope, presumably because the protein was released from the slide surface (Figure 2-1B).

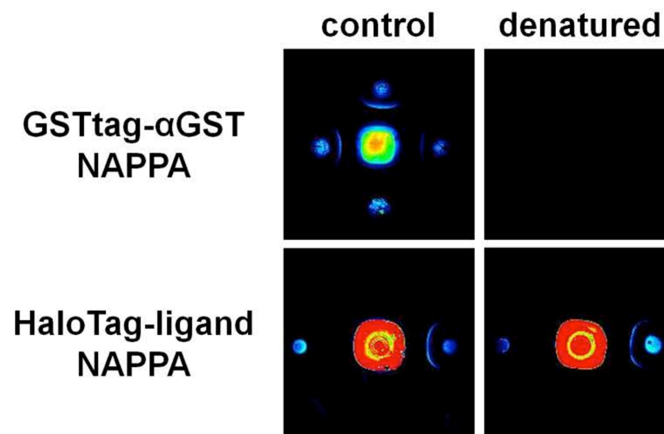
To solve this problem, we switched the immobilization chemistry from affinity capture to the covalent linkage between HaloTag and its ligand. HaloTag is a modified haloalkane dehalogenase developed to covalently bind to halogenated alkanes (Hurst et al., 2009). We hypothesized that protein microarrays with covalent immobilization chemistry would withstand harsh denaturation treatment without losing proteins from the slide surface.

To test this hypothesis, we designed and constructed pJFT7-nHalo vector that bears a Gateway™ death cassette next to an N-terminal fusion HaloTag protein (Figure 2-2). This vector has a T7 promoter and an encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES) sequence upstream of the gene of interest. We then transferred 69 genes from our DNASU repository from donor vector (pDONR221) to pJFT7-nHalo through Gateway™ LR reactions. DNA was miniprepped using a customized protocol as described (Ramachandran et al., 2008b).

A



B



Antibody: α-P53 mAb

Figure 2-1. HaloTag protein arrays can withstand harsh denaturing conditions. A. Picogreen staining of printed DNA (Left) and anti-HaloTag antibody detection of HaloTagged protein display (Right). The color scale for all protein array images is shown on right. B. Detection of p53 using an anti-p53 antibody on protein arrays where proteins were immobilized by GST/ anti-GST or HaloTag / HaloTag ligand with or without denaturation.

To manufacture protein microarrays, for each gene, the printing mix of BSA, HaloTag ligand, BS3 and DNA plasmid was prepared at concentrations of 3.7 mg/ml, 0.5 mM, 5 mM and ~1000 ng/µl, respectively, at a final volume of 30 µl. DNA samples were incubated overnight at 4°C before arrayed onto aminosilane coated slides by Genetix QArray2 with 350 µm solid pins. Each protein was fused to the n-terminal HaloTag. Just-in-time, *in vitro* expressed proteins were immobilized *in situ* to co-spotted HaloTag ligand that was linked to the slide surface using the amine chemistry. Picogreen staining of printed DNA and anti-HaloTag antibody detection of expressed proteins were carried out subsequently for quality control (Figure 2-1A).

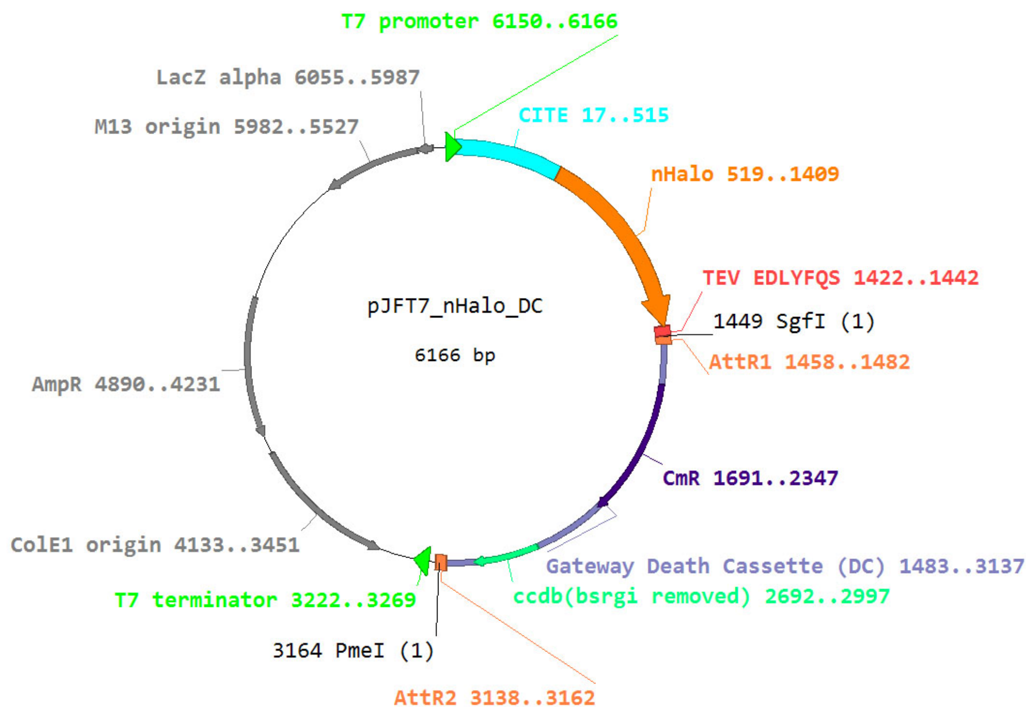


Figure 2-2. Vector Map of pJFT7-nHalo

In order to determine whether the protein microarrays displaying protein captured via covalent linkage would survive harsh washing conditions, we treated the slides with denaturing conditions and detected p53 using DO-1 monoclonal anti-p53 antibody, which

recognizes a normally exposed linear epitope QETFSDLWKL on p53 (Stephen et al., 1995). As a control, we did the exact same treatment on standard NAPPA which uses anti-GST antibody as the capture reagent. Similar p53 levels were detected in both denatured and non-denatured HaloTag-based protein microarrays, whereas p53 signal disappeared after denaturation on GST-tag NAPPA (Figure 2-1B).

2.4.2 E.coli lysate pre-block plasma reduces non-specific binding

During serum screening experiments, approximately 10% of samples show generalized background signals on all features with intensities that correlate with the serum sample (data not shown). Such background signals are particularly evident on features with DNA, and less on features with printing buffer only, and they do not correlate with the amount of displayed protein. To address this issue, we tested the possibility of using *E. coli* lysates as a blocking reagent for plasma, because it has been reported to reduce background signal (Vigil et al., 2011) (Figure 2-3A). We selected a plasma sample known to have high background and probed two identical NAPPA slides displaying around 2000 full-length human proteins as well as a domain of the Epstein-Barr virus nuclear antigen (EBNA) as a positive control. (In our experience, about 90% of individuals have sero-reactivity to EBNA (Ramachandran et al., 2008a). The proteins on the slides were produced by *in vitro* transcription and translation (IVTT) using rabbit reticulocyte lysate as a source of ribosomes. One of them was probed with plasma pre-blocked with *E. coli* lysate, the other one without blocking served as control. In the right panel of Figure 2-3A, the EBNA signal is shown in red. Although the "true signal" represented by the EBNA response was unchanged, the background signal on the other proteins was significantly reduced when blocked with *E. coli* lysate. We speculated that

trace amounts of bacterial proteins carried over through the DNA miniprep might contribute to this background in individuals who had immune responses to these *E. coli* proteins.

2.4.3 Evaluate Hela IVTT system in serum antibody detection

A common system used for the coupled transcription and translation of proteins has been rabbit reticulocyte-based IVTT mix (Promega), which has the advantage of using mammalian ribosomes and chaperones for protein production. This system is highly efficient, and has succeeded at producing thousands of different proteins, leading to the identification of disease-specific antibodies in cancers and infectious disease (Anderson et al., 2011a; Montor et al., 2009; Wright et al., 2012). A new mammalian cell free expression system derived from cultured human HeLa cells (HeLa IVTT; Thermo Scientific) was made available recently. We found that *E. coli* lysate blocking also worked with HeLa cell lysates (Figure 2-4). We also compared its performance with rabbit IVTT for plasma autoantibody detection after blocking plasma with *E. coli* lysate. We probed the same plasma sample onto two identical protein microarrays expressed with either rabbit IVTT or HeLa IVTT. Arrays expressed with HeLa IVTT gave much higher EBNA response (Figure 2-3B). The signal to noise ratio (EBNA signal intensity: median signal intensity of all spots on one slide) was calculated for each system and compared. As indicated in Figure 2-3B, we achieved a 7 fold increase of signal to noise ratio in the HeLa IVTT. Furthermore, some AAbs showed reactivity only on the HeLa IVTT expressed slide. Signal intensities for proteins IFT81, MKL1, MRPL28, SOX17 (red dots) were greater than the background in the HeLa IVTT expressed slide, but they were hidden in the background on the rabbit IVTT expressed slide (Figure 2-3C).

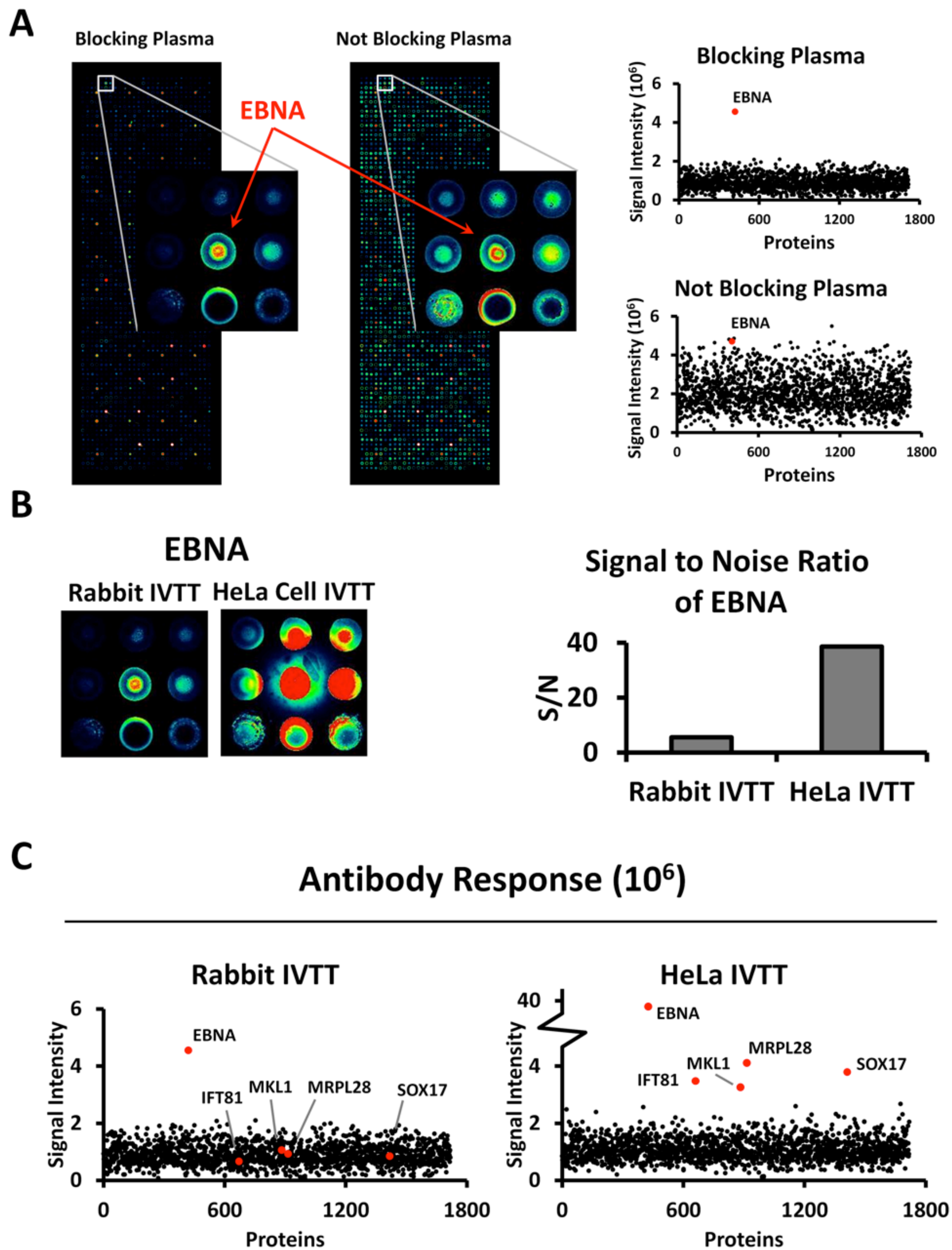


Figure 2-3. Optimization of the plasma screening protocol by blocking plasma samples with E.coli lysate and using HeLa IVTT expression system. **A.** Blocking plasma with E.coli lysate before probing onto arrays could reduce background signals. Two identical arrays were probed with the same plasma sample. Left, plasma pre-blocked by *E.coli*

lysate; right, no pre-blocking. Plots of antibody response signal intensity of all protein spots are shown on the right: top, plasma pre-blocked by E.coli lysates; bottom, no pre-blocking. **B.** HeLa IVTT improved the detection of antibody response. Zoomed-in images of antibody response to EBNA on NAPPA arrays expressed by either Rabbit IVTT or HeLa IVTT are shown on the left. Experiment condition was the same as that in Figure 2A, with plasma pre-blocked by E.coli lysate. Signal to noise ratio plot of EBNA antibody responses is shown on the right. **C.** Plots of antibody response signal intensity of all protein spots. Increased autoantibody responses against several human proteins when using HeLa IVTT expressed NAPPA arrays were detected (shown in red dots).

We then applied the optimized AAb screening protocol to evaluate the effect of denaturation on AAb detection. AAb profiles for the same plasma sample on denatured and non-denatured HaloTag NAPPA arrays were compared. For each condition, an additional "mock-expressed" slide, which used HeLa IVTT lacking T7 polymerase (thus incapable of expressing proteins), was used as a negative control to accurately measure background signals unrelated to protein expression. AAb responses were calculated by subtracting spot intensities on slides without protein expression from those on slides with expression. Spots with calculated signals less than 0 were set to 0. We detected AAb response changes against several proteins when comparing denatured and non-denatured slides. The AAb response against TPD52 (red dots) was lower in the denatured slide (Figure 2-5A; suggesting a conformational epitope), whereas AAb responses against MYC and CCNA1 (green dots) were higher (suggesting linear epitopes). Images of antibody responses to CCNA1 on both denatured and non-denatured HaloTag protein microarrays are shown in Figure 2-5B as an example (AAb response to TPD52 and MYC are shown in the Figure 2-6). It is obvious that AAb responses to CCNA1 were much stronger when the protein was denatured. Denaturation of proteins may expose hidden linear epitopes as well as unfold conformational epitopes that only exist in the non-

denatured form. TPD52 may lose its epitope after denaturation, whereas MYC and CCNA1 may gain their epitopes through denaturation. The adaptation of HaloTag to protein microarrays not only maintains its utility for studying of proteins where conformational presentation is needed, but also, more importantly, enables applications where denatured proteins are preferred. Combining the information from both natural and denatured protein microarrays may increase both the diagnostic sensitivity and the number of potential biomarkers, as more responses can be discovered.

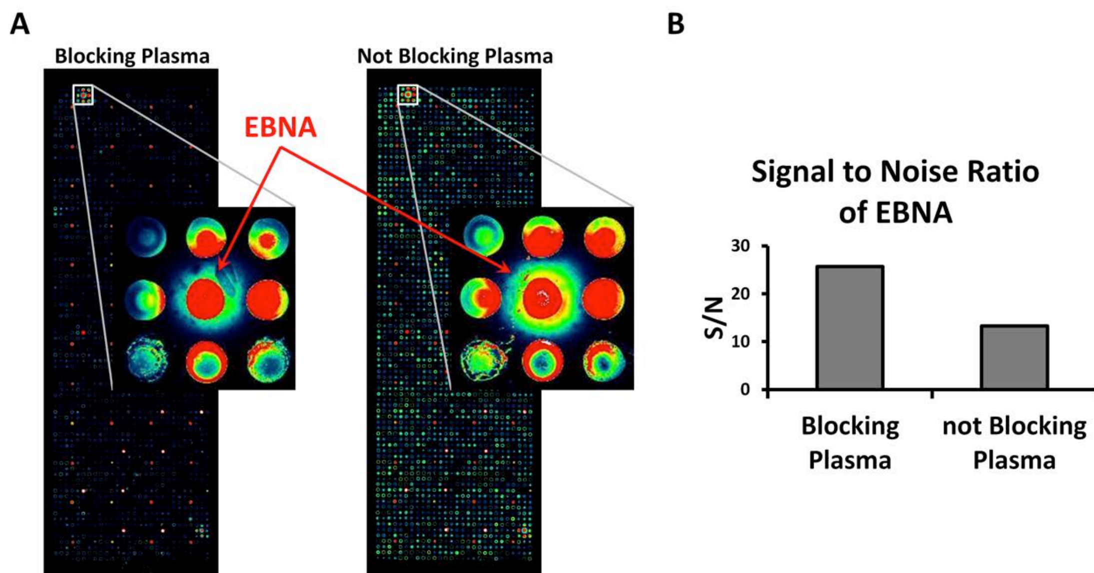


Figure 2-4. Blocking plasma with E. coli lysate can reduce background signal on HeLa IVTT expressed protein array. **A.** Left, plasma pre-blocked by *E.coli* lysate; right, no pre-blocking. **B.** Signal to noise ratio plot of two conditions in A. Background was represented by median signal intensity of all spots on one slide. Feature color represents an artificial color scheme to indicate signal intensity (red > yellow > green > blue).

2.5 Conclusion

In conclusion, we have developed a protein microarray platform where proteins are covalently linked to the matrix. We demonstrated that protein microarrays using the HaloTag and its ligand immobilization chemistry could survive denaturing buffer condition and produce distinct AAb response profiles from the non-denatured protein microarrays when probed with plasma samples. We recognize that more work needs to be done to analyze the differences of AAb profiles systematically between denatured and native protein microarrays using a larger sample size and determine its utility in identifying AAb biomarkers with better clinical performance. We believe that this method can easily extend beyond AAbs detection in cancers to automimmune and infectious diseases. Moreover, covalent attachment of proteins on the matrix for denaturation can also be adapted to other types of protein microarrays.

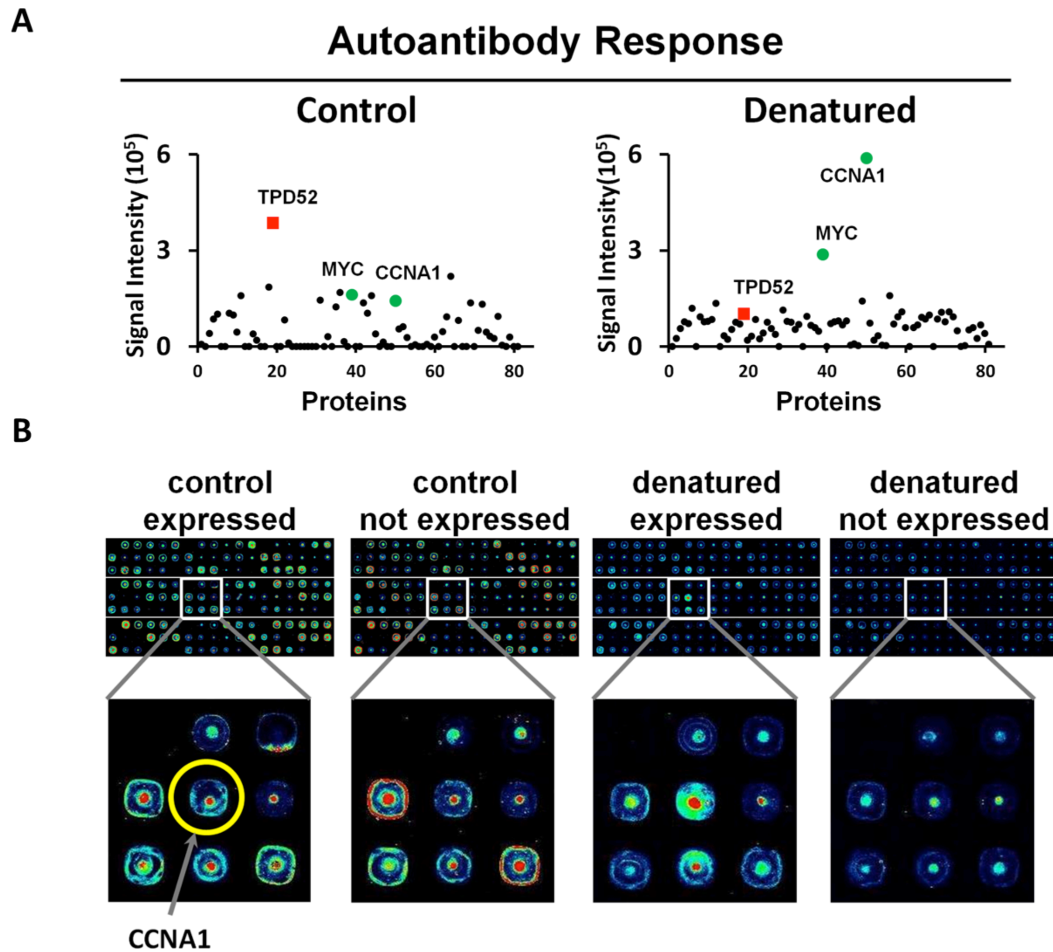


Figure 2-5. NAPPA using covalent capture chemistry enables the detection of AAb against denatured proteins. **A.** AAb response profile against proteins on denatured NAPPA was different from that on non-denatured NAPPA. Red squares, AAb response decreased after denaturation. Green dots, AAb response increased after denaturation. **B.** An example of different AAb response between the native protein array and the denatured protein array. NAPPA requires transcription of cDNA by T7 polymerase into mRNA before translation. T7 polymerase was not added to arrays labeled with "not expressed" to assess non-specific background signals.

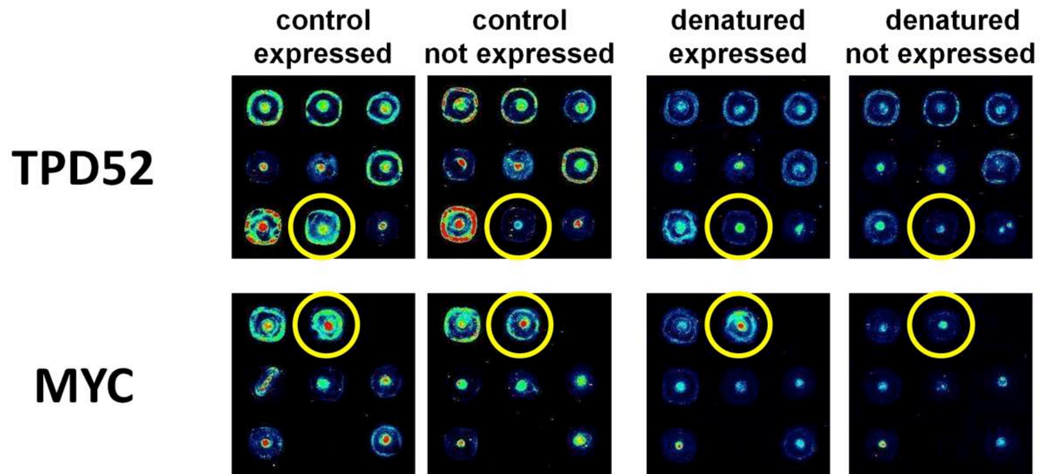


Figure 2-6. Different AAb response to TPD52 and MYC between native protein array and denatured protein array.

CHAPTER 3

PLASMA AUTOANTIBODIES ASSOCIATED WITH BASAL-LIKE BREAST CANCERS

3.1 Abstract

Background: Basal-like breast cancer (BLBC) is a rare aggressive subtype of breast cancer that is less likely to be detected through mammographic screening. Identification of circulating markers associated with BLBC could have promise in detection and management of this deadly disease.

Methods: Using samples from the Polish Breast Cancer study, a high-quality population-based case-control study of breast cancer, we screened 10,000 antigens on protein arrays using 45 BLBC patients and 45 controls, and identified 748 promising plasma autoantibodies (AAbs) associated with BLBC. ELISA assays of promising markers were performed on a total of 145 BLBC cases and 145 age-matched controls. Sensitivities at 98% specificity were calculated and a BLBC classifier was constructed.

Results: We identified a 13-AAbs (CTAG1B, CTAG2, TP53, RNF216, PPHLN1, PIP4K2C, ZBTB16, TAS2R8, WBP2NL, DOK2, PSRC1, MN1, TRIM21) that distinguished BLBC from controls with 33% sensitivity and 98% specificity. We also discovered a strong association of TP53 AAb with its protein expression (p=0.009) in BLBC patients. In addition, MN1 and TP53 AAbs were associated with worse survival (MN1 AAb marker HR=2.25 95%CI= 1.03-4.91 p=0.04; TP53, HR=2.02, 95%CI 1.06-3.85, p=0.03). We did not find that AAb levels differed significantly by demographic characteristics.

Conclusions: These AAbs warrant further investigation in clinical studies to determine their value for further understanding the biology of BLBC and possible detection.

3.2 Introduction

Despite significant improvements in early detection by routine mammography, breast cancer remains a global health challenge (Jemal et al., 2011). Current screening mammography detects only 70% of breast cancers (Esserman et al., 2009). Breast tumors associated with high breast density and highly proliferative cancers are frequently not detected by screening (Esserman et al., 2007; Tamimi et al., 2007). Therefore, there is an urgent need for biomarkers that can detect potentially invasive breast cancer in their early stages.

It is well established that breast cancer is heterogeneous, comprising multiple molecular subtypes with unique clinical and molecular characteristics that impact prognosis, response to treatment and risk of recurrence (Sorlie et al., 2001; Sorlie et al., 2003). This heterogeneity affects biomarker discovery, requiring both larger sample sizes and different statistical approaches from traditional methods of evaluation (Wallstrom et al., 2013). We have previously shown that patients with HER2 negative breast cancer developed AAbs against survivin, but not patients with HER2-enriched breast cancer (Anderson et al., 2008). Unique plasma protein signatures have also been reported in other breast cancer subtypes (Gonzalez et al., 2011; Li et al., 2012). Notably, as the sensitivity of a subtype-specific biomarker can never exceed the prevalence of that subtype in the population where it is tested, biomarkers might reveal very different performance as the proportions of different subtypes vary in populations. For disease subtypes that are rare in the overall population, it is essential to limit the sample heterogeneity to discover biomarkers with significant sensitivity for clinical use.

Basal-like tumor is a breast cancer subtype associated with African American ethnicity, *BRCA1/2* mutation and younger age (Rakha et al., 2008). It has about 80% overlap with triple negative breast cancer (TNBC) (Foulkes et al., 2010), a clinical pathological subtype characterized by negative tissue staining of estrogen receptor (ER), progesterone receptor (PR), and the absence of human epidermal growth factor receptor 2 (HER2) amplification (Foulkes et al., 2010; Reis-Filho and Tutt, 2008). The basal subtype includes expression of epidermal growth factor receptor (EGFR) or basal cytokeratin 5/6 (CK5/6) and has worse prognosis than other patients with TNBC (Cheang et al., 2008). This five marker subtyping (ER-, PR-, HER2-, and either EGFR+ or CK5/6+) is highly correlated with basal-like disease defined by gene expression profiling (Cheang et al., 2008; Metzger et al., 2012).

Current screening mammography has limitations in detecting basal-like tumors, which is often diagnosed at advanced stage with poor prognosis (Dent et al., 2007). Basal-like breast cancer is under-diagnosed by routine mammography and more frequently detected as interval breast cancer (Collett et al., 2005; Sihto et al., 2008). Detection is compromised partly by the high proliferation rate of basal-like subtype and lack of common suspicious radiological features (Dogan et al., 2010; Dogan and Turnbull, 2012; Yang et al., 2008). Dent et al. reported that patients with TNBC are less likely to be first detected by mammography or ultrasound than patients with other breast cancers (19.6% versus 36%), and they are more likely to have grade III tumors at time of diagnosis (Dent et al., 2007). Moreover, basal-like breast cancer often occurs in women less than 50 years old, who are not recommended for routine mammography (Calonge et

al., 2009). Thus, a relatively young population with basal-like tumors might benefit from a molecular test for the disease.

In practical terms, molecular diagnostic tests for cancer detection in apparently healthy individuals are best performed on accessible samples, like plasma. However, the biomarker concentration is often low because of limited secretion by a small number of cancer cells in the pre-clinical stage, of which only a fraction is distributed into the large plasma volume (Anderson and Hunter, 2006; Lutz et al., 2008). One strategy to amplify signal is to exploit the ability of the immune system to detect the presence of tumor cells through the generation of AAbs to either the cancer cells or products they release (Anderson and LaBaer, 2005; Anderson et al., 2011a; Anderson et al., 2010; Goodell et al., 2006; Hanash et al., 2011; Mange et al., 2012; Wang et al., 2005; Xu et al., 2014). AAbs are easier to detect and have been observed years before the clinical diagnosis of tumor (Lu et al., 2012; Qiu et al., 2008). In addition to early detection, AAbs in the blood may also help distinguish different disease subtypes or predict clinical course.

We have previously identified specific antibodies in diseases (Anderson et al., 2011b; Anderson et al., 2010; Miersch et al., 2013; Montor et al., 2009; Wright et al., 2012) by using nucleic acid programmable protein arrays (NAPPA), which display thousands of freshly-produced full-length human proteins on glass slides without the need of protein purification (Ramachandran et al., 2004; Ramachandran et al., 2008b). In particular, we previously discovered 28 AAb biomarkers for breast cancer (Anderson et al., 2011a) with sensitivities in the 10-30% range and specificities from 80-100% in a blinded study. However, that sample cohort was a mixed population of predominantly women with ER+/PR+ breast cancer potentially limiting the use of those markers in

subtypes like basal-like or HER2-enriched breast cancer. Here, we aimed to identify basal-like specific AAbs by profiling humoral immune responses of basal-like patients against 10,000 human proteins.

3.3 Materials and methods

3.3.1 Study samples

Subjects were selected from a population-based breast cancer case-control study of 2386 cases and 2502 age and study site matched controls, between ages 20 and 74 years who resided in Warsaw or Łódź, Poland from 2000-2003 (Garcia-Closas et al., 2006). Pathology for all the study cases was reviewed centrally as previously described to provide standardized classification. Basal-like subtype was defined by PAM50 signature when mRNA expression profiles are available (n=18); the rest (n=127) were identified by immunohistochemical (IHC) staining for the five markers (ER, PR, HER2, CK5/6, EGFR) as previously described (Sherman et al., 2007; Yang et al., 2007). Luminal A, Luminal B, and HER2-enriched subtypes were classified using PAM50 signature. We identified 145 cases with tumor tissues and plasma samples available (Sun et al., 2014). Similar to previous reports (Nielsen et al., 2010), we observed an 80% concordance rate between the five-IHC marker panel and mRNA expression profiles. Each case was individually matched on age (5 years) and study site with population based controls. To determine the specificity of AAbs identified for BLBC, we selected an additional set of age-matched non-basal cases (age matched to sample sets 2 and 3) classified by mRNA expression profiles (30 Lum A, 22 LumB and 18 HER2). All subjects provided informed consent and the study was approved by IRB boards in Poland and NCI.

3.3.2 Protein array experiments

Open reading frames were obtained from DNASU (https://dnasu.org/). Production of the protein array and array quality control experiments were performed as previously described (Festa et al., 2013; Wang et al., 2013). In brief, arrays displaying 10,000

human proteins (distributed evenly on five array sets) were manufactured. A common control plasma sample was repeated in every experiment to assess reproducibility. Consistency among experiments was determined with scatter plots comparing spot intensity measurements of the same plasma sample tested on different experiments. Specifically, Slides were incubated with SuperBlock (Pierce) at room temperature for 1 h with rocking. To express proteins, 160 µl 1-Step Human Coupled *in vitro* Expression system (Thermo) was injected into HybriWell (Grace BIO-LABS) sealed slides and incubated in the incubator (EchoTherm) for 1.5 h at 30°C for protein expression and 0.5 h at 15°C for protein capture. Expressed slides were rinsed with de-ionized water and dried by centrifugation at 4°C, 1200 rpm for 3 min. Plasma antibody profiling was then performed on HS 4800™ Pro hybridization station (Tecan). Plasma samples were diluted at 1:50 in 5% milk prepared with 100% E.coli lysate, and incubated for 3 h at room temperature before hybridized with protein arrays (Wang et al., 2013). Slides with displayed proteins were placed in the hybridization chamber of Tecan HsPro and programed with 1 h blocking with 5% milk-PBST (0.2% Tween), 16 h of plasma sample hybridization at 4°C followed by 3 times wash with 5% milk-PBST (0.2% Tween). Then slides were incubated with Dylight649 labeled goat anti-human IgG (Jackson ImmunoResearch Laboratories) at 23°C for 1 h. Slides were then washed, dried and scanned by Tecan scanner under consistent settings. To minimize technical variation from array production and sample handling, we grouped BLBC cases and matched controls, and probed the two samples from each pair side by side on consecutively printed protein arrays (blinded to case-control status as needed by the experimental design).

3.3.3 Protein array image analysis and quantification

We measured spot intensity of the scanned slides using ArrayPro Analyzer (MediaCybernetics). Raw intensity values were normalized by subtracting the background signal for the slide, which was estimated by the first quartile of signal intensity in spots with no printed DNA, and dividing by the median of background-subtracted intensity from non-control spots. In addition, to capture diffused signal (ring) that cannot be quantified by the image analysis software, one researcher qualitatively examined all images to identify and confirm positive responses, which was described previously (Figure 3-2) (Montor et al., 2009). Briefly, the researcher adjusted raw images to extreme contrast and brightness using ArrayPro Analyzer, and graded each spot at a scale of 0 to 5 based on ring's intensity and morphology.

3.3.4 Generating AAb target protein sets for enrichment analysis

To identify positive AAb responses for enrichment analysis using the quantitative data, we developed cutoff calculated by mean plus 3 times standard deviation of all spots in sample set 1. Spots with higher intensity than the cutoff were recorded as AAb positive (Table S2). Spots were excluded from analysis when they were affected by diffused signal from neighboring EBNA spots.

To perform GSEA and Gene Ontology (GO) term analysis, we used data from both quantitative and visual analysis. For quantitative analysis, we selected proteins with positive AAbs observed in at least 2 samples regardless of disease status, resulting in 253 proteins; for visual analysis, we selected proteins with positive AAbs observed in at least 3 samples regardless of disease status, resulting in 871 proteins.

3.3.5 Gene set enrichment analysis (GSEA)

GSEA was performed using GSEA Preranked package (Subramanian et al., 2005). Antigenicity scores of each individual peptide window with default size were calculated using B cell epitope prediction tools from Immune Epitope Database And Analysis Resource (IEDB; http://tools.immuneepitope.org/main/html/bcell_tools.html). Specifically, for each protein in each method, cutoff of 1.0 was used to assign whether a peptide window is antigenic. Then, we calculated the ratio between number of antigenic peptide windows and total number of peptide windows. This ratio was used as a score to quantify predicted overall antigenicity of each protein. Other protein properties, such as isoelectric point, protein length, aromaticity, fraction of helice, fraction of sheets, and fraction of turns were obtained using biopython tools (Bio.SeqUtils.ProtParam). GSEA was performed on AAb target protein sets against each of the above properties. Proteins were ranked in decreasing order of metric associated with each protein property. Normalized enrichment scores and *P* values were calculated based on 10,000 permutations. More details are provided in the Supplementary Materials. Parameters that differed from default settings were set as follows, and kept consistent throughout the analysis.

producer_class xtools.gsea.GseaPreranked

param collapse false

param scoring_scheme classic

param nperm 10000

param set_max 800

3.3.6 Gene Ontology cellular component analysis

We calculated numbers of shared gene symbols between proteins displayed on screening array (reference set) and each GO cellular component term. The same procedure was applied to both quantitatively and visually selected protein sets. Specifically, 6317 gene symbols from protein array, 178 quantitatively selected proteins, and 613 visually selected proteins were successfully mapped to GO terms of cellular component. GO terms represented by more than 100 genes on our protein array without apparent overlap were included for enrichment/depletion analysis and pie chart construction (Figure S8). *P* values were calculated by two-sided Fisher exact test, and adjusted multiple testing by Benjamini-Hochberg method.

3.3.7 Antigen selection for focused array

Using the normalized array data from the screening, we calculated sensitivity at 95% specificity based on data generated from printing batch 1 of each array set, area under receiver operating characteristic curve (AUC), partial AUC above 95% specificity (pAUC), as well as Welch's t test *P* value for each tested protein antigen. In addition, we designed a novel metric, named K, to measure antigens with strong antibody responses in a fraction of BLBC patients while remain consistent in controls. K is calculated using the formula below.

$$K = \frac{q_{cases}(0.975) - q_{cases}(0.800)}{q_{controls}(0.975) - q_{controls}(0.025)}$$

where q_{cases} and $q_{controls}$ denote the empirical quantile functions of normalized data from cases and controls, respectively. For antibodies with the same classification

performance, a high K value indicated greater separation of sero-reactivity of positive cases and negative controls.

We created focused protein arrays for stringent evaluation of antigens that met at least one of the following criteria (The corresponding metric values for each protein are listed in Table S4): 1). Antigens ranked in approximately top 2% of antigens on each array set based on any of these metrics: sensitivity at 95% specificity (n=228), AUC (n=185), pAUC (n=197), or *P* value of Welch's t test (n=197). 2). Antigens with K > 1.2 and sensitivity at least 15% at 95% specificity (n=63). 3). Antigens presented greater prevalence in cases than that in controls by visual analysis (n=198). Specifically, frequency in cases minus frequency in controls is greater than or equal to 2, and frequency in cases divided by frequency in controls is greater than or equal to 1.5. 4). Antigen presented in greater prevalence in controls than that in cases by visual analysis (n=16). Specifically, frequency in controls minus frequency in cases is greater than or equal to 5, and frequency in controls divided by frequency in cases is greater than 1.5. Accounted for proteins selected by multiple methods, totally 748 proteins were included for manufacturing focused array.

3.3.8 Power analysis for the biomarker discovery

We calculated the power for antigen selection using a homogeneous disease model and a heterogeneous disease model (Miersch et al., 2013). Using each model and Monte Carlo simulation, we calculated the proportion of markers with 20% sensitivity and 95% specificity that met criteria 1 or 3 in above section "Antigen selection for focused array". The visual inspection criterion was not considered in the power analysis. Under the homogeneous disease model, 95% of markers with 20% sensitivity and 95%

specificity met the selection criteria, and 5% of non-markers with 5% sensitivity at 95% specificity met the criteria. Under the heterogeneous model, 73% of markers and 6% of non-markers met the criteria. Hence, nearly all such markers with 20% sensitivity and 95% specificity would be selected by our screening process if basal-like subtype is homogeneous, and if it is itself heterogeneous, our process would still be expected to select 73% of these markers.

3.3.9 Antigen selection for ELISA verification

Protein antigens were selected for subsequent ELISA verification when they showed higher prevalence in basal-like breast cancer (BLBC) in sample set 1 based on visual analysis. Specifically, they had to meet all of the following criteria: 1). their frequency in BLBC minus frequency in controls is greater than or equal to 3; 2). frequency in cases divided by frequency in controls is greater than or equal to 2; and 3). frequency in cases is greater than or equal to 4. Totally, eighty-two unique proteins met these criteria and we successfully developed programmable ELISA assay for 71 of them. Two pairs of samples (PBCS-1243, PBCS-2930; PBCS-1754, PBCS-1325) were not measured in ELISA verification experiments due to limited amount of plasma at the time of experiment.

3.3.10 ELISA assays

ELISA assays were performed to verify selected AAb responses towards protein antigens using freshly produced human proteins as previously described (Ramachandran et al., 2008a). Specifically, 96-well highbind ELISA plates (Corning) were coated with goat anti-GST antibody (GE Healthcare) at 10 μg/ml in 0.2 M sodium bicarbonate buffer pH9.4 overnight at 4°C 1 day prior to experiment. All high-throughput liquid handling

were performed using a BioMek NxP Laboratory Automation Workstation (Beckman Coulter). On the next day, 40 μg/ml DNA plasmids encoding protein antigens were expressed using 1-Step Human Coupled *in vitro* Expression system (Thermo) at 30°C for 1.5 h. At the same time, coated plates were washed 3 times with PBST (0.2% Tween) and blocked with 5% milk PBST (0.2% Tween) at room temperature for approximately 2 h. Expressed antigens were then diluted at 1:50 and incubated in ELISA plates at room temperature (RT) with shaking at 500 rpm for 1 h, in order to capture expressed antigens. After the incubation, plates were washed 5 times with PBST (0.2% Tween). Plates were then incubated with 1:300 diluted plasma samples at RT with shaking at 500 rpm for 1 h, and washed 5 times again with PBST (0.2% Tween). After incubated with 1:8000 diluted HRP-conjugated goat anti-human IgG (Jackson ImmunoResearch Laboratories), plates were developed with 1-Step Ultra TMB - ELISA Substrate (Thermo scientific). Reaction was stopped with addition of 2 M sulfuric acid after 15 min. OD_{450} was measured by Envision Multilabel Plate Readers (PerkinElmer). Expression of antigens was confirmed by using anti-GST (Cell Signaling) and HRP-conjugated anti-mouse IgG (Jackson ImmunoResearch Laboratories). All measurements in the training and testing sets were performed in duplicates. ELISA relative absorbance of each plasma sample-antigen reaction was calculated using OD_{450} of expressed antigens over the median OD_{450} of all antigens measured for that sample. The median value was used to normalize systematic background of each plasma sample. Subjects were excluded from analysis when failed quality control. During ELISA training and blinded testing, one pair of samples (PBCS-1243, PBCS-2930) was not measured for TP53, CTAG1B, CTAG2, RNF216, SNRK, and PSRC1 AAb responses due to limited amount of plasma at the time of experiment.

3.3.11 Statistical analysis

Frequencies of tumor characteristics and demographics between cases were compared using chi-square tests. Associations with known breast cancer risk factors were determined using multivariable logistic regression models as previously described (Garcia-Closas et al., 2006).

ROC analysis was performed without feature selection (to avoid over fitting) using leave-one-out cross validation. The 13-AAb classifier was developed by classifying samples as positive if they exceeded antigen-specific cutoffs for at least 2 of the 13 antigens. Antigen-specific cutoffs were set to achieve 98% classifier specificity by adjusting the specificity at the antigen level to 98.7%. In this cross validation, for a given antigen-level specificity, we calculated the cutoffs for each antigen using the remaining samples and used these cutoffs to classify the held-out sample. The ROC curve was calculated by adjusting the antigen-level specificity. 95% confidence intervals of ROC curve and AUC were estimated by bootstrapping within BLBC or controls.

To determine AAb responders from ELISA analysis, we categorized subjects as responders to specific antigens of interest using 95-percentile cut-point using data from control subjects. This method was used to determine the association of AAb responses with tissue abundance of TP53 protein, as well as the overall survival.

To analyze the mRNA expression level of the candidate AAb targets, data set were generated by TCGA using Illumina HiSeq, and obtained from UC Santa Cruz Cancer Genome Browser (https://genome-cancer.ucsc.edu/) (TCGA_BRCA_exp_HiSeqV2-2013-12-18). All intensities were normalized by

subtracting mean of each mRNA from each sample. *P* values were calculated using ANOVA comparing basal-like subtype with other subtypes and normal tissue respectively, and were adjusted for multiple comparisons by Sidak correction.

The Kaplan-Meier (KM) method was used to generate survival curves for categories of the AAb responders/non-responders (Kaplan and Meier, 1958). HR and 95% confidence intervals (CI) associated with AAb markers adjusted for age, tumor size, grade, and node status, were estimated using Cox proportional hazard models (Cox, 1972). Survival analysis was performed using Stata/SE v11.2 for Windows (College Station, TX).

3.4 Results

3.4.1 Sample tumor characteristics and risk factors

Evaluation of established breast cancer risk factors for 145 basal-like cases and 145 healthy controls showed early age at menarche to be protective for breast cancer, positive family history of breast cancer, history of benign breast disease and ever having a screening mammogram associated with increased breast cancer risk (Table 3-1). These samples were randomly divided into three groups for the biomarker discovery and validation (Figure 3-1; Table 3-2). Samples included in the screen of 10,000 proteins (sample set 1) were more likely to be of higher grade and less likely to be node positive compared to sample set 2 and 3 (Table 3-2). Parity did not show a protective association that has been noted to be protective for ER-positive breast cancers but not ER-negative breast cancers(Yang et al., 2011b).

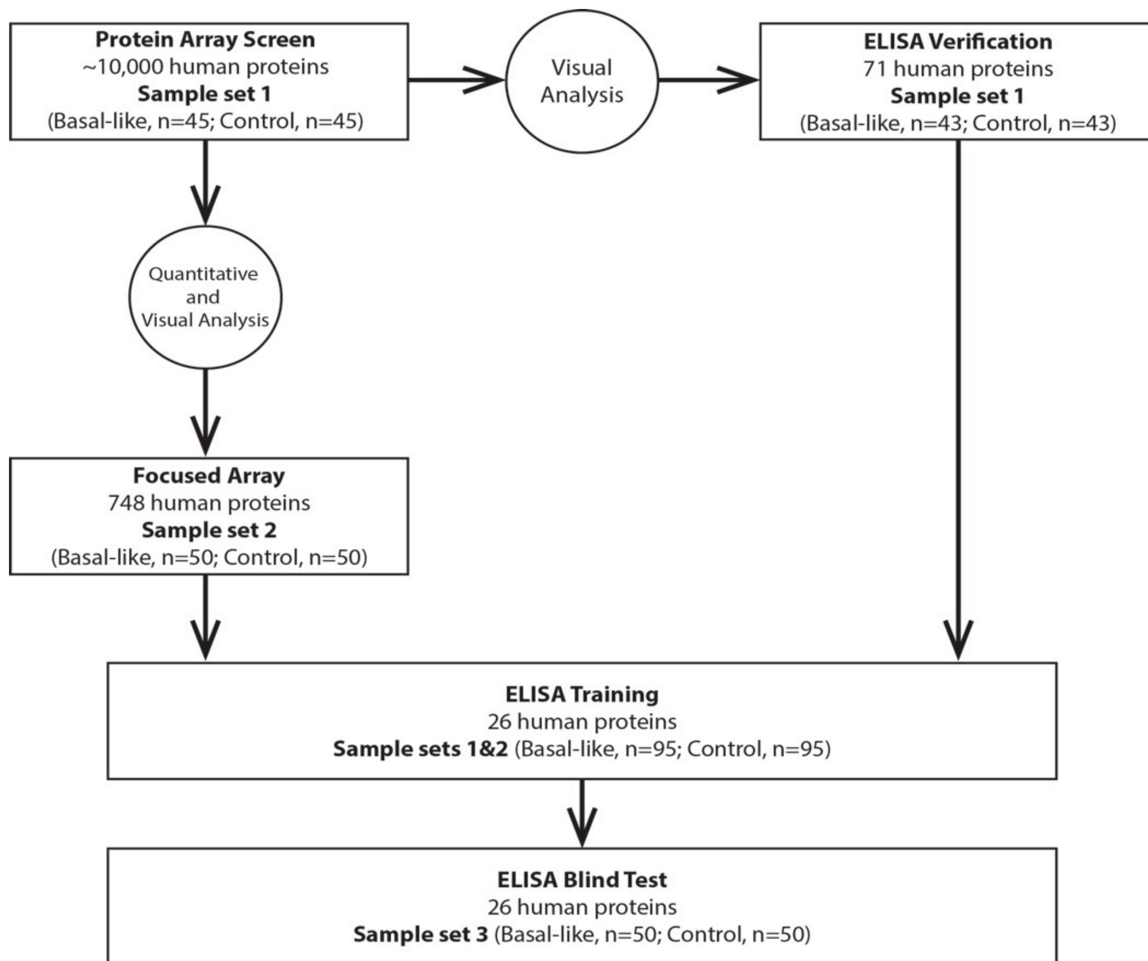


Figure 3-1. Overview of study design.

3.4.2 Characterizing the global autoantibody response in patients with basal-like breast cancer and healthy individuals

We performed a comprehensive profiling of plasma AAbs against 10,000 full-length human proteins in sample set 1, 45 patients with basal-like disease and 45 age-collection site- matched healthy individuals (Figure 3-1). Array production was quality-controlled to achieve consistent high level protein display (Figure 3-3B). The average Pearson correlation coefficient among all technical repeats across all experiments was 0.98 (SD=0.01) (Figure 3-3D; Figure 3-4). Duplicate spots on the same array, which were

situated away from each other to avoid local effects, revealed minimal within-slide variation, with an average for all QC experiments of 0.98 (SD=0.01) (Figure 3-3E).

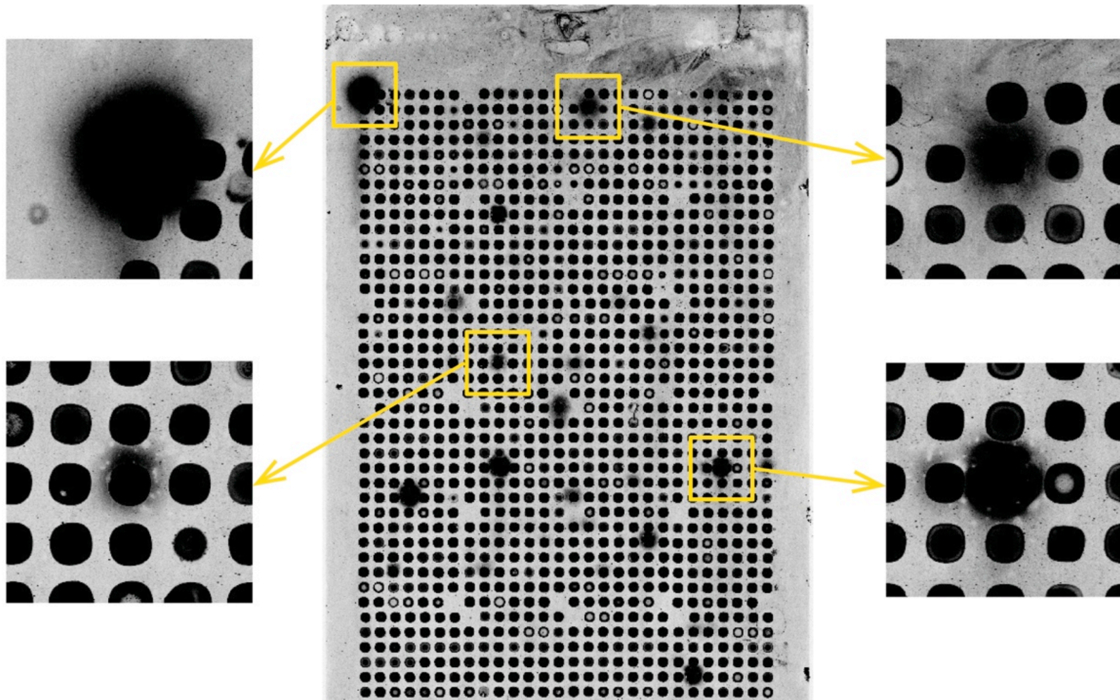


Figure 3-2. Representative array images for “ring” structure surrounding spots. “Ring” was considered as positive response during visual image analysis.

Table 3-1. Established breast cancer risk factors in the Polish Breast Cancer Study (basal-like, n=145; controls, n=145)[a].

	Basal-like		Controls						
	N	%	N	%	OR[b]	95% CI[b]		p	p-trend
Education level									
Less than high school	56	39%	49	34%	1				
High school	53	37%	58	40%	0.47	0.22	0.99	0.05	
Some tech training or college	11	8%	12	8%	0.32	0.1	1.08	0.07	
College degree	23	16%	26	18%	0.51	0.19	1.32	0.16	
Missing	2		0						
Age at menarche									
< 12	54	38%	28	20%	1				
13	31	22%	34	24%	0.33	0.15	0.75	0.01	
14	37	26%	43	30%	0.32	0.15	0.69	0.003	
15	14	10%	23	16%	0.11	0.04	0.33	0.0001	
≥ 16	7	5%	15	10%	0.16	0.05	0.53	0.003	0.0001
Missing	2		2						
No. of full-term births									
nulliparous	20	14%	37	26%	1				
1	46	32%	65	45%	1.89	0.46	7.82	0.38	
2	58	40%	33	23%	1.41	0.4	5	0.59	
≥ 3	21	14%	10	7%	1.13	0.29	4.42	0.86	0.81
Missing	0		0						
Age at first full-term birth among parous women									
< 20	37	26%	37	26%	1				
20-24	69	48%	65	45%	0.82	0.29	2.33	0.71	
25-29	27	19%	33	23%	0.73	0.22	2.43	0.61	
≥ 30	12	8%	10	7%	1.04	0.24	4.6	0.96	0.5
Missing	0		0						
Menopausal status									
Pre-menopausal	50	34%	65	45%	1				
Post-menopausal	95	66%	80	55%	9.54	1.96	46.46	0.01	
Missing	0		0						
Age at menopause among post-menopausal women									

< 45	64	45%	77	53%	1				
45-49	28	20%	24	17%	1.09	0.33	3.6	0.89	
≥ 50-54	51	36%	44	30%	1.78	0.57	5.59	0.32	0.06
Missing	2		0						
Hormone replacement therapy use among post-menopausal									
Never	70	82%	52	69%	1				
Current/recent use	7	8%	10	13%	0.17	0.05	0.64	0.01	
Past use	4	5%	5	7%	0.32	0.06	1.82	0.2	
Ever used E or P alone	4	5%	8	11%	0.06	0.01	0.33	0.001	
Missing	13		16		0.58	0.21	1.62	0.3	
Current BMI among pre-menopausal									
< 25	22	42%	29	43%	1				
25 - < 30	19	36%	23	34%	1.19	0.43	3.25	0.74	
≥ 30	12	23%	15	22%	0.9	0.28	2.97	0.87	0.94
Missing									
Current BMI among post-menopausal									
< 25	34	35%	25	30%	1				
25 - < 30	26	27%	31	37%	0.47	0.18	1.2	0.11	
≥ 30	36	38%	27	33%	0.73	0.28	1.95	0.54	0.2
Missing	0		0						
Family history of breast cancer in first-degree relatives									
No	121	83%	139	96%	1				
Yes	24	17%	6	4%	7.65	2.52	23.23	0.0003	
Missing	0		0						
History of benign breast disease									
No	118	83%	130	91%	1				
Yes	24	17%	13	9%	3.11	1.25	7.74	0.01	
Missing	3		2		1.99	0.21	19.02	0.55	
Ever had a screening mammogram									
No	69	49%	79	54%	1				
Yes	73	51%	66	46%	2	1.03	3.88	0.04	
Missing	3		0						

[a]145 cases and 145 age-site matched controls in models after excluding missing.

[b]Odds ratios (OR) and 95% confidence intervals (CI) calculated using logistic regression models adjusted for all covariates.

Table 3-2. Descriptive characteristics of samples used to identify AAbs associated with BLBC.

	Sample Set 1		Sample Set 2		Sample Set 3		Non-basal subtypes		
	basal-like	healthy	basal-like	healthy	basal-like	healthy	LumA	LumB	HER2
	n=45	n=45	n=50	n=50	n=50	n=50	N=30	N=22	N=18
Age, mean (SD)	53.1(9.8)	51.6 (9.6)	54.1 (10.4)	54.1(10.4)	52.6 (11.2)	52.6 (11.2)	54.3 (11.0)	56.5 (9.6)	58.0 (7.1)
Age min-max	32-74	30-70	31-72	31-72	34-74	34-74	36-73	39-73	46-71
Parous (%)	93	84	84	82	84	82	87	82	83
Menopausal (%)	64	56	74	58	58	52	70	73	83
Poorly differentiated (%)	76		43		48		3	23	34
Node positive (%)[a]	29		53		42		47	45	56
>2mm size (%)	62		62		62		33	50	72

[a]Chi2 $P<0.05$ between sample set

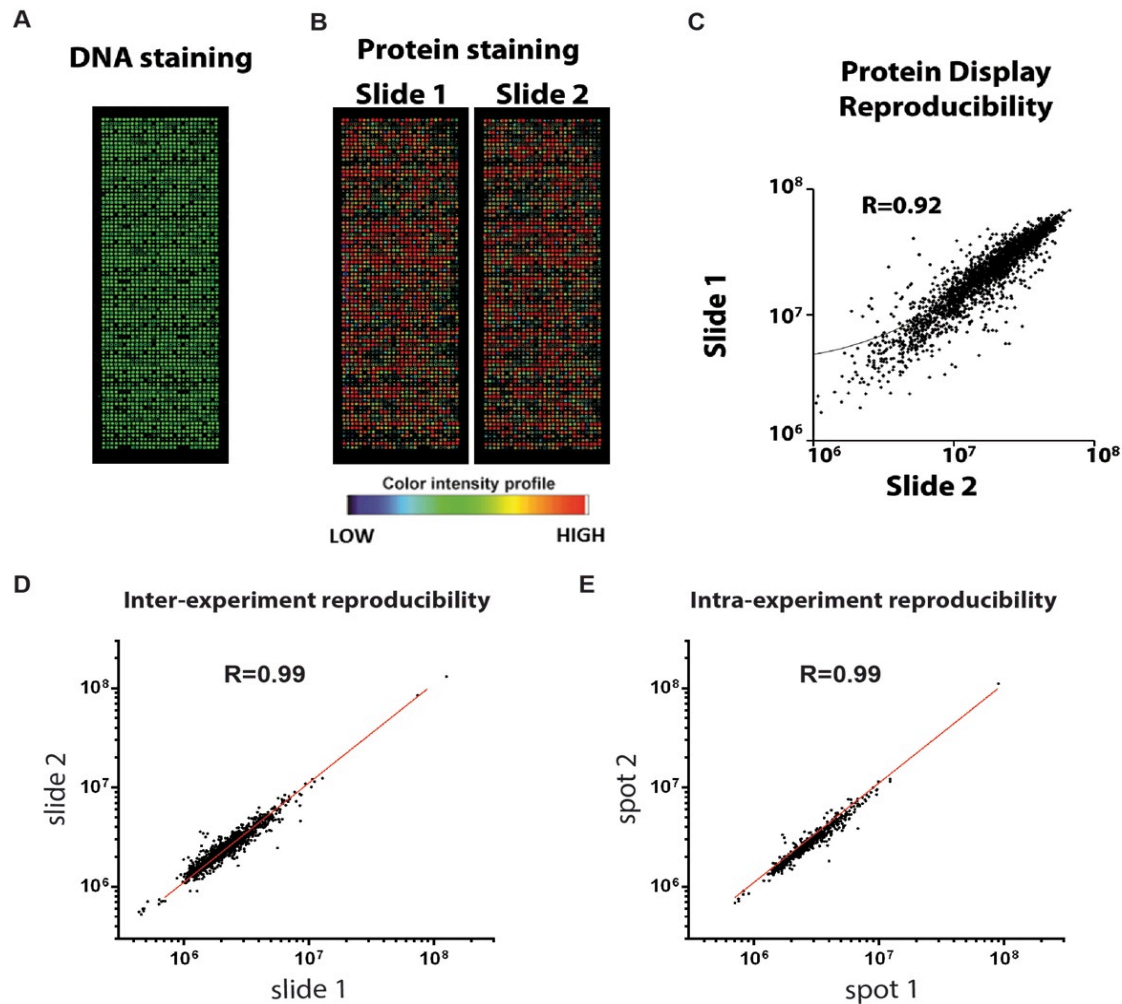


Figure 3-3. Protein array screening. Protein array was quality controlled by both DNA staining with picogreen (**A**), and anti-GST staining of *in vitro* produced proteins (**B**). **C.** Scatterplot of signal intensity measures of protein displays of two protein arrays produced in one single print batch. **D.** Scatterplot of signal intensity measures of two identical protein arrays probed with the same plasma sample. E. Scatterplot of signal intensity measures of two replicate spots with one array.

We first assessed the AAb immune responses in these 45 patient and control pairs using the quantified fluorescent intensity data for each feature. The median numbers of AAb responses per plasma sample across our basal-like and healthy sample sets were 14 and 17, respectively, with no statistical difference. The range for all samples was 1-121 (Figure 3-5A). AAbs against EPCAM (23.3%), AMY2A (20%), DPAGT1 (17.8%),

SLC10A1 (16.7%), GHRHR (16.7%), STMN4 (16.7%), RPSA (15.6%), RBPJ (14.4%), SSB (14.4%) were among the 22 proteins (0.2% of all proteins tested) with an overall prevalence above 10% among all 90 samples tested. AAbs against 876 proteins (7.6% of all proteins tested) were observed in at least one plasma sample.

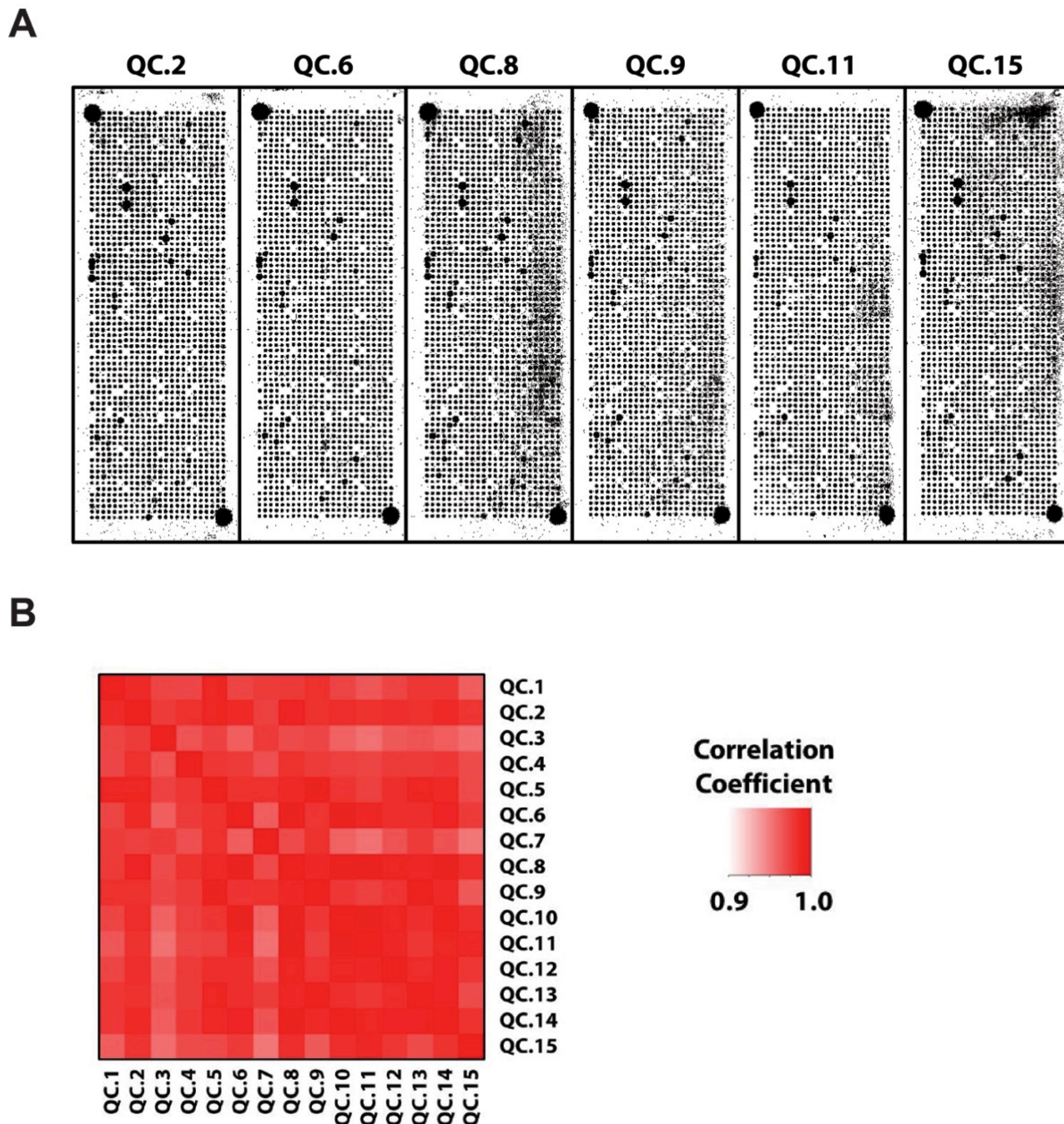


Figure 3-4. Array screening reproducibility. **A**. Representative images of the same plasma sample probed on identical protein arrays on different days. **B**. Heatmap of Pearson correlation coefficients between any two identical protein arrays probed with the same plasma sample on different days.

In parallel, we visually analyzed all images to document diffused signals that could not be accurately quantified by the image analysis software (Figure 3-2). The median numbers of AAb responses per plasma sample across our basal-like and healthy sample sets were 107 and 88, respectively, with no statistical difference. The range for all samples was 36-188 (Figure 3-5B). AAbs against ODF2 (100%), AMY2A (97%), RBPJ (78%), LENG1 (73%), CSF3 (63%), EPCAM (56%), POLDIP3 (54%), DGCR14 (53%) had overall prevalence above 50% among all 90 samples tested, and presented similar prevalence in both basal-like cases and controls. AAbs against 172 proteins (1.5% of all proteins tested) had overall prevalence above 10% among all 90 samples tested. AAbs against 2701 proteins (23.6% of all proteins tested) were observed in at least one plasma sample.

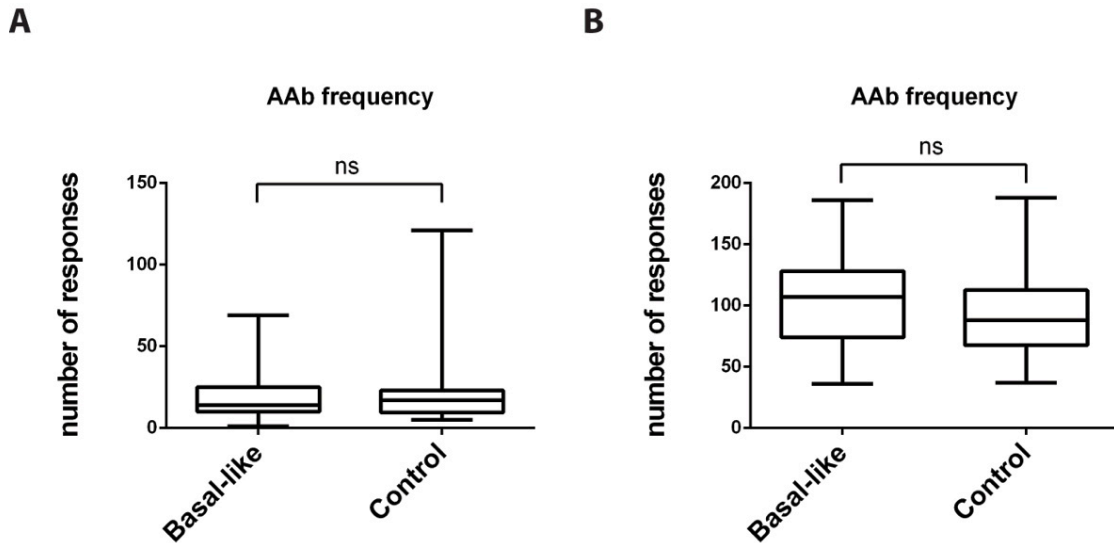
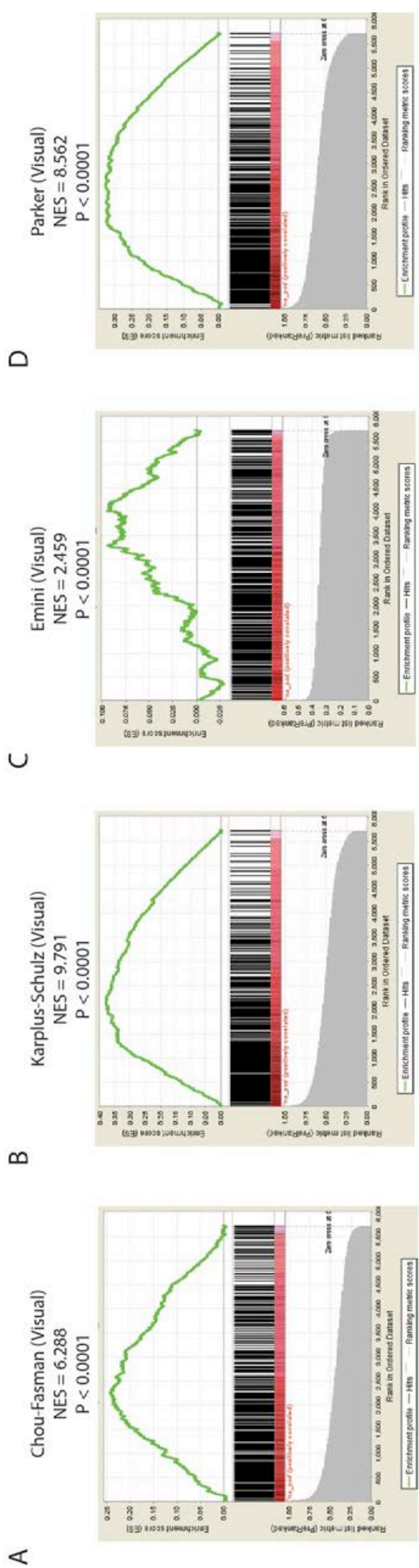


Figure 3-5. AAb response counts per sample are similar between BLBC and controls. **A**. Boxplot of AAb counts per sample using data from quantitative analysis. **B**. Boxplot of AAb counts per sample using data from visual analysis. (ns: not significant based on Wilcoxon Rank-Sum test)


A
Chou-Fasman (Visual)
NES = 6.288
P < 0.0001
B
Karplus-Schulz (Visual)
NES = 9.791
P < 0.0001
C
Emini (Visual)
NES = 2.459
P < 0.0001
D
Parker (Visual)
NES = 8.562
P < 0.0001

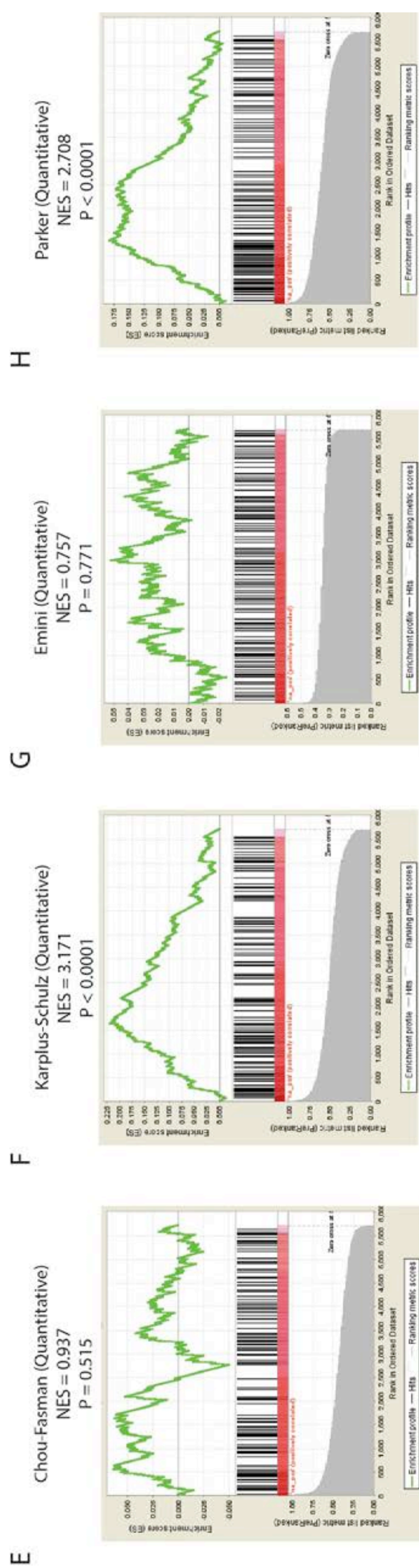

E
Chou-Fasman (Quantitative)
NES = 0.937
P = 0.515
F
Karplus-Schulz (Quantitative)
NES = 3.171
P < 0.0001
G
Emini (Quantitative)
NES = 0.757
P = 0.771
H
Parker (Quantitative)
NES = 2.708
P < 0.0001

Figure 3-6. GSEA analysis of AAb targets against predicted antigenicity. Proteins were ranked in decreasing order by ratios of antigenic peptide windows predicted using Chou and Fasman beta turn prediction (**A, E**), Karplus and Schulz flexibility scale (**B, F**), Emini surface accessibility scale (**C, G**), or Parker Hydrophilicity Scale (**D, H**). NES: Normalized Enrichment Score. Higher absolute value of NES indicates stronger association.

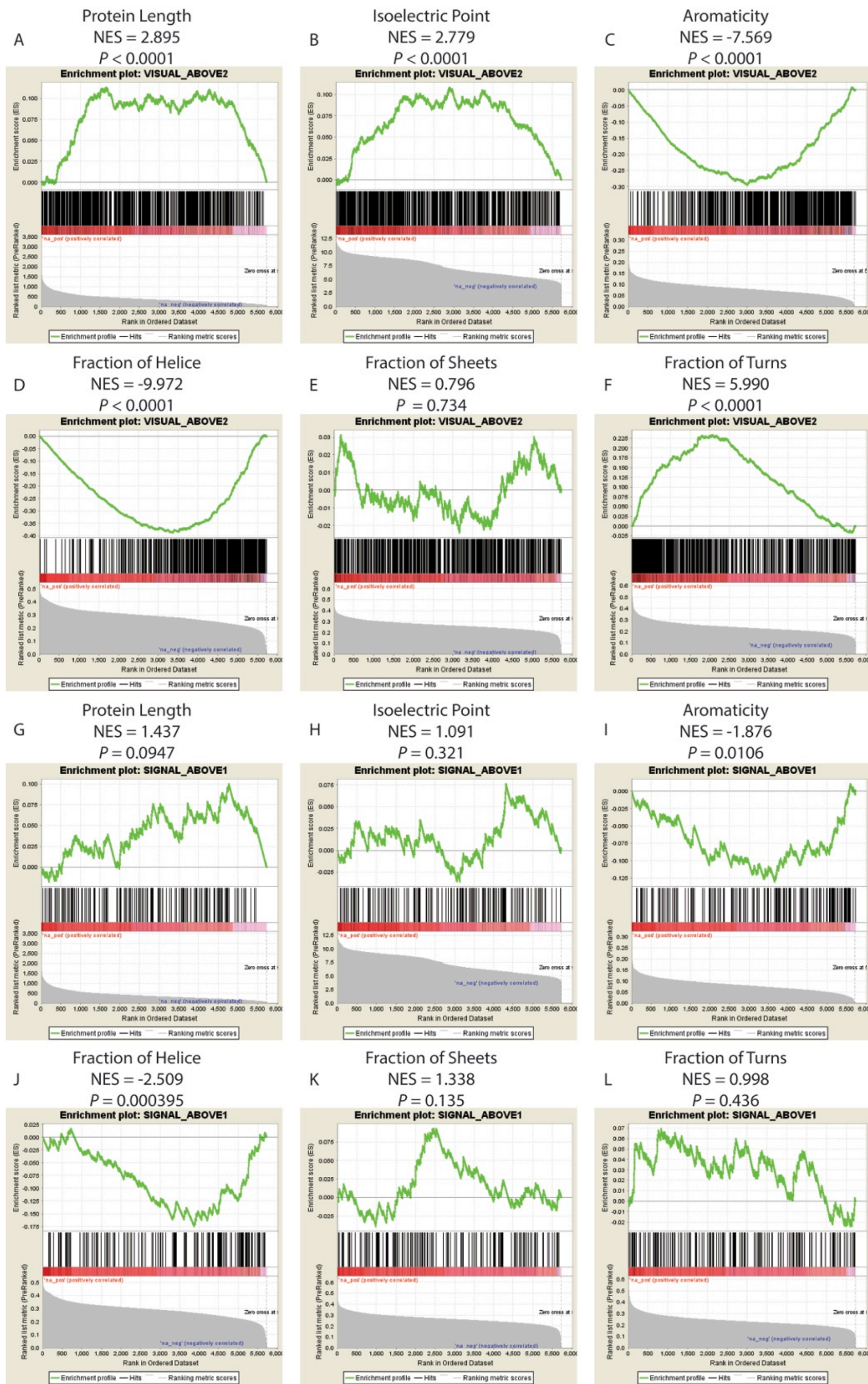
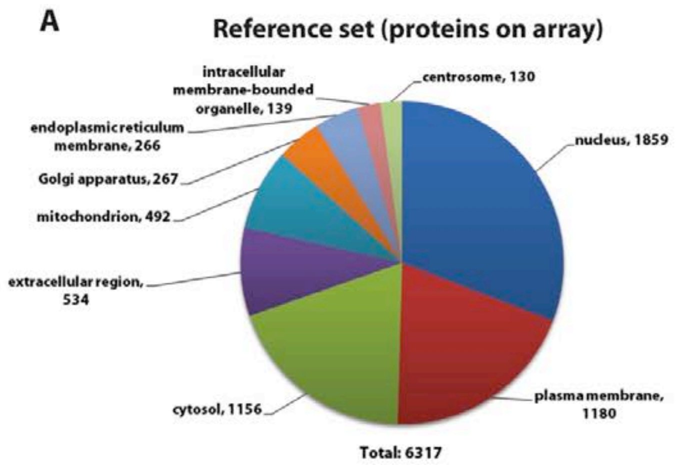
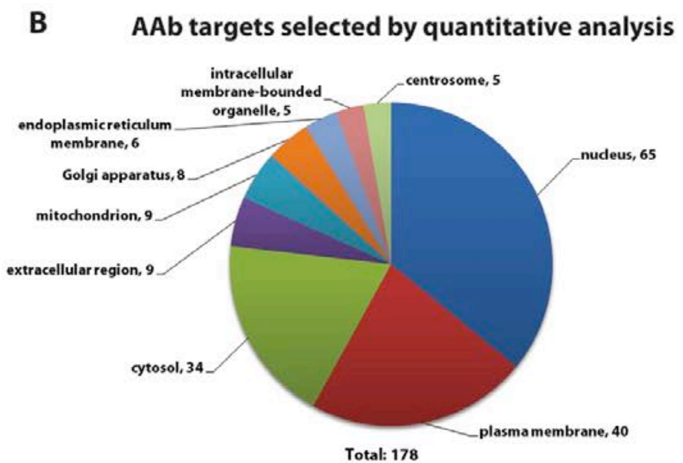


Figure 3-7. GSEA analysis of AAb protein targets against their biochemical properties. Proteins were ranked in decreasing order by protein length, isoelectric point, aromaticity, or fractions of secondary structures using AAb targets selected by either visual analysis (**A-F**) or quantitative analysis (**G-L**). Aromaticity and fraction of helices appeared negatively associated with the observed antigenicity.


A
Reference set (proteins on array)
intracellular membrane-bounded organelle, 139
centrosome, 130
endoplasmic reticulum membrane, 266
nucleus, 1859
Golgi apparatus, 267
mitochondrion, 492
extracellular region, 534
cytosol, 1156
plasma membrane, 1180
Total: 6317


B
AAb targets selected by quantitative analysis
intracellular membrane-bounded organelle, 5
centrosome, 5
endoplasmic reticulum membrane, 6
nucleus, 65
Golgi apparatus, 8
mitochondrion, 9
extracellular region, 9
cytosol, 34
plasma membrane, 40
Total: 178

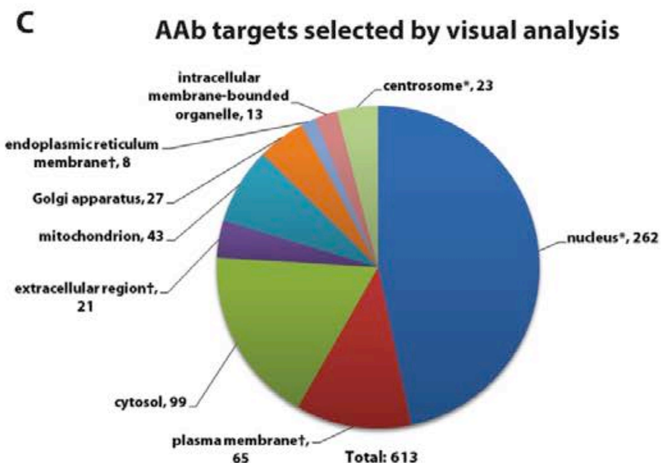

C
AAb targets selected by visual analysis
intracellular membrane-bounded organelle, 13
centrosome*, 23
endoplasmic reticulum membrane†, 8
Golgi apparatus, 27
mitochondrion, 43
nucleus*, 262
extracellular region†, 21
cytosol, 99
plasma membrane†, 65
Total: 613

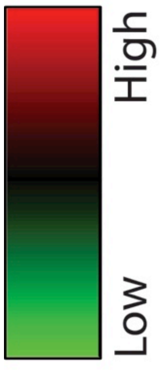
Figure 3-8. GO cellular component analysis. Note that numbers do not add up to total, because: 1. There are overlapping genes among GO terms; 2. Only top GO terms were used to construct the pie chart (see Supplementary Materials). *significantly enriched (BH adjusted $P<0.05$). †significantly depleted (BH adjusted $P<0.05$).

In order to determine if there were common protein properties that may contribute to autoimmune responses, we assessed the association between AAb responses and their target proteins' properties, such as protein length, isoelectric point, aromaticity and fractions of predicted secondary structures, as well as predicted antigenicity by common algorithms. Gene set enrichment analysis (GSEA) indicated significant over representation of protein array discovered AAb target proteins in proteins predicted with high antigenicity (Figure 3-6). We also observed AAb responses are positively associated with protein length, isoelectric point, and fraction of turns, and negatively associated with aromaticity, fraction of α helice (Figure 3-7). Gene Ontology cellular component analysis indicated that AAb target proteins identified by visual analysis were significantly enriched in nucleus and centrosome; while significantly depleted in plasma membrane, extracellular region and endoplasmic reticulum membrane (Figure 3-8).

3.4.3 Identification of candidate AAb biomarkers for basal-like breast cancer

In order to find which AAbs behaved as biomarkers, we took two complementary approaches, using focused protein arrays and ELISA. To design the focused array, quantified intensity data for each protein in sample set 1 were analyzed by multiple statistical tests with the goal of eliminating uninformative proteins and selecting antigens with higher reactivity in basal-like cases. This allowed us to reduce the number of biomarker candidates such that each antigen could be studied in duplicate on the focused array, reducing both cost and the risk of overfitting (Anderson et al., 2011a; Wallstrom et

al., 2013). The statistical filters included but were not limited to: sensitivity at 95% specificity (228 proteins); area under receiver operating characteristic curve (AUC; 185 proteins); partial AUC above 95% specificity (pAUC; 197 proteins); p value for Welch's t test (197 proteins). In addition, by visually analyzing array images, we also observed antigens with higher prevalence of AAb responses in basal-like disease than that in


High
Low

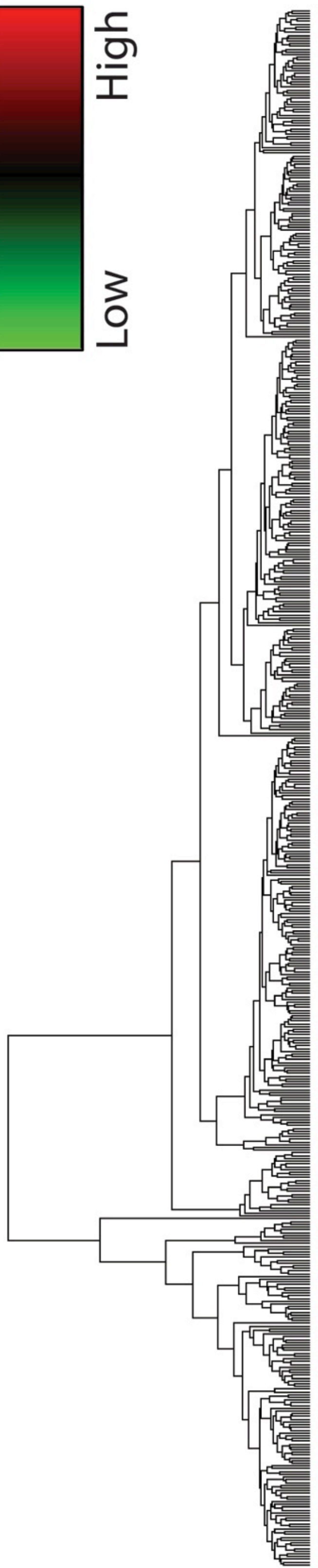

Basal-like
Control

Figure 3-9. Heatmap of plasma AAb responses against selected proteins based on quantitative analysis in sample set 1 (basal-like, n=45; controls, n=45). This figure displays the log2-transformed normalized intensity data of 565 proteins selected from the screening study to be included on the focused array. For graphical purposes (to avoid having the outliers drive the display), heatmap constrains log2-transformed values to lie between +/- 3. Individual spots are z-transformed and clustered using Euclidean distance.

healthy controls. At individual protein levels, TP53 (basal-like: 24%; healthy: 4%), RNF216 (basal-like: 22%; healthy: 4%) were among the AAb targets that were significantly associated with basal-like disease. We selected 748 antigens based on the above analysis (Figure 3-9; Figure 3-10), produced a set of focused arrays, and probed it with sample set 2 (Table 3-2). Fourteen antigens with sensitivities above 10% at 98% specificity and K>0.9 were selected for the blinded test (Table 3-3).

Table 3-3. List of 14 antigens selected from focused array experiments. Individual sensitivities were determined at 98% specificity (Sample set 2: basal-like, n=50; controls, n=50).

Antigen	Sensitivity	Specificity	K
CTAG1B	0.280	0.980	84.425
TP53	0.200	0.980	1.046
CTAG2	0.200	0.980	54.890
TAS2R8	0.180	0.980	1.042
LMO4	0.160	0.980	0.928
KCNIP3	0.160	0.980	0.948
PVRL4	0.140	0.980	1.280
POU4F1	0.120	0.980	3.270
ZBTB16	0.120	0.980	1.558
CCDC68	0.120	0.980	1.052
PPHLN1	0.120	0.980	0.949
TSGA13	0.120	0.980	0.917
TRIM21	0.120	0.980	2.309
WBP2NL	0.120	0.980	1.378

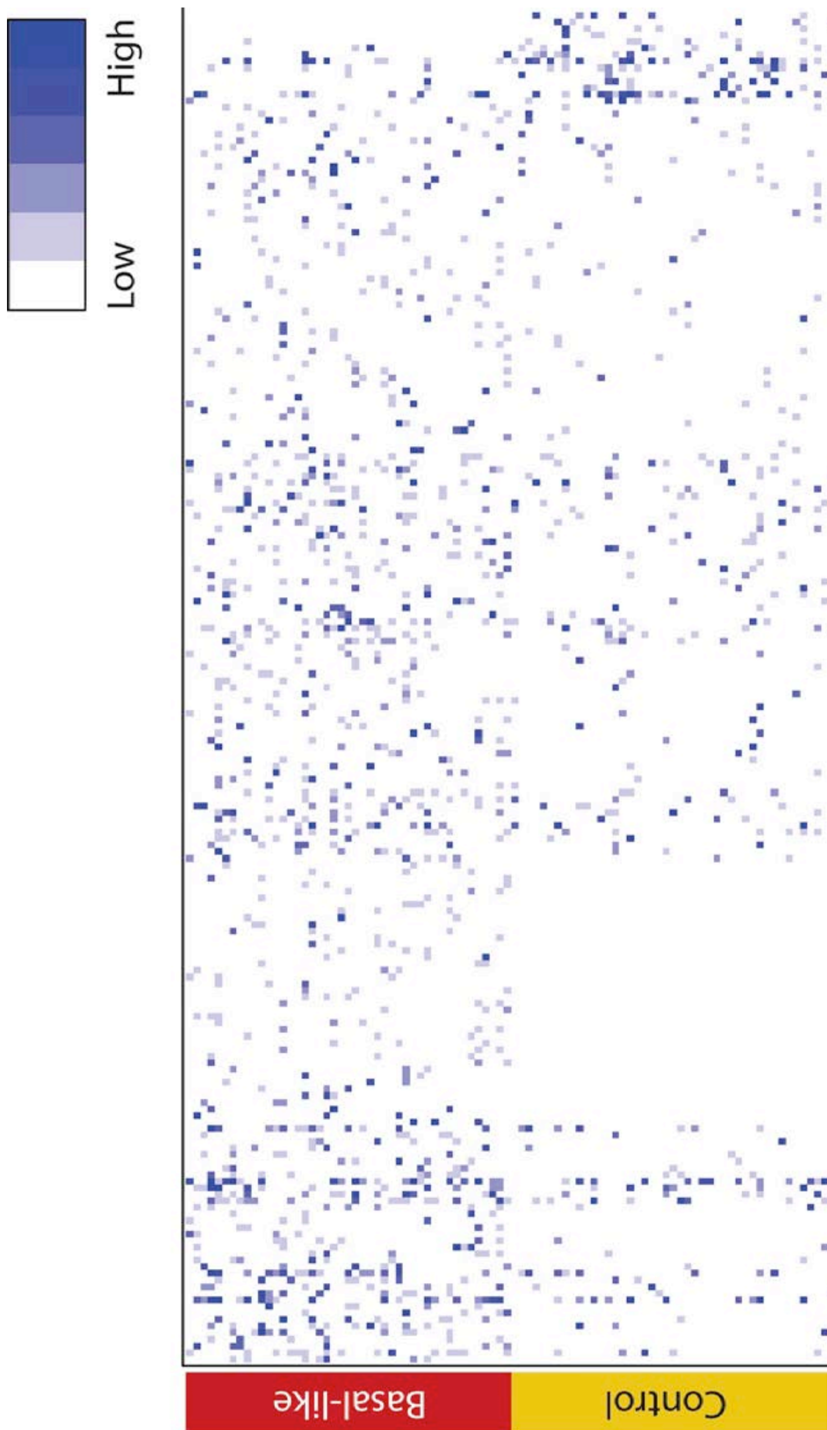


Figure 3-10. Heatmap of plasma AAb responses against selected proteins based on visual image analysis in sample set 1 (basal-like, n=45; controls, n=45). This figure displays the visual analysis of 205 proteins selected from the screening study to be included on the focused array. Spots are ordered first by decreasing ring (grade >= 1) prevalence across all subjects, then by decreasing mean ring grade across all subjects. Proteins towards the right end showed higher AAb responses in controls.

In parallel, we also selected 82 antigens that showed the most differential prevalence in sample set 1 between basal-like cases and controls from the visual analysis to verify the same samples by programmable ELISA. We successfully developed programmable ELISA assay for 71 of them. From these, 15 antigens with sensitivities above 10% at 98% specificity and K>0.9 (Table 3-4) were selected. Combining these two sources and accounting for duplication, AAbs against 26 unique proteins were identified for a subsequent blind test by ELISA.

Table 3-4. List of 15 antigens selected from ELISA verification. Individual sensitivities were determined at 98% specificity. (Sample set 1: basal-like, n=43; controls, n=43).

Antigen	Sensitivity	Specificity	K
SNRK	0.209	0.977	2.261
RNF216	0.163	0.977	4.950
MN1	0.163	0.977	2.109
CTAG2	0.140	0.977	8.544
DYRK3	0.140	0.977	4.798
TP53	0.140	0.977	8.932
BCL2	0.140	0.977	10.584
SSMEM1	0.140	0.977	6.851
PIP4K2C	0.116	0.977	8.732
KRT8	0.116	0.977	1.460
DOK2	0.116	0.977	2.924
CTAG1B	0.116	0.977	12.712
RNF32	0.116	0.977	1.354
JUNB	0.116	0.977	2.679
PSRC1	0.116	0.977	2.212

3.4.4 Blind test of AAb biomarkers for basal-like breast cancer

To test these 26 candidate antigens, a training study using ELISA was performed using sample sets 1 and 2, in order to set threshold values for each antigen. We then blindly tested their performance using an independent and untouched sample set 3 (Table

3-2, Figure 3-1). In the training set, 16 antigens showed sensitivity above 5% at 98% specificity (Table 3-5). Plasma antibodies against CTAG1B and CTAG2 proteins demonstrated the best ability to differentiate patients from controls. In the test set, all 26 antigens were assessed using cutoff values defined in the training study. In summary, antibodies against TP53, CTAG1B, PPHLN1, WBP2NL, DOK2 showed sensitivities above 5% at 98% specificity in both training and test phases. AAbs against CTAG2 showed lower specificity (96%) in the test phase, but its sensitivity remained at 18% (Table 3-5).

3.4.5 An AAb signature for diagnosing basal-like breast cancer

To evaluate the diagnostic performance of the candidate biomarkers as a panel, we combined ELISA results of all basal-like patients and healthy controls (Table1, sample sets 1-3). We included all antigens with a minimum sensitivity of 5% at 98% specificity, yielding 13 antigens that met our inclusion criteria (Table 3-6). The 13 AAb classifier has an AUC of 0.68, which was calculated using leave-one-out (LOO) cross validation (Table 3-6). The receiver operating characteristic (ROC) curve was computed under LOO cross validation by varying the cutoff values in the prediction model (Figure 3-11). The plasma AAb model predicted basal-like patients from healthy controls at a sensitivity of 33% and a specificity of 98%.

Table 3-5. Training and test statistics for potential BLBC autoantibody biomarkers.

Antigen	Training (Sample sets 1&2: basal, n=95; healthy, n=95)			Blinded Test (Sampel set 3: basal, n=50; healthy, n=50)	
	sensitivity	specificity	cutoffs[a]	sensitivity	specificity
CTAG1B	0.213	0.979	1.606	0.200	1.000
CTAG2	0.191	0.979	1.149	0.180	0.960
TRIM21	0.158	0.979	1.208	0.140	0.860
RNF216	0.110	0.978	1.369	0.043	0.956
MN1	0.105	0.979	1.311	0.060	0.920
PIP4K2C	0.105	0.979	1.200	0.020	1.000
TP53	0.084	0.979	3.171	0.200	1.000
ZBTB16	0.084	0.979	1.393	0.040	0.980
DOK2	0.074	0.979	1.164	0.060	1.000
PPHLN1	0.063	0.979	3.394	0.080	1.000
TAS2R8	0.063	0.979	1.064	0.080	0.940
SSMEM1	0.063	0.979	1.562	0.060	0.960
DYRK3	0.063	0.979	1.462	0.040	0.940
KRT8	0.053	0.979	1.645	0.060	0.960
LMO4	0.053	0.979	1.199	0.020	0.980
WBP2NL	0.053	0.979	1.991	0.060	0.980
JUNB	0.042	0.979	1.165	0.020	0.960
TSGA13	0.042	0.979	1.313	0.020	0.980
PVRL4	0.042	0.979	0.899	0.020	0.920
CCDC68	0.042	0.979	2.438	0.000	0.940
BCL2	0.042	0.979	1.160	0.000	1.000
SNRK	0.032	0.979	4.127	0.020	0.960
PSRC1	0.032	0.979	1.372	0.120	0.960
KCNIP3	0.032	0.979	0.973	0.000	0.960
POU4F1	0.032	0.979	0.992	0.080	0.940
RNF32	0.021	0.979	1.445	0.040	0.980

[a]ELISA relative absorbance at 98 percentile of controls.

Table 3-6. 13-AAb classifier.

Antigen	Sensitivity[a]	Specificity	Cutoff[b]
CTAG1B	0.208	0.979	1.606
CTAG2	0.188	0.979	1.176
TP53	0.124	0.979	3.171
RNF216	0.088	0.978	1.459
PPHLN1	0.083	0.979	3.448
PIP4K2C	0.076	0.979	1.201
ZBTB16	0.069	0.979	1.925
TAS2R8	0.069	0.979	1.178
WBP2NL	0.069	0.979	2.120
DOK2	0.069	0.979	1.164
PSRC1	0.063	0.979	1.461
MN1	0.062	0.979	1.687
TRIM21	0.055	0.979	5.509

[a]Individual sensitivities were determined at 98% specificity. (basal-like, n=145; control, n=145).
[b]Antigen-specific cutoffs of ELISA relative absorbance used in the 13-AAb classifier. A sample was classified as positive if it exceeds the cutoff value for two or more antigens.

Table 3-7. Performance of 13 AAbs in multiple breast cancer subtypes. (Basal-like, n=145; luminal A, n=30; luminal B, n=22; her2-enriched, n=18; controls, n=145).

Antigen	Sensitivity				Specificity
	basal-like	luminal A	luminal B	Her2-enriched	
CTAG1B	0.208	0.033	0.045	0.056	0.979
CTAG2	0.188	0.000	0.000	0.000	0.979
TP53	0.124	0.033	0.000	0.056	0.979
RNF216	0.088	0.133	0.095	0.000	0.978
PPHLN1	0.083	0.100	0.182	0.000	0.979
PIP4K2C	0.076	0.100	0.091	0.111	0.979
ZBTB16	0.069	0.000	0.000	0.000	0.979
TAS2R8	0.069	0.000	0.000	0.056	0.979
WBP2NL	0.069	0.100	0.091	0.000	0.979
DOK2	0.069	0.133	0.091	0.056	0.979
PSRC1	0.063	0.033	0.045	0.056	0.979
MN1	0.062	0.100	0.000	0.056	0.979
TRIM21	0.055	0.033	0.000	0.056	0.979

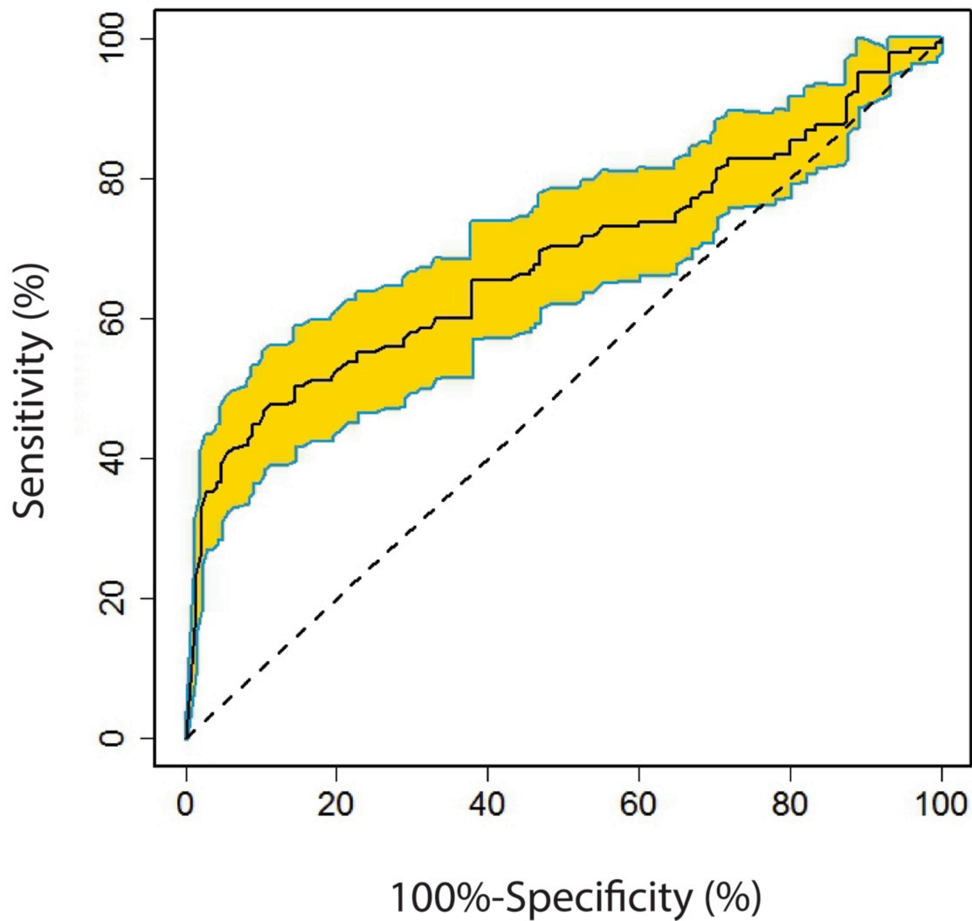


Figure 3-11. ROC curve of the 13-AAb classifier. 95% confidence interval was also computed (yellow).

3.4.6 AAb response in other breast cancer subtypes

We then examined the specificity of these top markers in basal-like subtype compared with other breast cancer subtypes. Specifically, we assayed AAb responses against the 13 proteins in plasma samples from 30 Luminal A, 22 Luminal B, and 18 HER2-enriched patients by ELISA (Table 3-2). Results indicated that AAbs targeting CTAG1B, CTAG2 and TP53 were significantly higher in basal-like patients' plasma (Figure 3-12; Table 3-7) relative to other breast cancer subtype

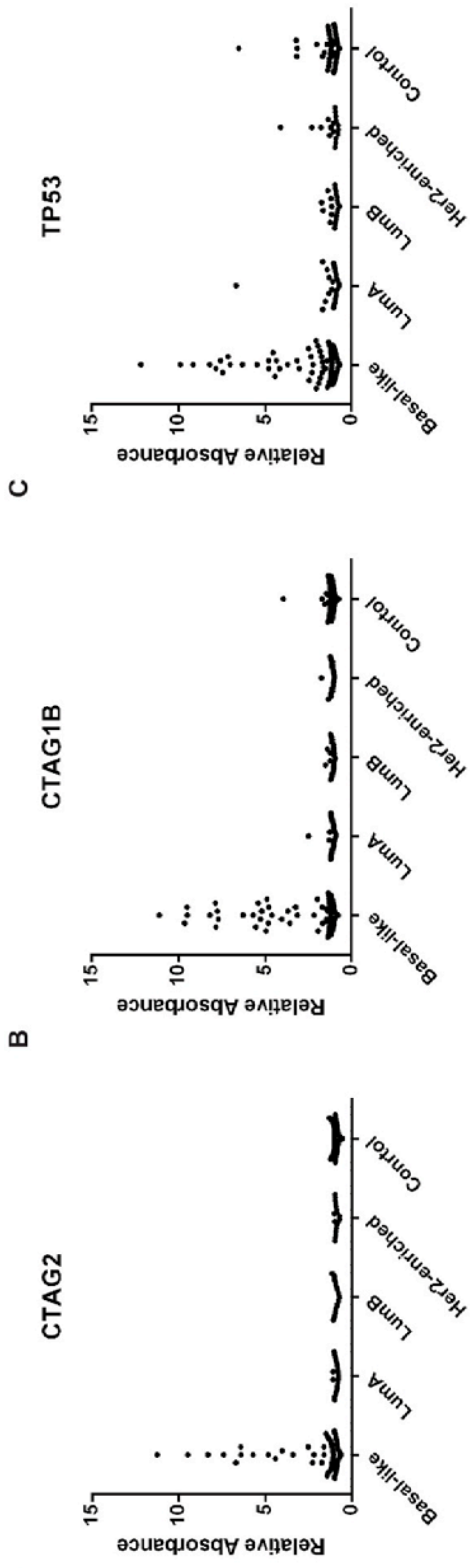


Figure 3-12. AAb responses in various breast cancer subtypes.

3.4.7 TP53 AAb response and TP53 protein levels

Because disease-specific AAbs are usually associated with presence of the corresponding antigens in the tumor tissue (Reuschenbach et al., 2009), we examined the protein level of TP53 in the same basal-like population used for this study. This allows direct comparison of AAb response and protein levels. Immunohistochemistry staining of TP53 was available on tumor sections from 79 basal-like patients using tissue microarrays. We divided these 79 patients into two groups based on their antibody response to TP53, namely greater or lower than the 95 percentile of the healthy controls. We found a significant association between TP53 protein levels and AAb responses (p=0.009). Specifically, 16 out of 54 patients with positive TP53 staining in their tissue sections also had positive TP53 AAb responses; while only 1 out of 25 patients with negative staining of TP53 proteins in their tissue sections had positive TP53 AAb responses (Table 3-8).

Table 3-8. Relationship between TP53 AAb response and TP53 tissue IHC score in basal-like breast cancers.

TP53 IHC	TP53 AAb				
	non-response		response		
	N	%	N	%	P^a
Negative	24	39	1	6	
Positive	38	61	16	94	0.009

[a]*P*-value calculated from Fisher's exact test

For the remaining 12 AAb biomarkers, using TCGA breast cancer data we found CTAG1B, RNF216 and PSRC1 to show significantly elevated mRNA levels in BLBC compared with other subtypes (Cline et al., 2013) (Figure 13). Other markers did not show any significant changes in mRNA expression from TCGA.

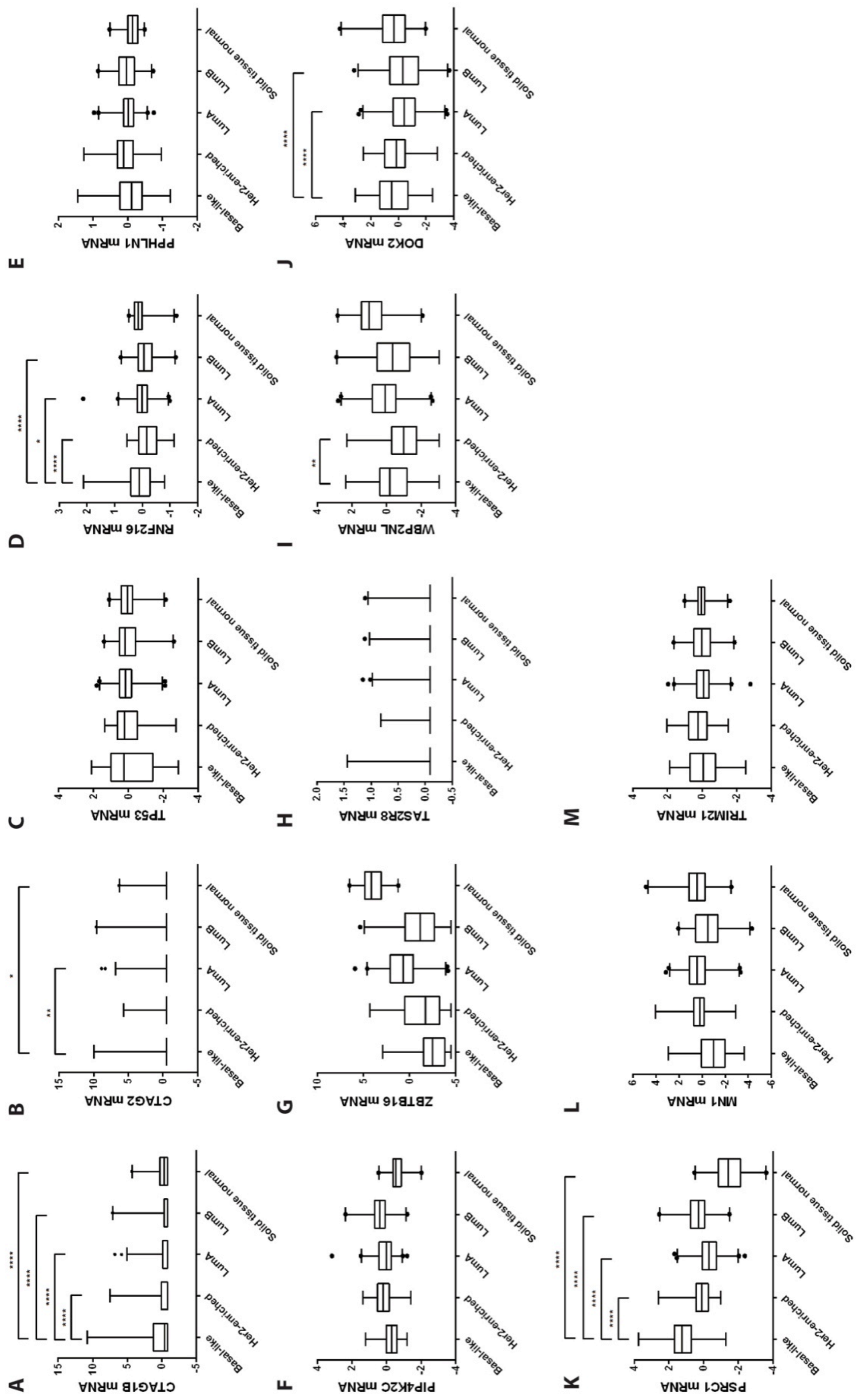

A
CTAG1B mRNA
B
CTAG2 mRNA
C
TP53 mRNA
D
RNF216 mRNA
E
PPHLN1 mRNA
F
PIP4K2C mRNA
G
ZBTB16 mRNA
H
TAS2R8 mRNA
I
WBP2NL mRNA
J
DOK2 mRNA
K
PSRC1 mRNA
L
MN1 mRNA
M
TRIM21 mRNA
Basal-like
Her2-enriched
LumA
LumB
Solid tissue normal

Figure 3-13. mRNA expression levels of candidate biomarkers in breast invasive carcinoma grouped by subtypes (TCGA). *P* values were obtained using ANOVA comparing basal-like subtype with other subtypes and normal tissue respectively, and only significantly higher mRNA levels in basal-like disease were labeled (*:$P<0.05$, **: $P<0.01$, ***: $P<0.001$, ****: $P<0.0001$). *P* values were adjusted for multiple comparisons by Sidak correction. **A**. CTAG1B mRNA; **B**. CTAG2 mRNA; **C**. TP53 mRNA; **D**. RNF216 mRNA; **E**. PPHLN1 mRNA; **F**. PIP4K2C mRNA; **G**. ZBTB16 mRNA; **H**. TAS2R8 mRNA; **I**. WBP2NL mRNA; **J**. DOK2 mRNA; **K**. PSRC1 mRNA; **L**. MN1 mRNA; **M**. TRIM21 mRNA.

3.4.8 TP53 AAb response and *TP53* mutation

Mutation could provide another plausible source of targets of immune responses(Reuschenbach et al., 2009; Soussi, 2000). We then evaluated whether mutation in the *TP53* gene was associated with AAb development. *TP53* mutation information was available for 21 basal-like patients from our study cohort. Among them, 16 had mutations in the *TP53* gene. However, none of them had AAb responses higher than 95 percentile of healthy controls, whereas 1 out of 5 patients with wildtype *TP53* gene had an AAb response against TP53 protein. Therefore, we did not observe significant association between *TP53* mutation and its AAb responses in our study cohort of patients with basal-like breast cancer (Table 3-9).

Table 3-9. Relationship between TP53 AAb response and TP53 mutation in basal-like breast cancers.

	TP53 autoantibody				
	non-response		**response**		
	N	**%**	**N**	**%**	***P****
TP53 mutation					
Negative	4	20	1	100	
Positive	16	80	0	0	0.238

[a]*P*-value calculated from Fisher's exact test

3.4.9 AAbs and overall survival in basal-like breast cancer

To address whether a humoral immune response against these antigens can predict clinical outcome, 13 antigens with at least 5% sensitivity at 98% specificity were evaluated by Kaplan-Meier analysis comparing overall survival in patients with elevated AAbs and those without (Table 3-6). Basal-like cases were divided into two groups based on their antibody level to each antigen, greater or lower than the 95 percentile of the healthy controls. The results indicated that AAbs against TP53 and MN1 proteins were associated with overall survival. Patients with AAbs against TP53 protein presented worse prognosis than those without responses (p=0.03) (Figure 3-14A). Patients with AAbs against MN1 protein also presented worse survival than those without responses (p=0.04) (Figure 3-14B). After adjustments for age, menopausal status, grade, node status and tumor size, there was attenuation of results. Nonetheless, although not significant, they were still related to worse outcome (Table 3-10). We did not observe any association between survival and AAb responses in the remaining 11 antigens.

Table 3-10. Hazard ratios for 2 antigen AAb responses that are significantly associated with survival amongst 145 basal like breast cancers.

Antigen	**non-response**	**response**	**Univariate**		**Multivariate[b]**	
	alive/dead[a]	**alive/dead[a]**	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
MN1	95/30	9/8	2.25 (1.03-4.91)	0.04	2.13 (0.86-5.25)	0.10
TP53	84/23	20/13	2.02 (1.06-3.85)	0.03	1.5 (0.70-3.20)	0.30

[a]note N may not sum to 145 due to missing data
[b]Multivariate models adjusted for age, menopausal status, grade, node status, and tumor size

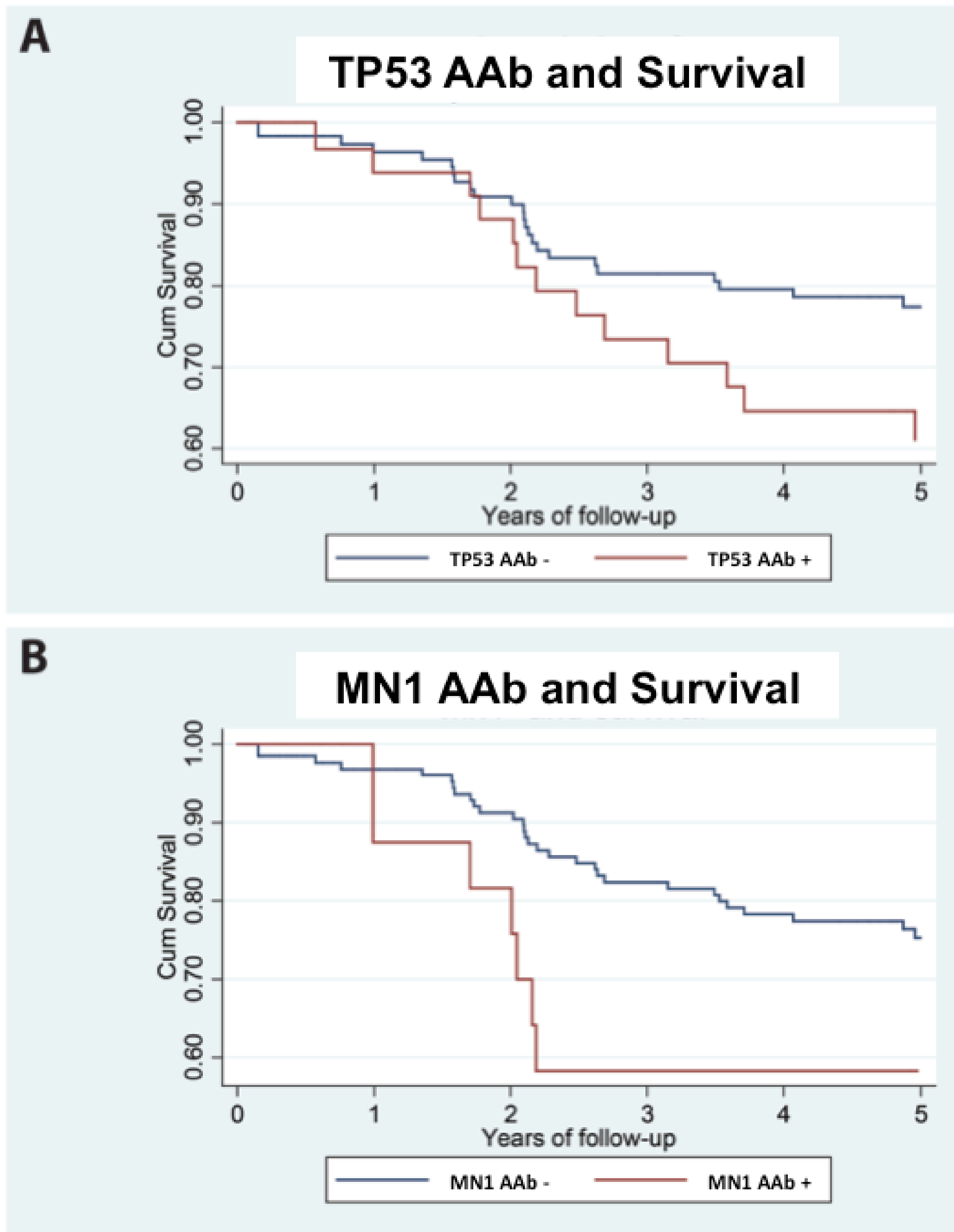


Figure 3-14. Overall survival of BLBC patients stratified by AAb responses. **A.** TP53 (*P*=0.03)**; B.** MN1 (*P*=0.04)**.**

3.5 Discussion

We have performed the largest biomarker discovery screen from women with basal-like breast cancer. The Polish Breast Cancer Study enabled the investigation of this low prevalence subtype with a large cohort of well-characterized samples to address the challenges associated with disease heterogeneity in biomarker studies. By using a highly reproducible protein microarray platform that tested about proteins encoded by half of the human genome, we estimate that most adults produce in the range of 50-250 antibodies against self-antigens, which did not differ significantly between cases and controls.

Our results were further confirmed by strong enrichment of proteins predicted with high antigenicity. We also observed both positive and negative association between autoantigenicity and various antigen properties. In particular, proteins in nucleus and centrosome were highly enriched in AAb targets. These results may provide insights for studying autoimmunity. Nevertheless, hypothesis driven researches are needed to further our knowledge.

From ~10,000 human proteins, we identified 13 AAbs preferentially enriched in basal-like disease when compared to healthy individuals; six of which passed a blinded test, representing the first verified blood biomarkers for basal-like breast cancer. Sensitivities ranged from 6% to 21% at 98% specificity, which are typical for AAbs. An AAb signature was constructed and that predicts basal-like disease with 33% sensitivity at 98% specificity. We further confirmed the specificity of CTAG1B, CTAG2 and TP53 AAb biomarkers to basal-like relative to other breast cancer subtypes, confirming the hypothesis that different cancer subtypes are likely to have unique biomarkers.

Significant effort has focused on searching for tissue biomarkers for TNBC; however, tissue markers are precluded from use in early detection (Cabezon et al., 2013; Koziol et al., 2003; Liu et al., 2014). Li et al. used antibody arrays to identify 29 plasma proteins that differentiated TNBC from healthy controls; although, the study had a relative small sample size (TNBC, n=28, Healthy, n=28)(Li et al., 2012). Serological analysis of breast cancer, without attention to subtype, has yielded a number of candidate AAb biomarkers.(Desmetz et al., 2009; Koziol et al., 2003; Lacombe et al., 2013) Most of these are awaiting validation, although one has been tested in an unblinded study (Lacombe et al., 2013). Using a study with 155 breast cancer cases, 77 benign cases and 53 healthy controls, we also described a signature of 28 AAbs that discriminated malignant breast cancer from benign breast disease in a blinded verification study (Anderson et al., 2011a). We included these reported biomarkers in this study; however, no AAbs against these antigens survived the selection criteria as markers for basal-like breast cancer. Presumably, most of the reported AAb biomarkers for these studies without subtyping would perform best in ER+ breast cancer.

By testing ~10,000 human proteins, we obtained an unbiased picture of the AAb responses in basal-like patients at the proteome level. The NY-ESO-1 antigen, the protein product of *CTAG1B* or *CTAG1A* gene, was first discovered in esophageal cancer by serological analysis of expression cDNA libraries (Chen et al., 1997). This is the first demonstration of AAb against NY-ESO-1 protein as biomarkers in basal-like patients' plasma samples, which emerged as the best performer with 20.8% sensitivity at 98% specificity in basal-like patients. In a recent targeted study, Ademuyiwa et al. observed elevated protein levels of NY-ESO-1 and the elicitation of antibodies in

TNBC(Ademuyiwa et al., 2012). This observational study did not include controls nor any validation, but they found that 8 out of 11 patients with detectable NY-ESO-1 protein in TNBC tissue had positive plasma AAb reactivity. CTAG2 AAb was the second best performer from our study with 18.8% sensitivity at 98% specificity, which is not surprising considering the 91% sequence homology between coding region of *CTAG2* and *CTAG1B*. 28 out of 32 (88%) patients with AAb responses against CTAG2 protein also had positive responses against CTAG1B protein, using the 98 percentile of controls as cutoff. TP53 AAb has been reported in many different cancers including breast cancer with varying prevalence (Reuschenbach et al., 2009). However, its performance in basal-like subtype has never been assessed before. Here, we demonstrated that TP53 AAb has a sensitivity of 12.4% at 98% specificity in basal-like tumors, compared with 3% sensitivity at 98% specificity in 70 patients with other breast cancer subtypes.

Additional markers had less sensitivity than the three discussed above. PPHLN1 encodes periphilin-1, which is involved in epithelial differentiation (Kazerounian and Aho, 2003) and induces AAbs in both gastric cancer and breast cancer (Line et al., 2002). PSRC1 encodes the mitotic proline/serine rich coiled-coil protein 1 (Jang et al., 2008) and has AAbs included in a classifier for ductal carcinoma in situ to invasive breast carcinoma transition (Mange et al., 2012). AAbs to TRIM21, an E3 ubiquitin ligase that promotes p27 degradation, were initially associated with autoimmune rheumatic disease (Ghillani et al., 2011), but their appearance in cancer patients' sera was observed in subsequent studies (Kuboshima et al., 2006; Vazquez-Del Mercado et al., 2013). It also participates in destabilization of TP53 protein according to a recent study (Reddy et al., 2014).

The importance of disease heterogeneity on biomarker studies has only recently been emphasized (Wallstrom et al., 2013). We specifically designed this study using well-characterized plasma samples from a specific breast cancer subtype and statistical approaches that account for heterogeneous diseases. We have evaluated our validated biomarkers in patients with other breast cancer subtypes in order to determine their selectivity. Our results confirmed subtype specific biomarker performance and the advantage of biomarker discovery on a more homogeneous patient population. Specifically, we found that AAbs against CTAG1B, CTAG2 and TP53 proteins had significant selectivity towards basal-like subtype.

Interestingly, AAbs against TP53 and CTAG1B proteins have also been reported in patients with tumors from other organ sites, such as ovarian cancer(Anderson et al., 2010; Soussi, 2000; Stockert et al., 1998), colorectal cancer(Reuschenbach et al., 2009; Scanlan et al., 1998; Soussi, 2000), and lung cancer(Boyle et al., 2011; Reuschenbach et al., 2009; Soussi, 2000; Stockert et al., 1998). The rare occurrence of these two AAbs in other breast cancer subtypes (LumA, Lum B, HER2-enriched) may indicate the potential of AAbs as an indication of common underlying molecular mechanism of the development of basal-like disease and cancers from other organ sites. For example, it is well accepted that basal-like breast cancer and ovarian cancer share many common molecular features, including but not limited to *BRCA1* inactivation, *RB1* loss and *CCNE1* amplification, *MYC* amplification, high frequency of *TP53* mutation(Koboldt et al., 2012). According to the Cancer Genome Atlas (TCGA), transcriptomic profiles of basal-like tumors are also correlated with that of lung cancer as well as serous ovarian cancer (Koboldt et al., 2012). Improved understanding of the shared underlying

mechanism among different cancer types may help transfer the knowledge of treatment strategy among these cancers. Moreover, biomarkers that are selective for molecular subtypes may prove useful in clinical management.

The mechanism of AAb generation remains unclear. Possible explanations include high protein abundance in tumor tissues and mutations. Immunohistochemical staining of TP53 and CTAG1B proteins in previous studies indicated their increased tissue levels in TNBC (Ademuyiwa et al., 2012; Grigoriadis et al., 2009; Hamai et al., 2011; Rakha et al., 2007). In this study, we had both tissue microarray data and AAb data for TP53 for the same subjects. We found that nearly all of the patients who developed antibodies had elevated protein levels compared with the controls suggesting that elevated TP53 protein levels might be necessary for AAb responses. However, there were 38 patients with positive IHC score who did not develop TP53 antibodies, indicating that other factors, such as tumor microenvironment, antigen processing machinery, and genetic background, may also contribute to the development of AAbs (Vesely et al., 2011). For the remaining biomarkers reported in this study, except for RNF216 and PSRC1, we did not find any evidence of elevated mRNA levels in basal-like tumors compared with other subtypes using TCGA data, although this does not rule out the possibility that protein levels could be altered or that individual changes could have occurred at the single patient level. Detailed study comparing tumor antigen levels and AAb responses at single patient resolution would help clarify this puzzle.

There are inconclusive reports on whether specific *TP53* mutations were associated with AAb development (Davidoff et al., 1992; Soussi, 2000). We knew both the *TP53* mutation status and AAb responses to wildtype TP53 protein for a subset of 21

patients and did not find any relationship. Given that the number of patients used for this comparison is relatively small, additional studies are needed to fully address this association. Furthermore, AAbs could also rise against specific TP53 mutant protein, only by examining wildtype TP53 protein might not reflect the actual AAb responses.

In some reports, AAb responses predicted clinical outcomes in cancer patients (Gnjatic et al., 2010; Mange et al., 2012). We did a survival analysis by stratifying our patients into AAb responders and non-responders, and compared their overall survival. By doing this, we, for the first time, found that AAbs against TP53 and MN1 proteins are associated with a reduced prognosis in basal-like patients. Reduced survival of TP53 AAb-positive patients has been reported in studies of other cancers (Reuschenbach et al., 2009). In basal-like tumors, elevated TP53 protein levels have been associated with worse prognosis (Rakha et al., 2007). MN1 encodes meningioma 1, a probable tumor suppressor protein of unknown function. MN1 mRNA is a negative prognostic marker in acute myeloid leukemia (AML) (Langer et al., 2009; Metzeler et al., 2009). Moreover, low MN1 protein levels associate with better treatment response (Schwind et al., 2011). If elevated MN1 contributes to AAb development, this would be consistent with our observation of reduced overall survival in basal-like patients. In addition, Andreu et al. had also demonstrated that stromal accumulation of AAbs could promote neoplastic progression by activating Fcγ receptors on myeloid cells, providing evidence for AAb induced cancer progression (Andreu et al., 2010).

With further validation, these markers might contribute to the improved detection of basal-like breast cancer, an aggressive breast cancer subtype that afflicts younger women (Dent et al., 2007) who are not recommended for routine mammography

(Calonge et al., 2009). Given the poor prognosis of basal-like subtype, early intervention is likely to be beneficial. Blinded studies on independent cohorts are already underway through the Early Detection Research Network. The additional of independent markers and screening modalities would also contribute to a diagnostic test with better performance.

Aside from phosphorylation, these protein arrays displayed proteins with only limited post translational modifications (PTM). Distinct alterations in PTMs have been found between cancer and healthy controls (Arnold et al., 2011; Blixt et al., 2010; Whelan et al., 2009). Disease-specific AAbs targeting aberrant PTMs had also been reported previously (Tomaino et al., 2011; Wandall et al., 2010). While this limitation is not unique to our protein array platform, we are actively working to address this challenge and hope to profile antibodies against proteins with disease-relevant PTMs at the proteome level in the future. We are also working to expand our antigen repertoire from current 10,000 to cover even greater percentage of the proteome.

3.6 Conclusion

In summary, we have performed the largest proteomic screen using NAPPA technology and identified 13 AAb biomarkers associated with BLBC. With further validation, these markers might contribute to improved detection of BLBC, an aggressive subtype that afflicts younger women where mammography is less sensitive (Domingo et al., 2014; Foulkes et al., 2010; Kirsh et al., 2011). Our analysis of AAbs associated with BLBC represent promising markers for early detection for several reasons: 1) they may be detectable at early stages, 2) their sensitivity is not dependent on visualization, so the fact that young women have poorly imaged dense breasts is not a limitation, and 3) blood testing can be performed repeatedly without risk of radiation exposure or expensive techniques such as MRI, making this a good approach for young women and those who may require frequent testing. Future work in clinical and prospective observational studies is needed to determine the value of these markers for early detection, prognosis and response to treatment.

CHAPTER 4

COMPARATIVE STUDY OF AUTOANTIBODY RESPONSES BETWEEN LUNG ADENOCARCINOMA AND BENIGN PULMONARY NODULE

4.1 Abstract

The reduction in lung cancer mortality associated with CT screening has led to its increased use and a concomitant increase in the detection of benign pulmonary nodules. Many of these individuals undergo unnecessary, costly and invasive procedures. Therefore, there is a need for companion diagnostics that stratify individuals with pulmonary nodules into high risk or low risk groups. Lung cancers can trigger host immune responses and elicit antibodies against tumor antigens. The identification of these antibodies and their corresponding antigens may expand our knowledge on cancer immunity, leading to early diagnostics or even benefit immunotherapy. Previous studies were mostly performed in the context of comparing cancers and healthy (smoker) controls. We have performed one of the first studies in understanding humoral immune response in cancer patients, patients with benign nodules and healthy smokers.

We first profiled sero-reactivity to 10,000 full-length human proteins in 40 patients with early stage lung cancer and 40 smoker controls using nucleic acid programmable protein arrays (NAPPA) to identify candidate cancer specific AAbs. Seventeen antigens showing higher reactivity in lung cancer cases relative to controls were subsequently selected for evaluation in a large sample set (n=264) using enzyme-linked immunosorbance assay (ELISA). A 5-AAb classifier (TTC14, BRAF, ACTL6B, MORC2, CTAG1B) was developed that can differentiate lung cancers from smoker controls with a sensitivity of 30% at 89% specificity. We further tested AAb responses in

subjects with CT positive benign nodules (n=307), and developed a 5-AAb panel (KRT8, TTC14, KLF8, BRAF, TLK1) with a sensitivity of 30% at 88% specificity. Interestingly, mRNA levels of 6 AAb targets (TTC14, BRAF, MORC2, CTAG1B, KRT8, TLK1) were also found to increase in lung adenocarcinoma tissues based on TCGA data set.

We believe that these antibodies warrant future validation using a larger sample set and / or longitudinal samples individually or as a panel. They could potentially be part of companion molecular diagnostics modalities that will benefit subjects undergoing CT screening for lung cancer.

4.2 Introduction

Lung cancer has long been the leading cause of cancer deaths in the United States, with more than 150,000 deaths in year 2014 (Siegel et al., 2014). 5 year survival rate of lung cancer overall is only 17%, and 57% of lung cancers are diagnosed at advanced stage with 5 year survival rate as low as only 4%(Siegel et al., 2014). Currently, low-dose computed tomography (CT) scans are used to screen populations with extensive smoking history between 55 and 74 years old (Bach et al., 2012). CT scan had been proved to effectively reduce lung cancer mortality by 20%, but many from the millions of pulmonary nodules identified by CT remained undiagnosed as malignant or benign (Li et al., 2013). According to the National Lung Screening Trial (NLST), only 3.6% of the nodules detected by CT were confirmed to have lung cancer (Aberle et al., 2011), suggesting high false positive rate. Therefore, there is a need for diagnostic tests that differentiate malignant from benign nodules, improving the diagnostic performance when combined with CT screening. Practically, tests for such markers should rely on readily accessible samples, like plasma or sputum, because they are likely to be performed on individuals undergoing screening. Tremendous efforts have been spent on the identification of proteins, circulation tumor cells, circulating tumor DNAs, and circulating miRNA for this purpose (Hennessey et al., 2012; Hofman et al., 2011; Krebs et al., 2012).

The concentration of many molecular markers in blood tends to be very low because it relies upon secretion by cancer cells, which are few in number in the pre-clinical stage (Hori and Gambhir, 2011). Typically, only a fraction of the secreted biomarker gets distributed to the plasma where the biomarker gets diluted in a large

volume in blood (Anderson and Hunter, 2006; Lutz et al., 2008). An alternative strategy is to exploit the ability of the immune system to detect the presence of tumor cells through the generation of autoantibodies (AAb)(Anderson et al., 2011a; Goodell et al., 2006; Wang et al., 2005). These responses of the adaptive immune system against target tumor antigens effectively amplify the signals from the minute amount of tumor proteins released from cancer tissue (Anderson and LaBaer, 2005; Hanash et al., 2011). AAbs have been observed years before the clinical diagnosis of tumor (Lu et al., 2012; Qiu et al., 2008). Therefore, blood AAbs could serve as a good repertoire for early detection biomarker discovery.

Here, we reported one of first study focusing on comparing plasma AAb responses in lung adenocarcinoma (ADC) patients with heavy smoker subjects (SMC), as well as with benign nodules (BNC). We started with an unbiased screen for cancer-specific antibodies in patients with the adenocarcinoma subtype non-small cell lung cancer (NSCLC) and age, gender, smoking matched controls, using nucleic acid programmable protein arrays (NAPPA) displaying ~10,000 full length human proteins. Candidate lung cancer specific antibodies were further assessed in an independent set of cases and controls, including subjects with benign pulmonary nodules.

4.3 Materials and methods

4.3.1 Characteristics of plasma samples

A total of 434 plasma samples were obtained from NYU (Table 4-1) with 137 lung adenocarcinoma, 127 controls with smoking history, and 170 benign pulmonary nodules (granuloma, n=47; emphysema, n=50; stable nodules, n=73). In the discovery sample set for protein array experiment, 40 patients with lung adenocarcinoma were matched to 40 cancer free controls by age, gender and smoking history. 38 out of 40 patients from the discovery samples had stage I disease. For validation purpose, additional 97 patients with lung adenocarcinoma of different stages (47% stage I) and 87 controls as well as 170 patients with CT positive benign lung disease were included.

4.3.2 Protein array experiments

Open reading frames were obtained from DNASU (https://dnasu.org/). Production of the protein array and array quality control experiments were performed as previously described (Festa et al., 2013; Wang et al., 2013). In brief, arrays displaying 10,000 human proteins (distributed evenly on five array sets) were manufactured. A common control plasma sample was repeated in every experiment to assess reproducibility. Consistency among experiments was determined with scatter plots comparing spot intensity measurements of the same plasma sample tested on different experiments.

4.3.3 Protein array image analysis and quantification

The scanned protein array images were examined using ArrayPro Analyzer (MediaCybernetics). To capture real antibody responses that cannot be quantified by the image analysis software, two researchers qualitatively examined all images to identify and confirm positive responses, which were described previously (Montor et al., 2009).

Briefly, raw images were adjusted to extreme contrast and brightness using ArrayPro Analyzer (MediaCybernetics), and each spot was graded at a scale of 0 to 5 based on ring's intensity and morphology.

Table 4-1. Sample information

Characteristics	Discovery			Validation				
	LC	**SC**	***P***	**LC**	**SC**	***P***	**BC**	***P***
Subjects	**40**	**40**		**97**	**87**		**170**	
Age (years)	74 (66-77)	70 (64-76)	0.3355	70 (62-77)	63 (57-70)	0.0002	65 (59-71)	0.0008
Gender			1			1		0.6076
Male	19 (48)	19 (48)		39 (40)	35 (40)		74 (44)	
Female	21 (52)	21 (52)		58 (60)	51 (59)		95 (56)	
No data	0	0		0	1 (1)		1 (<1)	
Smoking history								
Status			0.5542			0.1087		0.0329
Never	0	0		11	4		11	
Former	28	31		63	60		102	
Current	5	9		15	22		49	
No data	7	0		8	1		8	
Pack-year	29 (7-45)	26 (8-46)	0.9115	29 (10-45)	30 (19-46)	0.4032	29 (17-45)	0.4469
Nodules								<0.0001
Size (cm)	1.5 (1.3-1.8)	–		2.3 (1.4-3.2)	–		0.4 (0.6-1.0)	
Stage								
IA	37	–		31	–		–	
IB	1	–		15	–		–	
IIA	2	–		24	–		–	
IIB	0	–		3	–		–	
IIIA	0	–		20	–		–	
IIIB	0	–		4	–		–	

4.3.4 Candidate selection

Protein antigens were selected for subsequent ELISA confirmation when they showed higher prevalence in lung adenocarcinoma based on visual analysis. Specifically, they had to meet all of the following criteria: 1). Their frequency in ADC minus

frequency in SMC is greater than or equal to 2; 2). Frequency in ADC divided by frequency in SMC is greater than or equal to 1.4. Totally, 57 protein antigens were selected.

4.3.5 Pathway Analysis

Gene Ontology term enrichment analysis was performed using Cytoscape with ClueGo plugins on all 57 proteins with customized reference of all proteins displayed on our protein array. Gene symbol was used as identifier for the analysis. Node size was set proportional to number of genes observed. Node color was coded to reflect Benjamini-Hochberg adjusted p value.

4.3.6 ELISA assays

ELISA assays were performed to verify selected AAb responses towards protein antigens using freshly produced human proteins as previously described (Ramachandran et al., 2008a). In brief, 96-well highbind ELISA plates (Corning) were coated with goat anti-GST antibody (GE Healthcare) at 10 μg/ml in 0.2 M sodium bicarbonate buffer pH9.4 overnight at 4°C 1 day prior to experiment. All high-throughput liquid handling were performed using a BioMek NxP Laboratory Automation Workstation (Beckman Coulter). See Supplementary Materials and Methods for additional details.

4.3.7 Statistics and Data Analysis

To combine AAbs into panels, we used the 98 percentiles of the relative absorbance of either smoker control subjects (Panel I) or benign control subjects (Panel II) as cutoffs. A sample is called positive for lung adenocarcinoma if the AAb responses to one of the panel candidates exceed its corresponding cutoff.

A heatmap was developed to display differential AAb responses of 17 selected targets in lung cancer patients and smoker controls using the confirmation ELISA results. The heatmap color was scaled according to each AAb, and constructed using the gplots package in R.

To determine AAb responders from ELISA analysis, we categorized subjects as responders to specific antigens of interest using 98-percentile cutoff using data from benign subjects. This method was used to determine the association of AAb responses with cancer stage, age, nodule size and smoking history.

A multivariate logistic regression model was constructed to study the association of AAb responses to age, nodule size and smoking history. To adjust for lung cancer status and better assess the relationship between tumor size and AAb responses, we constructed another model with lung cancer status as an additional variable. We then applied the same method to analyze the association of AAb responses to tumor size, node status, and tumor stage among lung cancer patients.

To compare TCGA mRNA expression levels between lung adenocarcinoma and normal tissues, we used one-sided Welch's t test. The TCGA lung adenocarcinoma data were generated by Illumina HiSeq, and obtained from UC Santa Cruz Cancer Genome Browser (https://genome-cancer.ucsc.edu/) TCGA_LUNG_exp_HiSeqV2-2014-08-22. All intensities were normalized by subtracting the mean value of each mRNA from each sample.

4.4 Results

4.4.1 Identification of candidate AAbs associated with lung adenocarcinoma

To identify lung adenocarcinoma-associated candidate AAbs, we first performed comprehensive profiling of antibodies against 10,000 full-length human proteins in plasma samples from 40 patients with lung adenocarcinoma and 40 heavy smoker controls on NAPPA. Based on the array data, we selected 57 antigens whose AAb responses were differentially presented in lung cancer patients compared to smoker controls (Figure 4-1, materials and methods). A gene ontology enrichment analysis of these 57 candidate AAb targets revealed their involvement in embryonic morphogenesis, organ development, kinase signaling, and intermediate filament cytoskeleton (Figure 4-2A). We then assessed these selected candidates by ELISA using the same samples. Based on ELISA, 17 antigens were confirmed to elicit differential AAb responses in lung cancer patients, and included for subsequent analysis (Figure 4-2B).

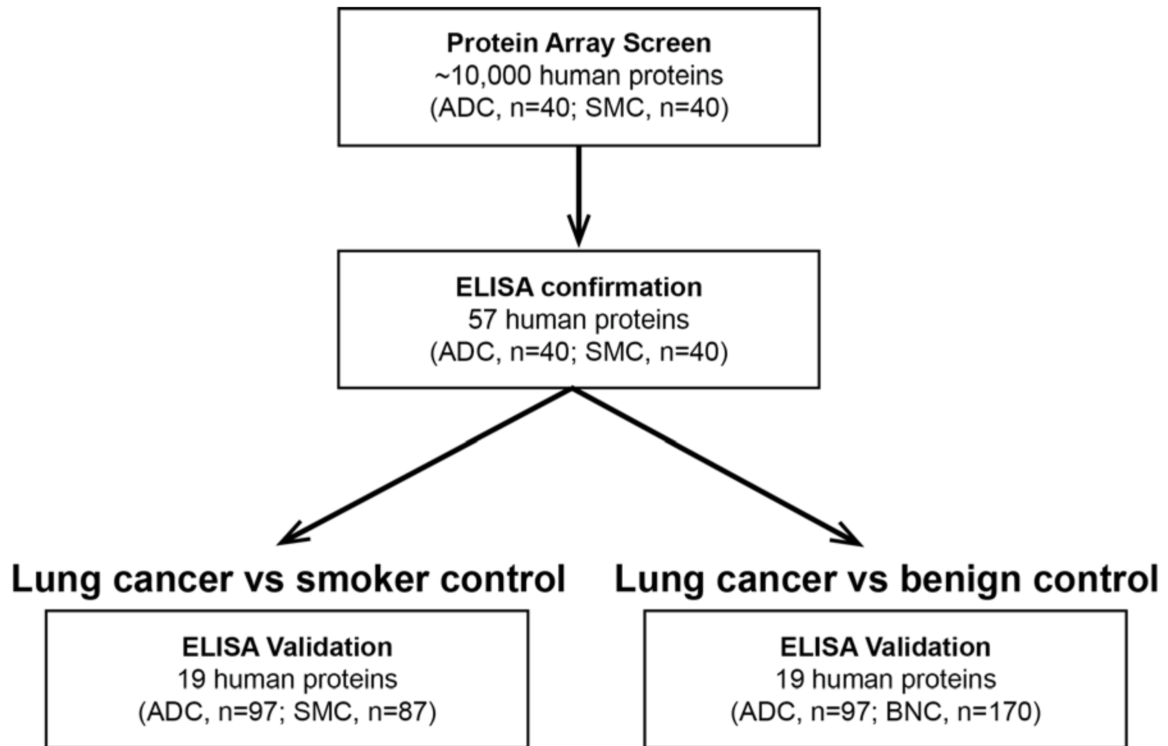
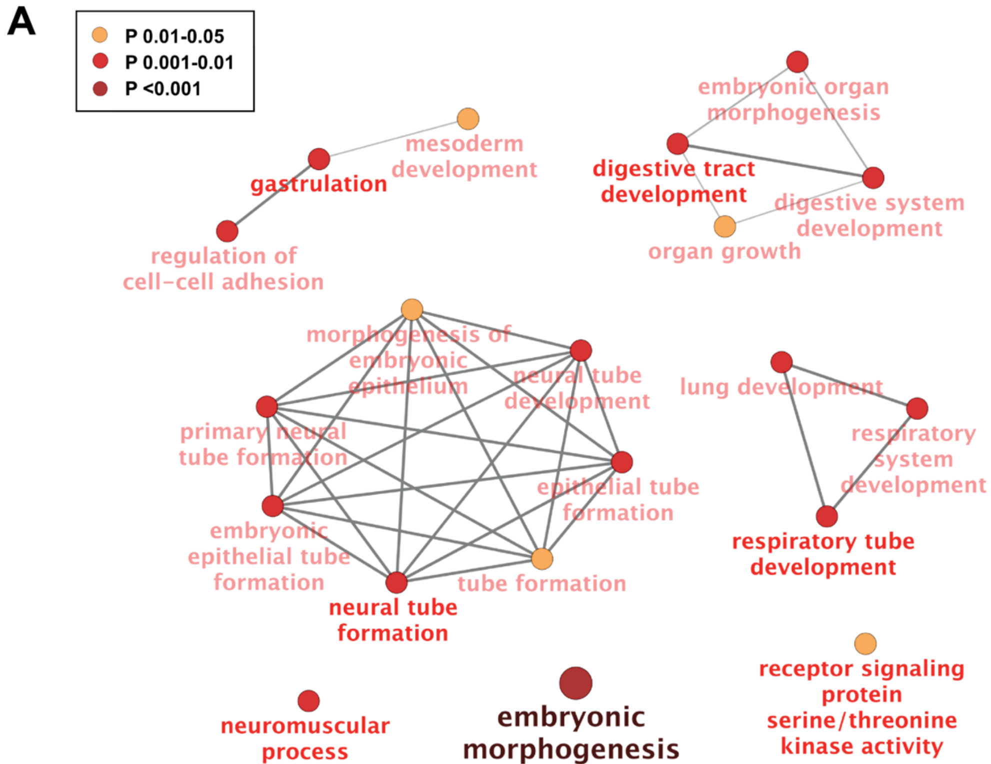


Figure 4-1. Overview of Study Design.



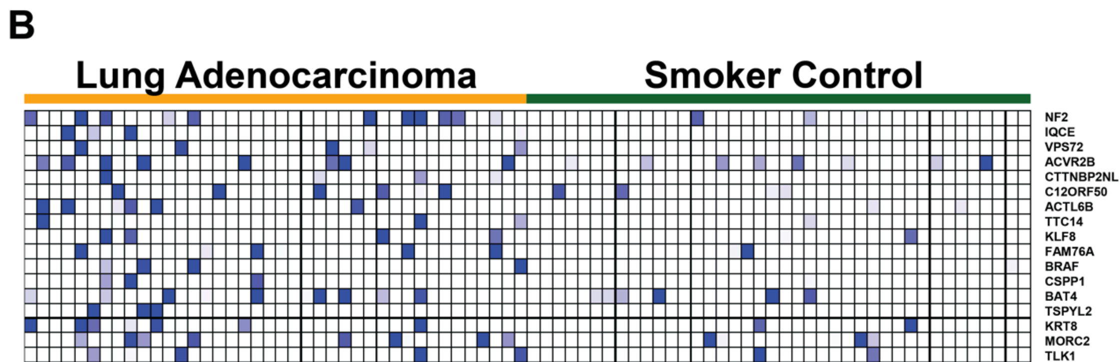


Figure 4-2. Summary of lung cancer associated antigens discovered from protein array screening. **A.** GO enrichment analysis of lung cancer associated AAb targets selected from protein array screening. Term embryonic morphogenesis has 20 genes, whereas the rest of the node has 5 genes each. **B.** Heatmap of differential AAb responses in lung cancer compared to smoker controls.

4.4.2 Validation in lung cancer patients versus healthy smoker controls

To verify the levels of these 17 AAbs in lung cancer patients, we measured these AAbs in 184 additional plasma samples from 97 cases and 87 controls. In addition, we also included TP53 and CTAG1B proteins as possible candidates according to previous

publications (Lam et al., 2011). Sero-positivity cutoffs of individual AAbs were set at 98 percentile of the ELISA absorbance in the 87 control samples. AAbs to TTC14, BRAF, and CTAG1B had sensitivity above 5% at 98% specificity when comparing lung cancer patients with smoker controls (Table 4-2). In addition, Sensitivities of AAbs to TTC14, BRAF, ACTL6B, MORC2 and CTAG1B were above 5% at 98% specificity in the entire sample set (Table 4-3). Further analysis of these 5 antigens using a standardized cutoff for each antigens of a relative absorbance greater than or equal to the 98 percentile of the relative absorbance in smoker controls, revealed a 5-AAb panel (panel I) with 30% sensitivity and 89% specificity.

Table 4-2. Discovery and validation statistics of selected AAbs.

Antigen	Discovery (ADC, n=40; SMC, n=40)		Validation (ADC, n=97; SMC, n=87)		Validation (ADC, n=97; BNC, n=170)	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
TTC14	17.5%	97.5%	11.3%	97.7%	5.2%	97.6%
VPS72	17.5%	97.5%	0.0%	97.7%	2.1%	97.6%
CTTNBP2NL	15.0%	97.5%	3.1%	97.7%	2.1%	97.6%
TSPYL2	15.0%	97.5%	2.1%	97.7%	1.0%	97.6%
ACTL6B	15.0%	97.5%	3.1%	97.7%	2.1%	97.6%
ACVR2B	15.0%	97.5%	4.1%	97.7%	2.1%	97.6%
BRAF	12.5%	97.5%	5.2%	97.7%	6.2%	97.6%
KLF8	12.5%	97.5%	1.0%	97.7%	5.2%	97.6%
BAT4	12.5%	97.5%	0.0%	97.7%	0.0%	97.6%
C12ORF50	10.0%	97.5%	2.1%	97.7%	2.1%	97.6%
IQCE	10.0%	97.5%	4.1%	97.7%	4.1%	97.6%
CSPP1	7.5%	97.5%	1.0%	97.7%	0.0%	97.6%
KRT8	7.5%	97.5%	0.0%	97.7%	7.2%	97.6%
MORC2	7.5%	97.5%	4.1%	97.7%	1.0%	97.6%
FAM76A	7.5%	97.5%	1.0%	97.7%	2.1%	97.6%
NF2	5.0%	97.5%	2.1%	97.7%	2.1%	97.6%
TLK1	5.0%	97.5%	4.1%	97.7%	6.2%	97.6%
TP53	2.5%	97.5%	3.1%	97.7%	4.1%	97.6%
CTAG1B	2.5%	97.5%	9.3%	97.7%	3.1%	97.6%

4.4.3 Classification of lung cancer versus benign controls

To test whether these 17 AAb together with AAb against TP53 and CTAG1B can differentiate lung cancer from benign disease identified by CT screening, we analyzed the AAb responses against these antigens from 267 plasma samples by ELISA. As above, cutoffs of individual AAb were set at 98 percentile of the relative absorbance in benign controls. KRT8, TTC14, KLF8, BRAF, TLK1 were confirmed for their association to lung cancer patients compared with benign controls (Table 4-2). They also had overall sensitivity above 5% at 98% specificity (Figure 4-3; Table 4-4). Further analysis of these 5 antigens using a standardized cutoff for each antigens of a relative absorbance greater than or equal to the 98 percentile of the relative absorbance in benign controls, revealed a 5-AAb panel (panel II) with 30% sensitivity and 88% specificity. Sensitivities of individual AAb using subjects with different benign lung nodules were also assessed (Table 4-5).

Table 4-3. Sensitivity and specificity of individual AAb from panel I.

Antigen	Overall (LC, n=137; SC, n=127)	
	Sensitivity	**Specificity**
TTC14	12.4%	97.6%
BRAF	8.0%	97.6%
ACTL6B	5.1%	97.6%
CTAG1B	5.1%	97.6%
MORC2	5.1%	97.6%

Table 4-4. Sensitivity and specificity of individual AAb from panel II.

Antigen	Overall (LC, n=137; BC, n=170)	
	Sensitivity	**Specificity**
KRT8	8.8%	97.6%
TTC14	8.0%	97.6%
KLF8	7.3%	97.6%
BRAF	6.6%	97.6%
TLK1	5.8%	97.6%

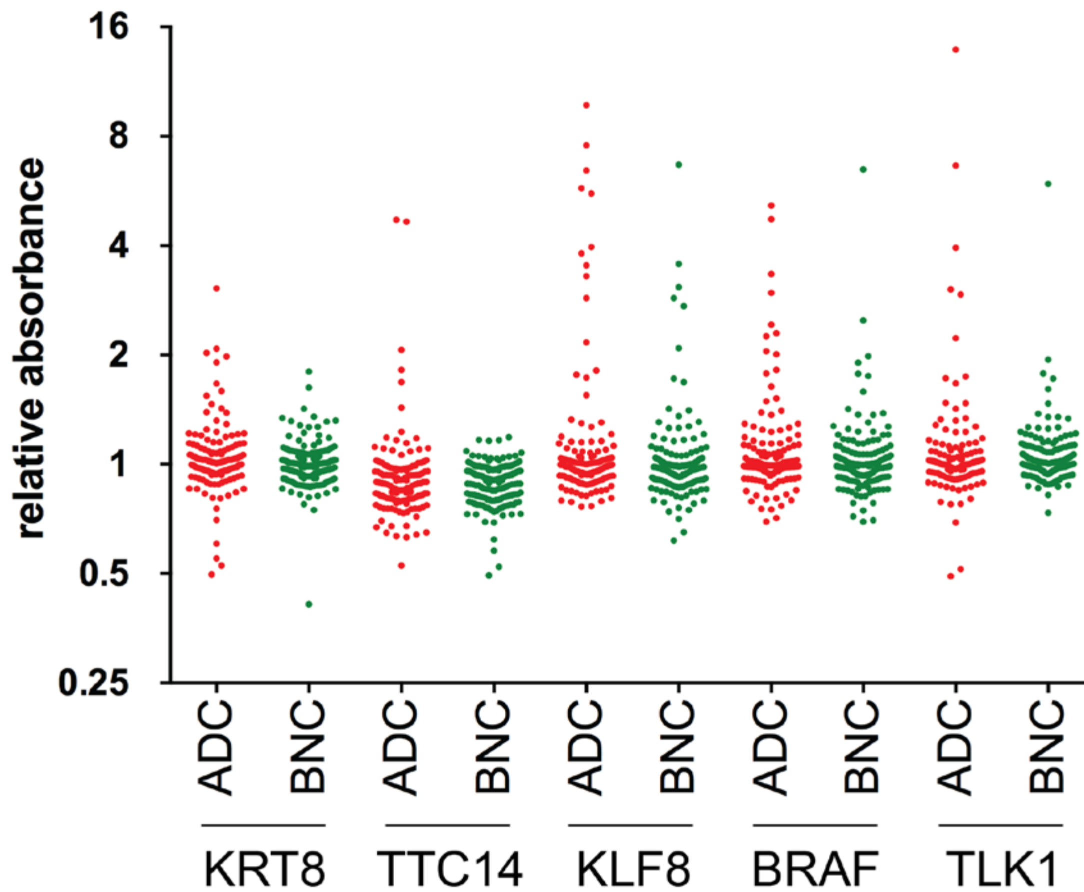


Figure 4-3. AAb responses of individual AAb from panel II.

4.4.4 Effect of patient and disease characteristics on AAb positivity

We compared clinical risk factors of smoking history, tumor size and age to the AAb responses. Using the defined cutoff derived from panel II, there are no significant differences in age (P=0.677) or smoking history (P=0.718) between AAb responders and nonresponders (Figure 4A). Whereas tumor size was found to be significantly associated with AAb responses (P=0.036), but when adjusted for disease status, it is no longer significant (P=0.858). This result indicates that the AAb panels provide additional

Table 4-5. Sensitivity of individual AAb using various benign lung nodules as control samples. Specificity was set at 98%.

Antigen	ADC vs Granuloma	ADC vs emphysema	ADC vs stable nodule
TTC14	7.3%	13.1%	12.4%
BRAF	10.2%	5.8%	8.0%
ACTL6B	2.2%	4.4%	2.9%
CTAG1B	5.1%	2.9%	2.9%
MORC2	2.2%	2.2%	0.7%
CTTNBP2NL	1.5%	12.4%	2.2%
VPS72	5.8%	5.8%	2.2%
IQCE	4.4%	4.4%	4.4%
TLK1	12.4%	4.4%	6.6%
ACVR2B	2.2%	5.1%	1.5%
TSPYL2	8.8%	2.9%	6.6%
TP53	6.6%	1.5%	4.4%
CSPP1	4.4%	2.9%	2.2%
C12ORF50	1.5%	1.5%	1.5%
NF2	1.5%	2.2%	9.5%
KLF8	10.9%	10.9%	7.3%
FAM76A	6.6%	4.4%	5.8%
BAT4	8.8%	1.5%	4.4%
KRT8	9.5%	4.4%	10.2%

information on lung cancer status and that the observed AAb responses are independent of the two known risk factors. In addition, we also analyzed the AAb responses among lung cancer patients and compared that with patients' characteristics including smoking

history, tumor size, node status, and tumor stage. No significant association was observed between these factors and AAb responses (Figure 4-4B). To assess this at individual AAb level, we observed that TTC14 AAb has higher prevalence in stage I lung cancer, whereas AAb against BRAF has higher prevalence in stage II and III (Table 4-6, Table 4-7).

Table 4-6. Positivity of individual autoantibody from panel I by stage.

Stage	Number of Samples	TTC14 Positive(%)	BRAF Positive(%)	ACTL6B Positive(%)	CTAG1B Positive(%)	MORC2 Positive(%)	Panel I Positive(%)
I	84	15.5%	6.0%	6.0%	6.0%	6.0%	32.1%
II	29	3.4%	13.8%	10.3%	3.4%	3.4%	20.7%
III	24	8.3%	12.5%	8.3%	8.3%	8.3%	33.3%

Table 4-7. Positivity of individual autoantibody from panel II by stage.

Stage	Number of Samples	KRT8 Positive(%)	TTC14 Positive(%)	KLF8 Positive(%)	BRAF Positive(%)	TLK1 Positive(%)	Panel II Positive(%)
I	84	9.5%	10.7%	7.1%	4.8%	6.0%	32.1%
II	29	3.4%	3.4%	3.4%	10.3%	3.4%	17.2%
III	24	12.5%	4.2%	12.5%	8.3%	8.3%	37.5%

4.4.5 Correlation of AAb targets and their mRNA level

We further investigated the tissue mRNA levels of protein antigens from both panels using TCGA data (Collisson et al., 2014). 6 out of 8 proteins showed significantly increased expression in lung adenocarcinoma tissues compared to normal tissue (Figure 4-5). This orthogonal analysis confirmed our discovery of these AAbs' association with lung adenocarcinoma.

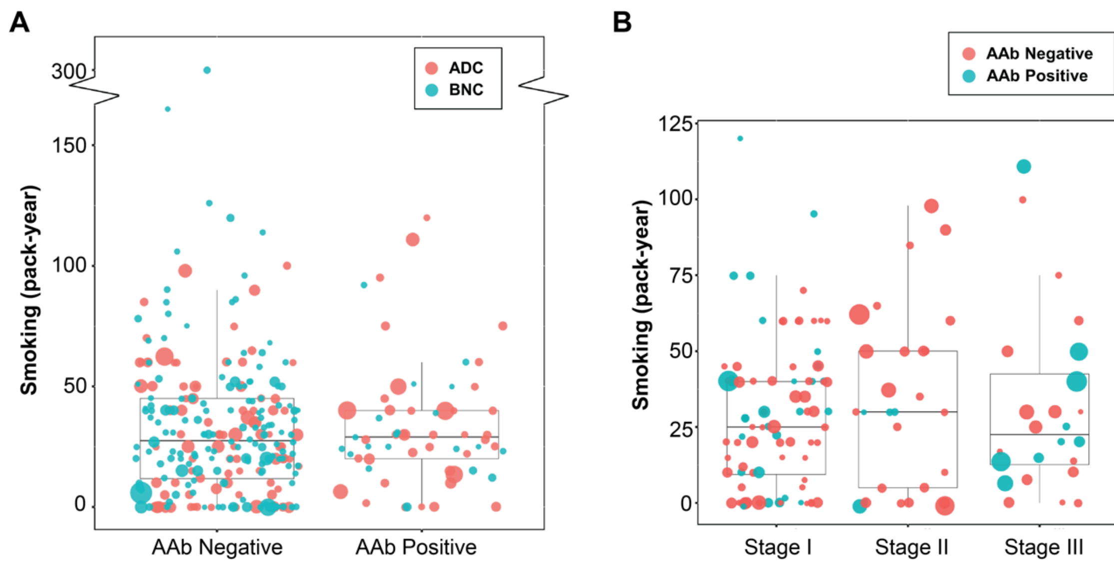


Figure 4-4. Multivariate analysis of clinical factors and AAb responses. **A.** Analysis of smoking history, AAb responses and nodule size in lung cancer and benign control. **B.** Analysis of smoking history, stage and AAb responses in lung cancer cases. (A and B, Smoking is measured by pack-year on the vertical axis. Nodule size is presented by circle diameter.)

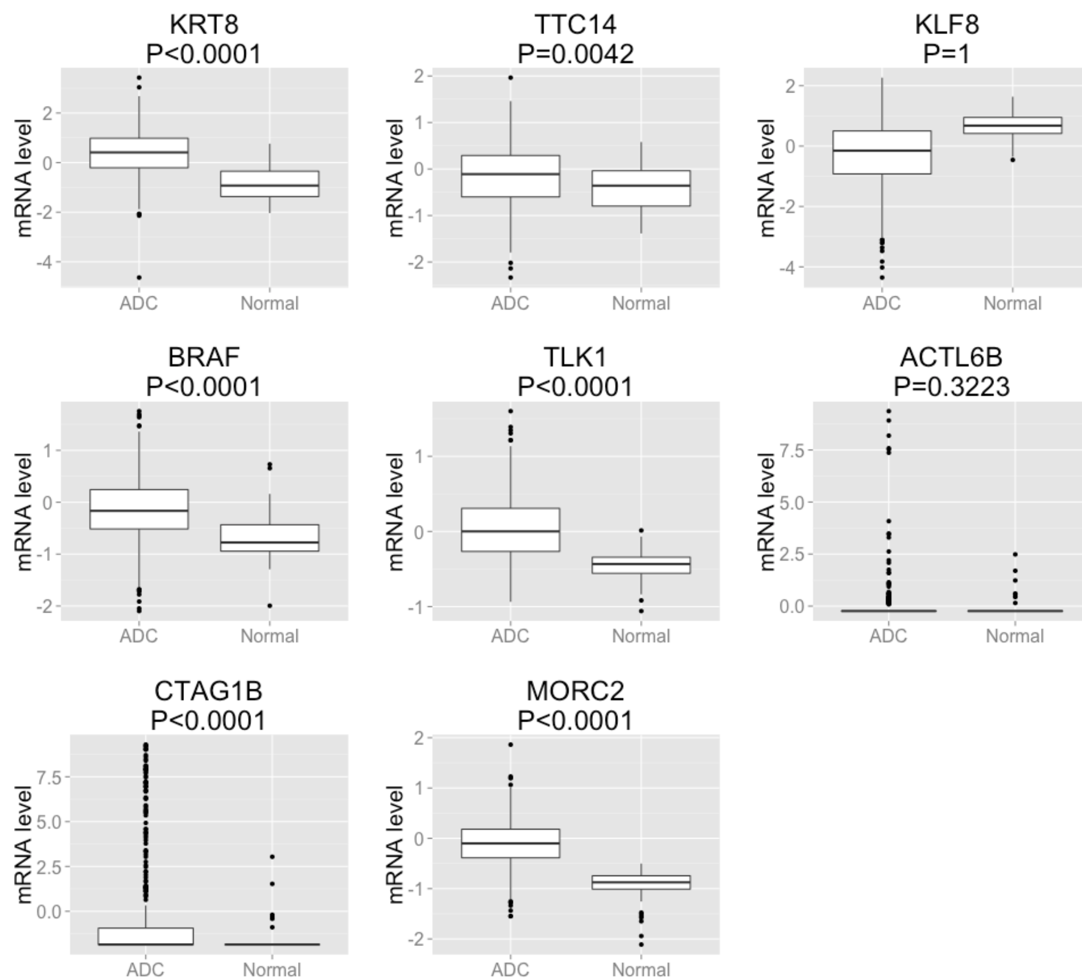


Figure 4-5. mRNA expression level of AAb targets from both panels (TCGA). Only mRNA levels in lung adenocarcinoma (ADC) and solid tissue normal (normal) were graphed.

4.5 Discussion

Using an immuno-proteomics approach, we profiled antibody responses in healthy heavy smoker controls and lung cancer patients. Bioinformatics analysis revealed significantly enriched pathways related to embryonic morphogenesis, organ development (including lung development) and receptor signaling and serine/threonine kinase pathways. The performance of a subset of antibodies was confirmed by ELISA using an expanded sample set including subjects with benign nodules, with sensitivities ranging from 5-10% at 98% specificity. We reported a 5-AAb panel (TTC14, BRAF, ACTL6B, MORC2, CTAG1B) that has 30% sensitivity at 89% specificity to distinguish lung cancer from high risk controls with smoking histories. A comparison of AAb responses between lung cancer and patients with CT positive pulmonary nodules revealed a related but different 5-AAb panel (TTC14, BRAF, KLF8, TLK1, KRT8) with a sensitivity of 30% at 88% specificity. Although these panels require further validation, they do provide information on the complementarities of these informative antigens. Further analysis revealed that AAbs do not associate with lung cancer stage. To our knowledge, this is one of the first studies that applied an immuno-proteomics approach on the identification of specific antibodies that might help stratify subjects with positive CT nodules into benign lung disease controls and lung adenocarcinoma patients. To ensure accurate estimations of responses when analyzing ELISA results, we also estimated the background associated with the supporting reagent for each plasma sample, which provides the most rigorous assay in similar studies.

Overall, our results are comparable with several validation studies reported by Chapman et al.(Boyle et al., 2011; Lam et al., 2011; Macdonald et al., 2012). The early

CDT-lung assay could differentiate patients with various subtypes of lung cancer from healthy controls at 49% sensitivity and 93% specificity (Macdonald et al., 2012), while sensitivity of adenocarcinoma alone was 25.6% at 88% specificity (Lam et al., 2011). Our previous study demonstrated AAb responses to Annexin I, 14-3-3 Theta and LAMR1 with a performance of 51% sensitivity at 82% specificity in lung cancer patients prior to the onset of symptoms and diagnosis (Qiu et al., 2008). Among the few studies that examined patients with CT positive pulmonary nodules, using multiple reaction monitoring mass spectrometry assay, Li, et al. reported a 13-protein classifier, which when coupled with CT scan, can improve negative predicted value, but independent validation presented a less attractive sensitivity and specificity (Li et al., 2013).

There are reports of marker panels with higher sensitivities for lung cancer; however, these studies have used less stringent study design and limited sample numbers. For example, Farlow et al. reported a biomarker panel that can differentiate NSCLC from controls at 88% sensitivity and 87% specificity (Farlow et al., 2010), and in a subsequent paper by the same group, they improved the diagnostic performance to 94% sensitivity and 97% specificity in the same patient cohort by adding several AAb markers(Murray et al., 2010). However, they used a limited number of control samples in the former study without any information on smoking history and no information on the controls was included on the follow up study. Wu et al. reported a phage-peptide detector that had over 92% sensitivity and 92% specificity for NSCLC (Wu et al., 2010). Patients involved in this study mainly had stage III/IV lung cancer, and control subjects were not matched by smoking, weakening their use as early detection markers. Moreover, all of the reported biomarkers above require further independent validation in larger sample sets.

In addition to the 17 AAbs identified in the protein array screening, we also evaluated two core AAbs (TP53, CTAG1B) reported in previous studies with 10-15% sensitivity in lung cancer (Lam et al., 2011; Macdonald et al., 2012; Shan et al., 2013). However, we did not observe specific AAb responses to TP53 in our patient cohort, although we have readily detected TP53 AAb responses in basal-like breast cancer and ovarian cancer (data not shown). CTAG1B AAb was among panel I (vs. smoker controls), but not panel II (vs. benign lung disease controls). Specifically, very few studies addressed AAb responses to CTAG1B and TP53 in early stage adenocarcinoma with benign lung disease as control samples. According to a previous report, TP53 AAb frequency appeared to be higher in squamous cell lung cancer and small cell lung cancer than that in adenocarcinoma, where it increased along with cancer stage (Lam et al., 2011). It is also possible that the lower AAb responses could be due to the effect of smoking as suggested (Palmer et al., 2005).

We were interested to find the association of lung adenocarcinoma with AAb against BRAF. The involvement of BRAF in progression of different types of cancer has been well studied (Davies et al., 2002; Pao and Girard, 2011). BRAF encodes a serine-threonine kinase that transduces regulatory signals from RAS to MAPK in the growth factor receptor signaling pathway (Davies et al., 2002). AAbs against BRAF protein have been reported in melanoma patients (Fensterle et al., 2004). Patients with Rheumatoid Arthritis also found to have BRAF AAb, but the association is under debate (Charpin et al., 2010; Li et al., 2011). It is also mutated in 10% of lung adenocarcinoma (Collisson et al., 2014). It is possible that BRAF's over expression in lung adenocarcinoma contributed to both its AAb level and advance of the disease. Unfortunately, tumor tissue

corresponding to the patient samples with BRAF AAb was not available to test this. Mutation in BRAF might also be a cause of its humoral immune responses. However, further study comparing its protein expression, mutation and AAb responses in the same patient is required to clarify these projections.

ACTL6B belongs to the chromatin remodeling brain-specific BAF (bBAF) complex (Olave et al., 2002). It acts as a transcription coactivator during postmitotic neural development and dendritic outgrowth (Yoo et al., 2009). KLF8 is a member of the KLF family of transcription factors. A recent study reported its role in promoting breast cancer metastasis and invasion through activating MMP9 and MMP14 (Lu et al., 2014; Wang et al., 2011). KRT8 has been reported to elicit AAb responses in lung cancer with a sensitivity of 4% and a specificity of 100% (Macdonald et al., 2012). Here, we have also found that AAbs against KRT8 protein are elevated in lung cancer patients. And further, our result indicated that this AAb is associated with lung cancer in population with pulmonary nodules with a sensitivity of 8.6% at 98% specificity. Mutations in MORC2 were found in colorectal cancer patients (Tuupanen et al., 2014). TTC14 is a protein that contains tetratricopeptide repeat domain, which was known as a scaffold protein that mediates protein-protein interaction (Allan and Ratajczak, 2011). TLK1 encodes a nuclear serine/threonine kinase, which involves in chromatin assembly during S phase of cell cycle (Li et al., 2007).

It is still not clear what factors determine the development of these humoral immune responses. Assuming AAb responses were linked to tissue overexpression of its protein target, only a small fraction of patients with the overexpressed protein will develop AAb responses at detectable level. We also examined the mRNA level of these

AAb targets in TCGA lung adenocarcinoma data set (Collisson et al., 2014). mRNA expressions of 6 out of the 8 proteins in both panels were significantly increased in lung adenocarcinoma tissues. This finding not only suggested that the development of AAbs in lung adenocarcinoma might be a result of protein overexpression, but also orthogonally verified the association of these AAbs to lung adenocarcinoma.

Strengths of this study include the use of a large number of plasma samples from adenocarcinoma of NSCLC with primarily stage I disease matched with smoker controls as well as controls with CT positive benign lung disease. We also used highly reproducible protein arrays for high-throughput screening of AAb candidates, which revealed informative pathways related to developmental processes and kinase signaling. To evaluate these AAbs' performance, we used more clinically relevant ELISA assays in large sample sets. Our results were also consistent with TCGA mRNA expression data.

4.6 Conclusion

In summary, we have performed an immune-proteomic screening of AAb responses using protein arrays, and identified two panels of AAb that can potentially differentiate lung adenocarcinoma from smoker controls as well as CT positive benign lung disease. BRAF, as a putative oncogene, was also found to elicit humoral immune responses in lung cancer patients. For this study, we focused on markers with high specificity so that high risk subjects with a positive CT screen and a positive serum test should get more invasive test such as needle biopsy for their timely cancer diagnosis. On the other hand, autoantibody markers with high sensitivity in CT positive population will help reduce false negatives to prevent the wrongly exclusion of subjects that have the disease, which will in turn reducing false positive diagnostics from CT screening. To achieve this in the future, we will be focusing on combining existing markers with our panel to achieve high sensitivity. Future studies using mutated antigens or post-translationally modified antigens may benefit the identification of better performance markers. The use of longitudinal samples to track antibody changes may also improve marker performance. We believe an integrated panel of markers of different molecular types reflecting the physiological state changes from benign lung diseases / healthy smokers to lung cancers will be necessary to make an impact on the reduction of CT false positivity.

CHAPTER 5

SUMMARY AND FUTURE DIRECTIONS

5.1 Summary

The work presented in the preceding chapters described the method development and optimization of protein microarray technology, and subsequently applying the method to discover autoantibody biomarkers in basal-like breast cancer and lung cancer. The conclusion and contribution for each project is summarized here.

The research described in chapter 2 sought to develop a protein microarray platform that is capable to display proteins in either their native or denatured conformation, allowing the binding of antibodies against either conformational or linear epitiopes. We first optimized our traditional plasma screening protocol to reduce the background binding of immunoglobulin. Switching from the old rabbit reticulocyte lysate to Hela cell lysate in vitro transcription and translation system, we demonstrated a higher antibody signal, as well as the ability to detect more antibody responses from plasma sample. We then successfully showed that by adapting the covalent linkage between HaloTag ligand and HaloTag to replace the traditional NAPPA chemistry, the protein products can endure harsh treatment of detergent and reducing agent without getting dissociated from the slide, indicating successful covalent immobilization. After probing a set of human plasma samples on either denatured protein array or native array, we also observed distinct antibody responses. This work presented the first protein microarray that is versatile to endure detergent wash. As a prove of concept, it can be used for antibody detection.

Chapter 3 described the largest proteomic scale screening of autoantibodies that are associated with basal-like breast cancer, an aggressive subtype of breast cancer that is often missed by routine mammography. The autoantibody profile of 45 BLBC cases and 45 controls revealed strong association between B cell antigenicity and protein properties, such as lower fraction of helice, lower aromaticity, and higher fraction of turns. Cellular localization analysis of the autoantigens revealed enrichment in nucleus, centrosome, and depletion in endoplasmic reticulum membrane, extracellular region. From the screening and subsequent verification, 26 antigens were select for blind validation. AAb against CTAG1B, CTAG2, TP53, WBP2NL, PPHLN1, DOK2 were confirmed in independent sample set. A combined analysis showed a 13-AAb classifier that can differentiate BLBC from controls with a sensitivity of 33% and a specificity of 98%. In addition, we also observed association of AAb responses with protein expression and survival.

AAb to CTAG1B, CTAG2 and TP53 have also gone on to a national validation trial of 80 candidate markers from 15 different research groups in their ability to detect triple negative breast cancer from matched controls. They are the only three markers that are validated in that study. In combination with CA125, these markers demonstrated a sensitivity of 35% at 95% specificity. At this cutoff level, women who test positive are seven times more likely to have BLBC and would benefit from mammography. Advanced studies to test this in longitudinal/pre-diagnostic samples are now underway.

In chapter 4, we carried out another large scale immunoproteomic screening of AAbs associated with lung adenocarcinoma. Using matched smoker controls as well as subjects with benign lung nodules detected by CT, we discover and verified the AAb responses that can differentiate lung cancer from each control group. Specifically, panel I

(TTC14, BRAF, ACTL6B, MORC2, CTAG1B) classified lung adenocarcinoma from smoker controls with 30% sensitivity and 89% specificity; whereas panel II (TTC14, BRAF, KLF8, TLK1, KRT8) classified lung adenocarcinoma from benign nodules with 30% sensitivity and 88% specificity. Further analysis of their mRNA levels using TCGA data revealed elevated expression of TTC14, BRAF, MORC2, CTAG1B, TLK1, KRT8 gene in lung adenocarcinoma, which suggested the contribution to their AAb responses.

5.2 Future Directions

The success in developing the versatile protein microarray using HaloTag technology enabled the potential to profile antibody responses against both conformational and linear epitopes. However, we haven't applied it to perform any large scale biomarker discovery. There is still more work needs to be done to analyze the differences of AAb profiles systematically between denatured and native protein microarrays using a larger sample size and determine its utility in identifying AAb biomarkers with better clinical performance. We believe that this method can easily extend beyond AAbs detection in cancers to automimmune and infectious diseases. Moreover, covalent attachment of proteins on the matrix for denaturation can also be adapted to other types of protein microarrays. These findings had motivated our lab to switch the capturing chemistry from traditional antibody dependent to HaloTag technology by transferring all human genes to the HaloTag vector.

The results in chapter 3 and 4 showed the discovery of several novel autoantibodies associated with cancer. A limitation of both studies is, although we screened for proteins encoded by ~50% of the human genome, these arrays do not display many proteins with post translational modifications that might also be important AAb

targets for distinguishing cases from controls (Blixt et al., 2010; Tomaino et al., 2011; Wandall et al., 2010; Whelan et al., 2009). Moreover, given that we performed this analysis in case-control study design, it is unclear how early these markers are present with respect to clinical diagnosis. According to the 5 phase biomarker development guideline (Pepe et al., 2001; Surinova et al., 2011). future studies evaluating these markers in longitudinal samples to track antibody changes may also improve marker performance. In addition, prospective cohorts are needed to determine the value of these markers for early detection, prognosis and response to treatment.

REFERENCES

Abd Hamid, U.M., Royle, L., Saldova, R., Radcliffe, C.M., Harvey, D.J., Storr, S.J., Pardo, M., Antrobus, R., Chapman, C.J., Zitzmann, N., *et al.* (2008). A strategy to reveal potential glycan markers from serum glycoproteins associated with breast cancer progression. Glycobiology *18*, 1105-1118.

Aberle, D.R., Adams, A.M., Berg, C.D., Black, W.C., Clapp, J.D., Fagerstrom, R.M., Gareen, I.F., Gatsonis, C., Marcus, P.M., Sicks, J.D., *et al.* (2011). Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. New England Journal of Medicine *365*, 395-409.

Ademuyiwa, F.O., Bshara, W., Attwood, K., Morrison, C., Edge, S.B., Ambrosone, C.B., O'Connor, T.L., Levine, E.G., Miliotto, A., Ritter, E., *et al.* (2012). NY-ESO-1 Cancer Testis Antigen Demonstrates High Immunogenicity in Triple Negative Breast Cancer. Plos One *7*, 9.

Allan, R.K., and Ratajczak, T. (2011). Versatile TPR domains accommodate different modes of target protein recognition and function. Cell Stress & Chaperones *16*, 353-367.

Anderson, K.S., Cramer, D.W., Sibani, S., Wallstrom, G., Wong, J., Park, J., Qiu, J., Vitonis, A., and LaBaer, J. (2015). Autoantibody Signature for the Serologic Detection of Ovarian Cancer. Journal of Proteome Research *14*, 578-586.

Anderson, K.S., and LaBaer, J. (2005). The sentinel within: Exploiting the immune system for cancer biomarkers. Journal of Proteome Research *4*, 1123-1133.

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

Anderson, K.S., Sibani, S., Wallstrom, G., Qiu, J., Mendoza, E.A., Raphael, J., Hainsworth, E., Montor, W.R., Wong, J., Park, J.G., *et al.* (2011a). Protein Microarray Signature of Autoantibody Biomarkers for the Early Detection of Breast Cancer. Journal of Proteome Research *10*, 85-96.

Anderson, K.S., Wong, J., D'Souza, G., Riemer, A.B., Lorch, J., Haddad, R., Pai, S.I., Longtine, J., McClean, M., LaBaer, J., *et al.* (2011b). Serum antibodies to the HPV16 proteome as biomarkers for head and neck cancer. British Journal of Cancer *104*, 1896-1905.

Anderson, K.S., Wong, J., Vitonis, A., Crum, C.P., Sluss, P.M., LaBaer, J., and Cramer, D. (2010). p53 Autoantibodies as Potential Detection and Prognostic Biomarkers in Serous Ovarian Cancer. Cancer Epidemiology Biomarkers & Prevention *19*, 859-868.

Anderson, L., and Hunter, C.L. (2006). Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Molecular & Cellular Proteomics *5*, 573-588.

Andreu, P., Johansson, M., Affara, N.I., Pucci, F., Tan, T.T., Junankar, S., Korets, L., Lam, J., Tawfik, D., DeNardo, D.G., *et al.* (2010). FcR gamma Activation Regulates Inflammation-Associated Squamous Carcinogenesis. Cancer Cell *17*, 121-134.

Andriole, G.L., Crawford, E.D., Grubb, R.L., 3rd, Buys, S.S., Chia, D., Church, T.R., Fouad, M.N., Isaacs, C., Kvale, P.A., Reding, D.J., *et al.* (2012). Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. J Natl Cancer Inst *104*, 125-132.

Arnold, J.N., Saldova, R., Galligan, M.C., Murphy, T.B., Mimura-Kimura, Y., Telford, J.E., Godwin, A.K., and Rudd, P.M. (2011). Novel Glycan Biomarkers for the Detection of Lung Cancer. Journal of Proteome Research *10*, 1755-1764.

Baade, P.D., Fritschi, L., and Eakin, E.G. (2006). Non-cancer mortality among people diagnosed with cancer (Australia). Cancer causes & control : CCC *17*, 287-297.

Bach, P.B., Mirkin, J.N., Oliver, T.K., Azzoli, C.G., Berry, D.A., Brawley, O.W., Byers, T., Colditz, G.A., Gould, M.K., Jett, J.R., *et al.* (2012). Benefits and Harms of CT Screening for Lung Cancer A Systematic Review. Jama-Journal of the American Medical Association *307*, 2418-2429.

Berger, F., and Reiser, M.F. (2013). Micro-RNAs as Potential New Molecular Biomarkers in Oncology: Have They Reached Relevance for the Clinical Imaging Sciences? Theranostics *3*, 932-941.

Bettegowda, C., Sausen, M., Leary, R.J., Kinde, I., Wang, Y.X., Agrawal, N., Bartlett, B.R., Wang, H., Luber, B., Alani, R.M., *et al.* (2014). Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. Science Translational Medicine *6*, 11.

Bill-Axelson, A., Holmberg, L., Ruutu, M., Garmo, H., Stark, J.R., Busch, C., Nordling, S., Haggman, M., Andersson, S.O., Bratell, S., *et al.* (2011). Radical prostatectomy versus watchful waiting in early prostate cancer. The New England journal of medicine *364*, 1708-1717.

Blixt, O., Clo, E., Nudelman, A.S., Sorensen, K.K., Clausen, T., Wandall, H.H., Livingston, P.O., Clausen, H., and Jensen, K.J. (2010). A High-Throughput O-Glycopeptide Discovery Platform for Seromic Profiling. Journal of Proteome Research *9*, 5250-5261.

Boyle, P., Chapman, C.J., Holdenrieder, S., Murray, A., Robertson, C., Wood, W.C., Maddison, P., Healey, G., Fairley, G.H., Barnes, A.C., *et al.* (2011). Clinical validation of an autoantibody test for lung cancer. Annals of Oncology *22*, 383-389.

Cabezon, T., Gromova, I., Gromov, P., Serizawa, R., Wielenga, V.T., Kroman, N., Celis, J.E., and Moreira, J.M.A. (2013). Proteomic Profiling of Triple-negative Breast Carcinomas in Combination With a Three-tier Orthogonal Technology Approach Identifies Mage-A4 as Potential Therapeutic Target in Estrogen Receptor Negative Breast Cancer. Molecular & Cellular Proteomics *12*, 381-394.

Calonge, N., Petitti, D.B., DeWitt, T.G., Dietrich, A.J., Gregory, K.D., Grossman, D., Isham, G., LeFevre, M.L., Leipzig, R.M., Marion, L.N., *et al.* (2009). Screening for Breast Cancer: US Preventive Services Task Force Recommendation Statement. Annals of Internal Medicine *151*, 716-W236.

Chapman, C., Murray, A., Chakrabarti, J., Thorpe, A., Woolston, C., Sahin, U., Barnes, A., and Robertson, J. (2007). Autoantibodies in breast cancer: their use as an aid to early diagnosis. Annals of Oncology *18*, 868-873.

Chapman, C.J., Healey, G.F., Murray, A., Boyle, P., Robertson, C., Peek, L.J., Allen, J., Thorpe, A.J., Hamilton-Fairley, G., Parsy-Kowalska, C.B.*, et al.* (2012). EarlyCDT(R)-Lung test: improved clinical utility through additional autoantibody assays. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine *33*, 1319-1326.

Charpin, C., Martin, M., Balandraud, N., Roudier, J., and Auger, I. (2010). Autoantibodies to BRAF, a new family of autoantibodies associated with rheumatoid arthritis. Arthritis Research & Therapy *12*, 7.

Cheang, M.C.U., Voduc, D., Bajdik, C., Leung, S., McKinney, S., Chia, S.K., Perou, C.M., and Nielsen, T.O. (2008). Basal-like breast cancer defined by five biomarkers has superior prognostic value then triple-negative phenotype. Clinical Cancer Research *14*, 1368-1376.

Chen, Y.T., Scanlan, M.J., Sahin, U., Tureci, O., Gure, A.O., Tsang, S.L., Williamson, B., Stockert, E., Pfreundschuh, M., and Old, L.J. (1997). A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. Proceedings of the National Academy of Sciences of the United States of America *94*, 1914-1918.

Chou, R., Croswell, J.M., Dana, T., Bougatsos, C., Blazina, I., Fu, R., Gleitsmann, K., Koenig, H.C., Lam, C., Maltz, A.*, et al.* (2011). Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med *155*, 762-771.

Cline, M.S., Craft, B., Swatloski, T., Goldman, M., Ma, S., Haussler, D., and Zhu, J. (2013). Exploring TCGA Pan-Cancer data at the UCSC Cancer Genomics Browser. Sci Rep *3*, 2652.

Collett, K., Stefansson, I.M., Eide, J., Braaten, A., Wang, H., Eide, G.E., Thoresen, S.O., Foulkes, W.D., and Akslen, L.A. (2005). A basal epithelial phenotype is more frequent in interval breast cancers compared with screen detected tumors. Cancer Epidemiology Biomarkers & Prevention *14*, 1108-1112.

Collisson, E.A., Campbell, J.D., Brooks, A.N., Berger, A.H., Lee, W., Chmielecki, J., Beer, D.G., Cope, L., Creighton, C.J., Danilova, L.*, et al.* (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543-550.

Cox, D.R. (1972). REGRESSION MODELS AND LIFE-TABLES. Journal of the Royal Statistical Society Series B-Statistical Methodology *34*, 187-+.

Croswell, J.M., Kramer, B.S., Kreimer, A.R., Prorok, P.C., Xu, J.L., Baker, S.G., Fagerstrom, R., Riley, T.L., Clapp, J.D., Berg, C.D.*, et al.* (2009). Cumulative Incidence of False-Positive Results in Repeated, Multimodal Cancer Screening. Annals of Family Medicine *7*, 212-222.

Davidoff, A.M., Iglehart, J.D., and Marks, J.R. (1992). IMMUNE-RESPONSE TO P53 IS DEPENDENT UPON P53/HSP70 COMPLEXES IN BREAST CANCERS. Proceedings of the National Academy of Sciences of the United States of America *89*, 3439-3442.

Davies, H., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M.J., Bottomley, W.*, et al.* (2002). Mutations of the BRAF gene in human cancer. Nature *417*, 949-954.

Dent, R., Trudeau, M., Pritchard, K.I., Hanna, W.M., Kahn, H.K., Sawka, C.A., Lickley, L.A., Rawlinson, E., Sun, P., and Narod, S.A. (2007). Triple-negative breast cancer: Clinical features and patterns of recurrence. Clinical Cancer Research *13*, 4429-4434.

Desmetz, C., Bascoul-Mollevi, C., Rochaix, P., Lamy, P.J., Kramar, A., Rouanet, P., Maudelonde, T., Mange, A., and Solassol, J. (2009). Identification of a New Panel of Serum Autoantibodies Associated with the Presence of In situ Carcinoma of the Breast in Younger Women. Clinical Cancer Research *15*, 4733-4741.

Disis, M.L., Pupa, S.M., Gralow, J.R., Dittadi, R., Menard, S., and Cheever, M.A. (1997). High-titer HER-2/neu protein-specific antibody can be detected in patients with early-stage breast cancer. Journal of Clinical Oncology *15*, 3363-3367.

Dogan, B.E., Gonzalez-Angulo, A.M., Gilcrease, M., Dryden, M.J., and Yang, W.T. (2010). Multimodality Imaging of Triple Receptor-Negative Tumors With Mammography, Ultrasound, and MRI. American Journal of Roentgenology *194*, 1160-1166.

Dogan, B.E., and Turnbull, L.W. (2012). Imaging of triple-negative breast cancer. Annals of Oncology *23*, 23-29.

Domingo, L., Salas, D., Zubizarreta, R., Bare, M., Sarriugarte, G., Barata, T., Ibanez, J., Blanch, J., Puig-Vives, M., Fernandez, A.B., *et al.* (2014). Tumor phenotype and breast density in distinct categories of interval cancer: results of population-based mammography screening in Spain. Breast Cancer Research *16*, 11.

Esserman, L., Shieh, Y., and Thompson, I. (2009). Rethinking Screening for Breast Cancer and Prostate Cancer. Jama-Journal of the American Medical Association *302*, 1685-1692.

Esserman, L.J., Shieh, Y., Park, J.W., and Ozanne, E.M. (2007). A role for biomarkers in the screening and diagnosis of breast cancer in younger women. Expert Review of Molecular Diagnostics *7*, 533-544.

Fan, B.C., Lu, K.Y., Sutandy, F.X.R., Chen, Y.W., Konan, K., Zhu, H., Kao, C.C., and Chen, C.S. (2014). A Human Proteome Microarray Identifies that the Heterogeneous Nuclear Ribonucleoprotein K (hnRNP K) Recognizes the 5 ' Terminal Sequence of the Hepatitis C Virus RNA. Molecular & Cellular Proteomics *13*, 84-92.

Farlow, E.C., Patel, K., Basu, S., Lee, B.S., Kim, A.W., Coon, J.S., Faber, L.P., Bonomi, P., Liptay, M.J., and Borgia, J.A. (2010). Development of a Multiplexed Tumor-Associated Autoantibody-Based Blood Test for the Detection of Non-Small Cell Lung Cancer. Clinical Cancer Research *16*, 3452-3462.

Fensterle, J., Becker, J.C., Potapenko, T., Heimbach, V., Vetter, C.S., Brocker, E.B., and Rapp, U.R. (2004). B-Raf specific antibody responses in melanoma patients. Bmc Cancer *4*, 9.

Festa, F., Rollins, S.M., Vattem, K., Hathaway, M., Lorenz, P., Mendoza, E.A., Yu, X.B., Qiu, J., Kilmer, G., Jensen, P., *et al.* (2013). Robust microarray production of freshly expressed proteins in a human milieu. Proteomics Clinical Applications *7*, 372-377.

Foulkes, W.D., Smith, I.E., and Reis, J.S. (2010). Triple-Negative Breast Cancer. New England Journal of Medicine *363*, 1938-1948.

Garcia-Closas, M., Brinton, L.A., Lissowska, J., Chatterjee, N., Peplonska, B., Anderson, W.F., Szeszenia-Dabrowska, N., Bardin-Mikolajczak, A., Zatonski, W., Blair, A*., et al.* (2006). Established breast cancer risk factors by clinically important tumour characteristics. British Journal of Cancer *95*, 123-129.

Ghillani, P., Andre, C., Toly, C., Rouquette, A.M., Bengoufa, D., Nicaise, P., Goulvestre, C., Gleizes, A., Dragon-Durey, M.A., Alyanakian, M.A*., et al.* (2011). Clinical significance of anti-Ro52 (TRIM21) antibodies non-associated with anti-SSA 60 kDa antibodies: Results of a multicentric study. Autoimmunity Reviews *10*, 509-513.

Ginsberg, R.J., and Rubinstein, L.V. (1995). Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. Lung Cancer Study Group. The Annals of thoracic surgery *60*, 615-622; discussion 622-613.

Gnjatic, S., Ritter, E., Buchler, M.W., Giese, N.A., Brors, B., Frei, C., Murray, A., Halama, N., Zornig, I., Chen, Y.T*., et al.* (2010). Seromic profiling of ovarian and pancreatic cancer. Proceedings of the National Academy of Sciences of the United States of America *107*, 5088-5093.

Gonzalez, R.M., Daly, D.S., Tan, R.M., Marks, J.R., and Zangar, R.C. (2011). Plasma Biomarker Profiles Differ Depending on Breast Cancer Subtype but RANTES Is Consistently Increased. Cancer Epidemiology Biomarkers & Prevention *20*, 1543-1551.

Goodell, V., Salazar, L.G., Urban, N., Drescher, C.W., Gray, H., Swensen, R.E., McIntosh, M.W., and Disis, M.L. (2006). Antibody immunity to the p53 oncogenic protein is a prognostic indicator in ovarian cancer. Journal of Clinical Oncology *24*, 762-768.

Grigoriadis, A., Caballero, O.L., Hoek, K.S., da Silva, L., Chen, Y.T., Shin, S.J., Jungbluth, A.A., Miller, L.D., Clouston, D., Cebon, J*., et al.* (2009). CT-X antigen expression in human breast cancer. Proceedings of the National Academy of Sciences of the United States of America *106*, 13493-13498.

Hamai, A., Duperrier-Amouriaux, K., Pignon, P., Raimbaud, I., Memeo, L., Colarossi, C., Canzonieri, V., Perin, T., Classe, J.M., Campone, M*., et al.* (2011). Antibody Responses to NY-ESO-1 in Primary Breast Cancer Identify a Subtype Target for Immunotherapy. Plos One *6*, 7.

Hanash, S.M., Baik, C.S., and Kallioniemi, O. (2011). Emerging molecular biomarkers-blood-based strategies to detect and monitor cancer. Nature Reviews Clinical Oncology *8*, 142-150.

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

He, P., Naka, T., Serada, S., Fujimoto, M., Tanaka, T., Hashimoto, S., Shima, Y., Yamadori, T., Suzuki, H., Hirashima, T.*, et al.* (2007). Proteomics-based identification of alpha-enolase as a tumor antigen in non-small lung cancer. Cancer Science *98*, 1234-1240.

Heneghan, H.M., Miller, N., Lowery, A.J., Sweeney, K.J., Newell, J., and Kerin, M.J. (2010). Circulating microRNAs as Novel Minimally Invasive Biomarkers for Breast Cancer. Annals of Surgery *251*, 499-505.

Hennessey, P.T., Sanford, T., Choudhary, A., Mydlarz, W.W., Brown, D., Adai, A.T., Ochs, M.F., Ahrendt, S.A., Mambo, E., and Califano, J.A. (2012). Serum microRNA Biomarkers for Detection of Non-Small Cell Lung Cancer. Plos One *7*, 6.

Hofman, V., Bonnetaud, C., Ilie, M.I., Vielh, P., Vignaud, J.M., Flejou, J.F., Lantuejoul, S., Piaton, E., Mourad, N., Butori, C.*, et al.* (2011). Preoperative Circulating Tumor Cell Detection Using the Isolation by Size of Epithelial Tumor Cell Method for Patients with Lung Cancer Is a New Prognostic Biomarker. Clinical Cancer Research *17*, 827-835.

Hong, B., and Zu, Y.L. (2013). Detecting Circulating Tumor Cells: Current Challenges and New Trends. Theranostics *3*, 377-394.

Hori, S.S., and Gambhir, S.S. (2011). Mathematical Model Identifies Blood Biomarker-Based Early Cancer Detection Strategies and Limitations. Science Translational Medicine *3*, 9.

Huang, Z.H., Huang, D., Ni, S.J.A., Peng, Z.L., Sheng, W.Q., and Du, X. (2010). Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. International Journal of Cancer *127*, 118-126.

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

Hurst, R., Hook, B., Slater, M.R., Hartnett, J., Storts, D.R., and Nath, N. (2009). Protein-protein interaction studies on protein arrays: Effect of detection strategies on signal-to-background ratios. Analytical Biochemistry *392*, 45-53.

Hyung, S.W., Lee, M.Y., Yu, J.H., Shin, B., Jung, H.J., Park, J.M., Han, W., Lee, K.M., Moon, H.G., Zhang, H.*, et al.* (2011). A Serum Protein Profile Predictive of the Resistance to Neoadjuvant Chemotherapy in Advanced Breast Cancers. Molecular & Cellular Proteomics *10*, 13.

Imperiale, T.F., Wagner, D.R., Lin, C.Y., Larkin, G.N., Rogge, J.D., and Ransohoff, D.F. (2000). Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. The New England journal of medicine *343*, 169-174.

Jager, D., Stockert, E., Gure, A.O., Scanlan, M.J., Karbach, J., Jager, E., Knuth, A., Old, L.J., and Chen, Y.T. (2001). Identification of a tissue-specific putative transcription factor in breast tissue by serological screening of a breast cancer library. Cancer Research *61*, 2055-2061.

Jang, C.Y., Wong, J., Coppinger, J.A., Seki, A., Yates, J.R., and Fang, G.W. (2008). DDA3 recruits microtubule depolymerase Kif2a to spindle poles and controls spindle dynamics and mitotic chromosome movement. Journal of Cell Biology *181*, 255-267.

Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., and Forman, D. (2011). Global Cancer Statistics. Ca-a Cancer Journal for Clinicians *61*, 69-90.

Kaplan, E.L., and Meier, P. (1958). NONPARAMETRIC-ESTIMATION FROM INCOMPLETE OBSERVATIONS. Journal of the American Statistical Association *53*, 457-481.

Kazerounian, S., and Aho, S. (2003). Characterization of periphilin, a widespread, highly insoluble nuclear protein and potential constituent of the keratinocyte cornified envelope. Journal of Biological Chemistry *278*, 36707-36717.

Kijanka, G., and Murphy, D. (2009). Protein arrays as tools for serum autoantibody marker discovery in cancer. Journal of Proteomics *72*, 936-944.

Kim, J., Namchuk, M., Bugawan, T., Fu, Q., Jaffe, M., Shi, Y.G., Aanstoot, H.J., Turck, C.W., Erlich, H., Lennon, V.*, et al.* (1994). HIGHER AUTOANTIBODY LEVELS AND RECOGNITION OF A LINEAR NH2-TERMINAL EPITOPE IN THE AUTOANTIGEN GAD(65), DISTINGUISH STIFF-MAN SYNDROME FROM INSULIN-DEPENDENT DIABETES-MELLITUS. Journal of Experimental Medicine *180*, 595-606.

Kirsh, V.A., Chiarelli, A.M., Edwards, S.A., O'Malley, F.P., Shumak, R.S., Yaffe, M.J., and Boyd, N.F. (2011). Tumor Characteristics Associated With Mammographic Detection of Breast Cancer in the Ontario Breast Screening Program. Journal of the National Cancer Institute *103*, 942-950.

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R.*, et al.* (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61-70.

Koziol, J.A., Zhang, J.Y., Casiano, C.A., Peng, X.X., Shi, F.D., Feng, A.C., Chan, E.K.L., and Tan, E.M. (2003). Recursive partitioning as an approach to selection of immune markers for tumor diagnosis. Clinical Cancer Research *9*, 5120-5126.

Krebs, M.G., Hou, J.M., Sloane, R., Lancashire, L., Priest, L., Nonaka, D., Ward, T.H., Backen, A., Clack, G., Hughes, A.*, et al.* (2012). Analysis of Circulating Tumor Cells in Patients with Non-small Cell Lung Cancer Using Epithelial Marker-Dependent and -Independent Approaches. Journal of Thoracic Oncology *7*, 306-315.

Kuboshima, M., Shimada, H., Liu, T.L., Nomura, F., Takiguchi, M., Hiwasa, T., and Ochiai, T. (2006). Presence of serum tripartite motif-containing 21 antibodies in patients with esophageal squamous cell carcinoma. Cancer Science *97*, 380-386.

Lacombe, J., Mange, A., Jarlier, M., Bascoul-Mollevi, C., Rouanet, P., Lamy, P.J., Maudelonde, T., and Solassol, J. (2013). Identification and validation of new autoantibodies for the diagnosis of DCIS and node negative early-stage breast cancers. International Journal of Cancer *132*, 1105-1113.

Lam, S., Boyle, P., Healey, G.F., Maddison, P., Peek, L., Murray, A., Chapman, C.J., Allen, J., Wood, W.C., Sewell, H.F.*, et al.* (2011). EarlyCDT-Lung: An Immunobiomarker Test as an Aid to Early Detection of Lung Cancer. Cancer Prevention Research *4*, 1126-1134.

Langer, C., Marcucci, G., Holland, K.B., Radmacher, M.D., Maharry, K., Paschka, P., Whitman, S.P., Mrozek, K., Baldus, C.D., Vij, R.*, et al.* (2009). Prognostic Importance of MN1 Transcript Levels, and Biologic Insights From MN1-Associated Gene and MicroRNA Expression Signatures in Cytogenetically Normal Acute Myeloid Leukemia: A Cancer and Leukemia Group B Study. Journal of Clinical Oncology *27*, 3198-3204.

Larman, H.B., Zhao, Z.M., Laserson, U., Li, M.Z., Ciccia, A., Gakidis, M.A.M., Church, G.M., Kesari, S., LeProust, E.M., Solimini, N.L.*, et al.* (2011). Autoantigen discovery with a synthetic human peptidome. Nature Biotechnology *29*, 535-U101.

Li, C.I., Mirus, J.E., Zhang, Y.Z., Ramirez, A.B., Ladd, J.J., Prentice, R.L., McIntosh, M.W., Hanash, S.M., and Lampe, P.D. (2012). Discovery and preliminary confirmation of novel early detection biomarkers for triple-negative breast cancer using preclinical plasma samples from the Women's Health Initiative observational study. Breast Cancer Research and Treatment *135*, 611-618.

Li, W.L., Wang, W., Sun, S.P., Sun, Y., Pan, Y., Wang, L.N., Zhang, R., Zhang, K., and Li, J.M. (2011). Autoantibodies against the Catalytic Domain of BRAF Are Not Specific Serum Markers for Rheumatoid Arthritis. Plos One *6*, 6.

Li, X.J., Hayward, C., Fong, P.Y., Dominguez, M., Hunsucker, S.W., Lee, L.W., McLean, M., Law, S., Butler, H., Schirm, M.*, et al.* (2013). A Blood-Based Proteomic Classifier for the Molecular Characterization of Pulmonary Nodules. Science Translational Medicine *5*, 10.

Li, Z., Gourguechon, S., and Wang, C.C. (2007). Tousled-like kinase in a microbial eukaryote regulates spindle assembly and S-phase progression by interacting with Aurora kinase and chromatin assembly factors. Journal of Cell Science *120*, 3883-3894.

Lieberman, D.A., Weiss, D.G., Bond, J.H., Ahnen, D.J., Garewal, H., and Chejfec, G. (2000). Use of colonoscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380. The New England journal of medicine *343*, 162-168.

Line, A., Stengrevics, A., Slucka, Z., Li, G., Jankevics, E., and Rees, R.C. (2002). Serological identification and expression analysis of gastric cancer-associated genes. British Journal of Cancer *86*, 1824-1830.

Liu, N.Q., Stingl, C., Look, M.P., Smid, M., Braakman, R.B.H., De Marchi, T., Sieuwerts, A.M., Span, P.N., Sweep, F., Linderholm, B.K.*, et al.* (2014). Comparative Proteome Analysis Revealing an 11-Protein Signature for Aggressive Triple-Negative Breast Cancer. Jnci-Journal of the National Cancer Institute *106*, 10.

Lu, D., Fall, K., Sparen, P., Ye, W., Adami, H.O., Valdimarsdottir, U., and Fang, F. (2013). Suicide and suicide attempt after a cancer diagnosis among young individuals. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO *24*, 3112-3117.

Lu, H., Hu, L., Yu, L., Wang, X., Urvalek, A.M., Li, T., Shen, C., Mukherjee, D., Lahiri, S.K., Wason, M.S.*, et al.* (2014). KLF8 and FAK cooperatively enrich the active MMP14 on the cell surface required for the metastatic progression of breast cancer. Oncogene *33*, 2909-2917.

Lu, H.L., Ladd, J., Feng, Z.D., Wu, M., Goodell, V., Pitteri, S.J., Li, C.I., Prentice, R., Hanash, S.M., and Disis, M.L. (2012). Evaluation of Known Oncoantibodies, HER2, p53, and Cyclin B1, in Prediagnostic Breast Cancer Sera. Cancer Prevention Research *5*, 1036-1043.

Lutz, A.M., Willmann, J.K., Cochran, F.V., Ray, P., and Gambhir, S.S. (2008). Cancer screening: A mathematical model relating secreted blood biomarker levels to tumor sizes. Plos Medicine *5*, 1287-1297.

MacBeath, G., and Schreiber, S.L. (2000). Printing proteins as microarrays for high-throughput function determination. Science *289*, 1760-1763.

Macdonald, I.K., Murray, A., Healey, G.F., Parsy-Kowalska, C.B., Allen, J., McElveen, J., Robertson, C., Sewell, H.F., Chapman, C.J., and Robertson, J.F.R. (2012). Application of a High Throughput Method of Biomarker Discovery to Improvement of the EarlyCDT (R)-Lung Test. Plos One *7*, 9.

Mange, A., Lacombe, J., Bascoul-Mollevi, C., Jarlier, M., Lamy, P.J., Rouanet, P., Maudelonde, T., and Solassol, J. (2012). Serum Autoantibody Signature of Ductal Carcinoma In Situ Progression to Invasive Breast Cancer. Clinical Cancer Research *18*, 1992-2000.

Metzeler, K.H., Dufour, A., Benthaus, T., Hummel, M., Sauerland, M.C., Heinecke, A., Berdel, W.E., Buchner, T., Wormann, B., Mansmann, U.*, et al.* (2009). ERG Expression Is an Independent Prognostic Factor and Allows Refined Risk Stratification in Cytogenetically Normal Acute Myeloid Leukemia: A Comprehensive Analysis of ERG, MN1, and BAALC Transcript Levels Using Oligonucleotide Microarrays. Journal of Clinical Oncology *27*, 5031-5038.

Metzger, O., Tutt, A., de Azambuja, E., Saini, K.S., Viale, G., Loi, S., Bradbury, I., Bliss, J.M., Azim, H.A., Ellis, P.*, et al.* (2012). Dissecting the Heterogeneity of Triple-Negative Breast Cancer. Journal of Clinical Oncology *30*, 1879-1887.

Miersch, S., Bian, X.F., Wallstrom, G., Sibani, S., Logvinenko, T., Wasserfall, C.H., Schatz, D., Atkinson, M., Qiu, J., and LaBaer, J. (2013). Serological autoantibody profiling of type 1 diabetes by protein arrays. Journal of Proteomics *94*, 486-496.

Montor, W.R., Huang, J., Hu, Y.H., Hainsworth, E., Lynch, S., Kronish, J.W., Ordonez, C.L., Logvinenko, T., Lory, S., and LaBaer, J. (2009). Genome-Wide Study of Pseudomonas aeruginosa Outer Membrane Protein Immunogenicity Using Self-Assembling Protein Microarrays. Infection and Immunity *77*, 4877-4886.

Mostert, B., Sleijfer, S., Foekens, J.A., and Gratama, J.W. (2009). Circulating tumor cells (CTCs): Detection methods and their clinical relevance in breast cancer. Cancer Treatment Reviews *35*, 463-474.

Moyer, V.A. (2012). Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med *157*, 120-134.

Murray, A., Chapman, C.J., Healey, G., Peek, L.J., Parsons, G., Baldwin, D., Barnes, A., Sewell, H.F., Fritsche, H.A., and Robertson, J.F.R. (2010). Technical validation of an autoantibody test for lung cancer. Annals of Oncology *21*, 1687-1693.

Nam, M.J., Madoz-Gurpide, J., Wang, H., Lescure, P., Schmalbach, C.E., Zhao, R., Misek, D.E., Kuick, R., Brenner, D.E., and Hanash, S.M. (2003). Molecular profiling of the immune response in colon cancer using protein microarrays: Occurrence of autoantibodies to ubiquitin C-terminal hydrolase L3. Proteomics *3*, 2108-2115.

Nielsen, T.O., Parker, J.S., Leung, S., Voduc, D., Ebbert, M., Vickery, T., Davies, S.R., Snider, J., Stijleman, I.J., Reed, J.*, et al.* (2010). A Comparison of PAM50 Intrinsic Subtyping with Immunohistochemistry and Clinical Prognostic Factors in Tamoxifen-Treated Estrogen Receptor-Positive Breast Cancer. Clinical Cancer Research *16*, 5222-5232.

Olave, I., Wang, W., Xue, Y., Kuo, A., and Crabtree, G.R. (2002). Identification of a polymorphic, neuron-specific chromatin remodeling complex. Genes & development *16*, 2509-2517.

Palmer, R.M., Wilson, R.F., Hasan, A.S., and Scott, D.A. (2005). Mechanisms of action of environmental factors - tobacco smoking. Journal of Clinical Periodontology *32*, 180-195.

Pao, W., and Girard, N. (2011). New driver mutations in non-small-cell lung cancer. Lancet Oncology *12*, 175-180.

Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X.P., Hu, Z.Y.*, et al.* (2009). Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. Journal of Clinical Oncology *27*, 1160-1167.

Pepe, M.S., Etzioni, R., Feng, Z., Potter, J.D., Thompson, M.L., Thornquist, M., Winget, M., and Yasui, Y. (2001). Phases of biomarker development for early detection of cancer. J Natl Cancer Inst *93*, 1054-1061.

Pereira-Faca, S.R., Kuick, R., Puravs, E., Zhang, Q., Krasnoselsky, A.L., Phanstiel, D., Qiu, J., Misek, D.E., Hinderer, R., Tammemagi, M.*, et al.* (2007). Identification of 14-3-3 theta as an antigen that induces a humoral response in lung cancer. Cancer Research *67*, 12000-12006.

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

Pietrowska, M., Marczak, L., Polanska, J., Behrendt, K., Nowicka, E., Walaszczyk, A., Chmura, A., Deja, R., Stobiecki, M., Polanski, A., *et al.* (2009). Mass spectrometry-based serum proteome pattern analysis in molecular diagnostics of early stage breast cancer. Journal of Translational Medicine *7*, 13.

Piura, E., and Piura, B. (2011). Autoantibodies to tailor-made panels of tumor-associated antigens in breast carcinoma. Journal of oncology *2011*, 982425.

Qiu, J., Choi, G., Li, L., Wang, H., Pitteri, S.J., Pereira-Faca, S.R., Krasnoselsky, A.L., Randolph, T.W., Omenn, G.S., Edelstein, C., *et al.* (2008). Occurrence of Autoantibodies to Annexin I, 14-3-3 Theta and LAMR1 in Prediagnostic Lung Cancer Sera. Journal of Clinical Oncology *26*, 5060-5066.

Qiu, J., and LaBaer, J. (2011). NUCLEIC ACID PROGRAMMABLE PROTEIN ARRAY: A JUST-IN-TIME MULTIPLEXED PROTEIN EXPRESSION AND PURIFICATION PLATFORM. In Methods in Enzymology, Vol 500: Methods in Systems Biology, D. Jameson, M. Verma, and H.V. Westerhoff, eds. (San Diego: Elsevier Academic Press Inc), pp. 151-163.

Qiu, J., Madoz-Gurpide, J., Misek, D.E., Kuick, R., Brenner, D.E., Michailidis, G., Haab, B.B., Omenn, G.S., and Hanash, S. (2004). Development of natural protein microarrays for diagnosing cancer based on an antibody response to tumor antigens. Journal of Proteome Research *3*, 261-267.

Rakha, E.A., El-Sayed, M.E., Green, A.R., Lee, A.H.S., Robertson, J.F., and Ellis, I.O. (2007). Prognostic markers in triple-negative breast cancer. Cancer *109*, 25-32.

Rakha, E.A., Reis-Filho, J.S., and Ellis, I.O. (2008). Basal-like breast cancer: A critical review. Journal of Clinical Oncology *26*, 2568-2581.

Ramachandran, N., Anderson, K.S., Raphael, J.V., Hainsworth, E., Sibani, S., Montor, W.R., Pacek, M., Wong, J., Eljanne, M., Sanda, M.G., *et al.* (2008a). Tracking humoral

responses using self assembling protein microarrays. Proteomics Clinical Applications *2*, 1518-1527.

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A.Y., Walter, J.C., and LaBaer, J. (2004). Self-assembling protein microarrays. Science *305*, 86-90.

Ramachandran, N., Raphael, J.V., Hainsworth, E., Demirkan, G., Fuentes, M.G., Rolfs, A., Hu, Y.H., and LaBaer, J. (2008b). Next-generation high-density self-assembling functional protein arrays. Nature Methods *5*, 535-538.

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

Reddy, B.A., van der Knaap, J.A., Bot, A.G.M., Mohd-Sarip, A., Dekkers, D.H.W., Timmermans, M.A., Martens, J.W.M., Demmers, J.A.A., and Verrijzer, C.P. (2014). Nucleotide Biosynthetic Enzyme GMP Synthase Is a TRIM21-Controlled Relay of p53 Stabilization. Molecular Cell *53*, 458-470.

Reis-Filho, J.S., and Tutt, A.N.J. (2008). Triple negative tumours: a critical review. Histopathology *52*, 108-118.

Reuschenbach, M., Doeberitz, M.V., and Wentzensen, N. (2009). A systematic review of humoral immune responses against tumor antigens. Cancer Immunology Immunotherapy *58*, 1535-1544.

Romanov, V., Davidoff, S.N., Miles, A.R., Grainger, D.W., Gale, B.K., and Brooks, B.D. (2014). A critical comparison of protein microarray fabrication technologies. Analyst *139*, 1303-1326.

Rosario, D.J., Lane, J.A., Metcalfe, C., Donovan, J.L., Doble, A., Goodwin, L., Davis, M., Catto, J.W., Avery, K., Neal, D.E., *et al.* (2012). Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: prospective evaluation within ProtecT study. BMJ (Clinical research ed) *344*, d7894.

Sahin, U., Tureci, O., Schmitt, H., Cochlovius, B., Johannes, T., Schmits, R., Stenner, F., Luo, G.R., Schobert, I., and Pfreundschuh, M. (1995). HUMAN NEOPLASMS ELICIT MULTIPLE SPECIFIC IMMUNE-RESPONSES IN THE AUTOLOGOUS HOST. Proceedings of the National Academy of Sciences of the United States of America *92*, 11810-11813.

Scanlan, M.J., Chen, Y.T., Williamson, B., Gure, A.O., Stockert, E., Gordan, J.D., Tureci, O., Sahin, U., Pfreundschuh, M., and Old, L.J. (1998). Characterization of human colon cancer antigens recognized by autologous antibodies. International Journal of Cancer *76*, 652-658.

Scanlan, M.J., Gure, A.O., Jungbluth, A.A., Old, L.J., and Chen, Y.T. (2002). Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. Immunological Reviews *188*, 22-32.

Schroder, F.H., Hugosson, J., Roobol, M.J., Tammela, T.L., Ciatto, S., Nelen, V., Kwiatkowski, M., Lujan, M., Lilja, H., Zappa, M.*, et al.* (2009). Screening and prostate-cancer mortality in a randomized European study. The New England journal of medicine *360*, 1320-1328.

Schroder, F.H., Hugosson, J., Roobol, M.J., Tammela, T.L., Ciatto, S., Nelen, V., Kwiatkowski, M., Lujan, M., Lilja, H., Zappa, M.*, et al.* (2012). Prostate-cancer mortality at 11 years of follow-up. The New England journal of medicine *366*, 981-990.

Schwartz, H.L., Chandonia, J.M., Kash, S.F., Kanaani, J., Tunnell, E., Domingo, A., Cohen, F.E., Banga, J.P., Madec, A.M., Richter, W.*, et al.* (1999). High-resolution autoreactive epitope mapping and structural modeling of the 65 kDa form of human glutamic acid decarboxylase. Journal of Molecular Biology *287*, 983-999.

Schwind, S., Marcucci, G., Kohlschmidt, J., Radmacher, M.D., Mrozek, K., Maharry, K., Becker, H., Metzeler, K.H., Whitman, S.P., Wu, Y.Z.*, et al.* (2011). Low expression of MN1 associates with better treatment response in older patients with de novo cytogenetically normal acute myeloid leukemia. Blood *118*, 4188-4198.

Seiler, C.Y., Park, J.G., Sharma, A., Hunter, P., Surapaneni, P., Sedillo, C., Field, J., Algar, R., Price, A., Steel, J.*, et al.* (2014). DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research. Nucleic acids research *42*, D1253-1260.

Seliger, B., and Kellner, R. (2002). Design of proteome-based studies in combination with serology for the identification of biomarkers and novel targets. Proteomics *2*, 1641-1651.

Shan, Q., Lou, X.M., Xiao, T., Zhang, J., Sun, H.Y., Gao, Y.N., Cheng, S.J., Wu, L., Xu, N.Z., and Liu, S.Q. (2013). A cancer/testis antigen microarray to screen autoantibody biomarkers of non-small cell lung cancer. Cancer Letters *328*, 160-167.

Sherman, M.E., Rimm, D.L., Yang, X.H.R., Chatterjee, N., Brinton, L.A., Lissowska, J., Peplonska, B., Szeszenia-Dbrowska, N., Zatonski, W., Cartun, R*., et al.* (2007). Variation in breast cancer hormone receptor and HER2 levels by etiologic factors: A population-based analysis. International Journal of Cancer *121*, 1079-1085.

Siegel, R., Ma, J.M., Zou, Z.H., and Jemal, A. (2014). Cancer Statistics, 2014. Ca-a Cancer Journal for Clinicians *64*, 9-29.

Sihto, H., Lundin, J., Lehtimaki, T., Sarlomo-Rikala, M., Butzow, R., Holli, K., Sailas, L., Kataja, V., Lundin, M., Turpeenniemi-Hujanen, T*., et al.* (2008). Molecular subtypes of breast cancers detected in mammography screening and outside of screening. Clinical Cancer Research *14*, 4103-4110.

Sima, C.S., Panageas, K.S., and Schrag, D. (2010). Cancer Screening Among Patients With Advanced Cancer. Jama-Journal of the American Medical Association *304*, 1584-1591.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S*., et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the United States of America *98*, 10869-10874.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S*., et al.* (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences of the United States of America *100*, 8418-8423.

Soussi, T. (2000). p53 antibodies in the sera of patients with various types of cancer: A review. Cancer Research *60*, 1777-1788.

Stephen, C.W., Helminen, P., and Lane, D.P. (1995). CHARACTERIZATION OF EPITOPES ON HUMAN P53 USING PHAGE-DISPLAYED PEPTIDE LIBRARIES - INSIGHTS INTO ANTIBODY PEPTIDE INTERACTIONS. Journal of Molecular Biology *248*, 58-78.

Stockert, E., Jager, E., Chen, Y.T., Scanlan, M.J., Gout, I., Karbach, J., Arand, M., Knuth, A., and Old, L.J. (1998). A survey of the humoral immune response of cancer patients to a panel of human tumor antigens. Journal of Experimental Medicine *187*, 1349-1354.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Sun, X., Sandhu, R., Figueroa, J.D., Gierach, G.L., Sherman, M.E., and Troester, M.A. (2014). Benign Breast Tissue Composition in Breast Cancer Patients: Association with Risk Factors, Clinical Variables, and Gene Expression. Cancer Epidemiol Biomarkers Prev.

Surinova, S., Schiess, R., Huttenhain, R., Cerciello, F., Wollscheid, B., and Aebersold, R. (2011). On the Development of Plasma Protein Biomarkers. Journal of Proteome Research *10*, 5-16.

Takulapalli, B.R., Qiu, J., Magee, D.M., Kahn, P., Brunner, A., Barker, K., Means, S., Miersch, S., Bian, X., Mendoza, A.*, et al.* (2012). High density diffusion-free nanowell arrays. J Proteome Res *11*, 4382-4391.

Tamimi, R.M., Byrne, C., Colditz, G.A., and Hankinson, S.E. (2007). Endogenous hormone levels, mammographic density, and subsequent risk of breast cancer in postmenopausal women. Journal of the National Cancer Institute *99*, 1178-1187.

Tao, S.C., and Zhu, H. (2006). Protein chip fabrication by capture of nascent polypeptides. Nat Biotechnol *24*, 1253-1254.

Thierry, A.R., Mouliere, F., El Messaoudi, S., Mollevi, C., Lopez-Crapez, E., Rolet, F., Gillet, B., Gongora, C., Dechelotte, P., Robert, B., *et al.* (2014). Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. Nature Medicine *20*, 430-+.

Tomaino, B., Cappello, P., Capello, M., Fredolini, C., Sperduti, I., Migliorini, P., Salacone, P., Novarino, A., Giacobino, A., Ciuffreda, L., *et al.* (2011). Circulating Autoantibodies to Phosphorylated alpha-Enolase are a Hallmark of Pancreatic Cancer. Journal of Proteome Research *10*, 105-112.

Tuupanen, S., Hanninen, U.A., Kondelin, J., von Nandelstadh, P., Cajuso, T., Gylfe, A.E., Katainen, R., Tanskanen, T., Ristolainen, H., Bohm, J., *et al.* (2014). Identification of 33 candidate oncogenes by screening for base-specific mutations. British Journal of Cancer *111*, 1657-1662.

USPSTF (2009). Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med *151*, 716-726, w-236.

Vazquez-Del Mercado, M., Daneri-Navarro, A., Martin-Marquez, B.T., Ramirez, R.V., Velasco-Sanchez, D., Chan, J.Y.F., Calise, S.J., Chan, E.K.L., and Satoh, M. (2013). Autoantibodies Associated With Inflammatory Myopathy and Other Systemic Autoimmune Rheumatic Diseases In Sera From Breast Cancer Patients. Arthritis and Rheumatism *65*, S881-S882.

Vesely, M.D., Kershaw, M.H., Schreiber, R.D., and Smyth, M.J. (2011). Natural Innate and Adaptive Immunity to Cancer. In Annual Review of Immunology, Vol 29, W.E. Paul, D.R. Littman, and W.M. Yokoyama, eds. (Palo Alto: Annual Reviews), pp. 235-271.

Vigil, A., Chen, C., Jain, A., Nakajima-Sasaki, R., Jasinskas, A., Pablo, J., Hendrix, L.R., Samuel, J.E., and Felgner, P.L. (2011). Profiling the Humoral Immune Response of Acute and Chronic Q Fever by Protein Microarray. Molecular & Cellular Proteomics *10*, 12.

Wallstrom, G., Anderson, K.S., and LaBaer, J. (2013). Biomarker Discovery for Heterogeneous Diseases. Cancer Epidemiology Biomarkers & Prevention *22*, 747-755.

Wandall, H.H., Blixt, O., Tarp, M.A., Pedersen, J.W., Bennett, E.P., Mandel, U., Ragupathi, G., Livingston, P.O., Hollingsworth, M.A., Taylor-Papadimitriou, J.*, et al.* (2010). Cancer Biomarkers Defined by Autoantibody Signatures to Aberrant O-Glycopeptide Epitopes. Cancer Research *70*, 1306-1313.

Wang, J., Barker, K., Steel, J., Park, J., Saul, J., Festa, F., Wallstrom, G., Yu, X.B., Bian, X.F., Anderson, K.S.*, et al.* (2013). A versatile protein microarray platform enabling antibody profiling against denatured proteins. Proteomics Clinical Applications *7*, 378-383.

Wang, X., Lu, H., Urvalek, A.M., Li, T., Yu, L., Lamar, J., DiPersio, C.M., Feustel, P.J., and Zhao, J. (2011). KLF8 promotes human breast cancer cell invasion and metastasis by transcriptional activation of MMP9. Oncogene *30*, 1901-1911.

Wang, X.J., Yu, J.J., Sreekumar, A., Varambally, S., Shen, R.L., Giacherio, D., Mehra, R., Montie, J.E., Pienta, K.J., Sanda, M.G.*, et al.* (2005). Autoantibody signatures in prostate cancer. New England Journal of Medicine *353*, 1224-1235.

Whelan, S.A., Lu, M., He, J.B., Yan, W.H., Saxton, R.E., Faull, K.F., Whitelegge, J.P., and Chang, H.R. (2009). Mass Spectrometry (LC-MS/MS) Site-Mapping of N-Glycosylated Membrane Proteins for Breast Cancer Biomarkers. Journal of Proteome Research *8*, 4151-4160.

Wright, C., Sibani, S., Trudgian, D., Fischer, R., Kessler, B., LaBaer, J., and Bowness, P. (2012). Detection of Multiple Autoantibodies in Patients with Ankylosing Spondylitis Using Nucleic Acid Programmable Protein Arrays. Molecular & Cellular Proteomics *11*, 10.

Wu, L.L., Chang, W.J., Zhao, J.F., Yu, Y.W., Tan, X.J., Su, T., Zhao, L.J., Huang, S.D., Liu, S.Y., and Cao, G.W. (2010). Development of Autoantibody Signatures as Novel Diagnostic Biomarkers of Non-Small Cell Lung Cancer. Clinical Cancer Research *16*, 3760-3768.

Xu, Y.W., Peng, Y.H., Chen, B., Wu, Z.Y., Wu, J.Y., Shen, J.H., Zheng, C.P., Wang, S.H., Guo, H.P., Li, E.M., *et al.* (2014). Autoantibodies as Potential Biomarkers for the Early Detection of Esophageal Squamous Cell Carcinoma. American Journal of Gastroenterology *109*, 36-45.

Yang, L.N., Guo, S.J., Li, Y., Zhou, S.M., and Tao, S.C. (2011a). Protein microarrays for systems biology. Acta Biochimica Et Biophysica Sinica *43*, 161-171.

Yang, W.T., Dryden, M., Broglio, K., Gilcrease, M., Dawood, S., Dempsey, P.J., Valero, V., Hortobagyi, G., Atchley, D., and Arun, B. (2008). Mammographic features of triple receptor-negative primary breast cancers in young premenopausal women. Breast Cancer Research and Treatment *111*, 405-410.

Yang, X.R., Chang-Claude, J., Goode, E.L., Couch, F.J., Nevanlinna, H., Milne, R.L., Gaudet, M., Schmidt, M.K., Broeks, A., Cox, A., *et al.* (2011b). Associations of Breast Cancer Risk Factors With Tumor Subtypes: A Pooled Analysis From the Breast Cancer Association Consortium Studies. Journal of the National Cancer Institute *103*, 250-263.

Yang, X.R., Pfeiffer, R.M., Garcia-Closas, M., Rimm, D.L., Lissowska, J., Brinton, L.A., Peplonska, B., Hewitt, S.M., Cartun, R.W., Mandich, D., *et al.* (2007). Hormonal markers in breast cancer: Coexpression, relationship with pathologic characteristics, and risk factor associations in a population-based study. Cancer Research *67*, 10608-10617.

Yoo, A.S., Staahl, B.T., Chen, L., and Crabtree, G.R. (2009). MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. Nature *460*, 642-646.

Yu, X., Decker, K.B., Barker, K., Neunuebel, M.R., Saul, J., Graves, M., Westcott, N., Hang, H., LaBaer, J., Qiu, J., *et al.* (2015). Host-Pathogen Interaction Profiling Using Self-Assembling Human Protein Arrays. J Proteome Res.

Yu, X.B., Bian, X.F., Throop, A., Song, L.S., Del Moral, L., Park, J., Seiler, C., Fiacco, M., Steel, J., Hunter, P., *et al.* (2014a). Exploration of Panviral Proteome: High-Throughput Cloning and Functional Implications in Virus-host Interactions. Theranostics *4*, 808-822.

Yu, X.B., Woolery, A.R., Luong, P., Hao, Y.H., Grammel, M., Westcott, N., Park, J., Wang, J., Bian, X.F., Demirkan, G., *et al.* (2014b). Copper-catalyzed azide-alkyne cycloaddition (click chemistry)-based Detection of Global Pathogen-host AMPylation on Self-assembled Human Protein Microarrays. Molecular & Cellular Proteomics *13*, 3164-3176.

Zhao, H., Shen, J., Medico, L., Wang, D., Ambrosone, C.B., and Liu, S. (2010). A Pilot Study of Circulating miRNAs as Potential Biomarkers of Early Stage Breast Cancer. Plos One *5*, 12.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., *et al.* (2001). Global analysis of protein activities using proteome chips. Science *293*, 2101-2105.

APPENDIX A

FIGURE PERMISSIONS

Figure 1-1 License

RightsLink®

Home | Account Info | Help

ACS Publications
Most Trusted. Most Cited. Most Read.

Title: On the Development of Plasma Protein Biomarkers

Author: Silvia Surinova, Ralph Schiess, Ruth Hüttenhain, et al

Publication: Journal of Proteome Research

Publisher: American Chemical Society

Date: Jan 1, 2011

Copyright © 2011, American Chemical Society

Logged in as:
Jie Wang
Account #:
3000899022

LOGOUT

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

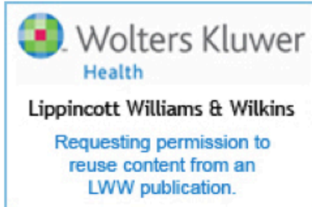
BACK | CLOSE WINDOW

Copyright © 2015 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Figure 1-2 License

RightsLink®

Home | Account Info | Help



Title: Clinical Relevance of Autoantibody Detection in Lung Cancer

Author: Jerome Solassol, Thierry Maudelonde, Alain Mange, et al

Publication: Journal of Thoracic Oncology

Publisher: Wolters Kluwer Health

Date: Jan 1, 2011

Copyright © 2011, (C) 2011International Association for the Study of Lung Cancer

Logged in as:
Jie Wang
Account #:
3000899022

LOGOUT

Order Completed

Thank you very much for your order.

This is a License Agreement between Jie Wang ("You") and Wolters Kluwer Health ("Wolters Kluwer Health"). The license consists of your order details, the terms and conditions provided by Wolters Kluwer Health, and the payment terms and conditions.

Get the printable license.

License Number	3595021368588
License date	Mar 23, 2015
Licensed content publisher	Wolters Kluwer Health
Licensed content publication	Journal of Thoracic Oncology
Licensed content title	Clinical Relevance of Autoantibody Detection in Lung Cancer
Licensed content author	Jerome Solassol, Thierry Maudelonde, Alain Mange, et al
Licensed content date	Jan 1, 2011
Volume number	6
Issue Number	5
Type of Use	Dissertation/Thesis
Requestor type	Individual
Portion	Figures/table/illustration
Number of figures/tables/illustrations	1
Figures/tables/illustrations used	Figure 1
Author of this Wolters Kluwer article	No
Title of your thesis / dissertation	Cancer autoantibody biomarker discovery and validation using nucleic acid programmable protein array
Expected completion date	May 2015
Estimated size(pages)	150
Total	0.00 USD

ORDER MORE... | CLOSE WINDOW

Copyright © 2015 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Figure 1-3A License

Copyright Clearance Center RightsLink®

Home | Account Info | Help

Title:	Design of proteome-based studies in combination with serology for the identification of biomarkers and novel targets
Author:	Barbara Seliger,Roland Kellner
Publication:	Proteomics
Publisher:	John Wiley and Sons
Date:	Dec 5, 2002

© 2002 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Logged in as:
Jie Wang
Account #:
3000899022

LOGOUT

Order Completed

Thank you for your order.

This Agreement between Jie Wang ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

Get the printable license.

License Number	3595031295655
License date	Mar 23, 2015
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Proteomics
Licensed Content Title	Design of proteome-based studies in combination with serology for the identification of biomarkers and novel targets
Licensed Content Author	Barbara Seliger,Roland Kellner
Licensed Content Date	Dec 5, 2002
Licensed Content Pages	11
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1
Will you be translating?	No
Title of your thesis / dissertation	Cancer autoantibody biomarker discovery and validation using nucleic acid programmable protein array
Expected completion date	May 2015
Expected size (number of pages)	150
Requestor Location	Jie Wang 1001 S McAllister Ave TEMPE, AZ 85287 United States Attn: Jie Wang

Billing Type	Invoice
Billing address	Jie Wang 1001 S McAllister Ave TEMPE, AZ 85287 United States Attn: Jie Wang
Total	0.00 USD

ORDER MORE CLOSE WINDOW

Copyright © 2015 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Figure 1-3B License

RightsLink®

Home | Account Info | Help

ACS Publications
Most Trusted. Most Cited. Most Read.

Title: The Sentinel Within: Exploiting the Immune System for Cancer Biomarkers†

Author: Karen S. Anderson, Joshua LaBaer

Publication: Journal of Proteome Research

Publisher: American Chemical Society

Date: Aug 1, 2005

Copyright © 2005, American Chemical Society

Logged in as:
Jie Wang
Account #:
3000899022

LOGOUT

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

Copyright © 2015 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

APPENDIX B

MANUSCRIPT COPYRIGHT PERSMISSIONS

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Mar 26, 2015

This Agreement between Jie Wang ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3590451094032
License date	Mar 15, 2015
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	PROTEOMICS - CLINICAL APPLICATIONS
Licensed Content Title	A versatile protein microarray platform enabling antibody profiling against denatured proteins
Licensed Content Author	Jie Wang,Kristi Barker,Jason Steel,Jin Park,Justin Saul,Fernanda Festa,Garrick Wallstrom,Xiaobo Yu,Xiaofang Bian,Karen S. Anderson,Jonine D. Figueroa,Joshua LaBaer,Ji Qiu
Licensed Content Date	May 17, 2013
Pages	6
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	Cancer autoantibody biomarker discovery and validation using nucleic acid programmable protein array
Expected completion date	May 2015
Expected size (number of pages)	150
Requestor Location	Jie Wang 1001 S McAllister Ave

	TEMPE, AZ 85287 United States Attn: Jie Wang
Billing Type	Invoice
Billing Address	Jie Wang 1001 S McAllister Ave TEMPE, AZ 85287 United States Attn: Jie Wang
Total	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking �accept� in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your Rightslink account (these are available at any time athttp://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this licence must be completed within two years of the date of the grant of this licence (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material.

Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto.

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC�s Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the

benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC� s Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC� s Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state� s conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses:: Creative Commons Attribution (CC-BY) license Creative Commons Attribution Non-Commercial (CC-BY-NC) license and Creative Commons Attribution Non-Commercial-NoDerivs (CC-BY-NC-ND) License. The license type is clearly identified on the article.

Copyright in any research article in a journal published as Open Access under a Creative Commons License is retained by the author(s). Authors grant Wiley a license to publish the article and identify itself as the original publisher. Authors also grant any third party the right to use the article freely as long as its integrity is maintained and its original authors, citation details and publisher are identified as follows: [Title of Article/Author/Journal Title and Volume/Issue. Copyright (c) [year] [copyright owner as specified in the Journal]. Links to the final article on Wiley� s website are encouraged where applicable.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-commercial re-use of an open access article, as long as the author is properly attributed.

The Creative Commons Attribution License does not affect the moral rights of authors, including without limitation the right not to have their work subjected to derogatory treatment. It also does not affect any other rights held by authors or third parties in the article, including without limitation the rights of privacy and publicity. Use of the article must not assert or imply, whether implicitly or explicitly, any connection with, endorsement or sponsorship of such use by the author, publisher or any other party associated with the article.

For any reuse or distribution, users must include the copyright notice and make clear to others that the article is made available under a Creative Commons Attribution license, linking to the relevant Creative Commons web page.

To the fullest extent permitted by applicable law, the article is made available as is and without representation or warranties of any kind whether express, implied, statutory or otherwise and including, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of defects, accuracy, or the presence or absence of errors.

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC) License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by non-commercial users

For non-commercial and non-promotional purposes, individual users may access, download, copy, display and redistribute to colleagues Wiley Open Access articles, as well as adapt, translate, text- and data-mine the content subject to the following

conditions:

- The authors' moral rights are not compromised. These rights include the right of "paternity" (also known as "attribution" - the right for the author to be identified as such) and "integrity" (the right for the author not to have the work altered in such a way that the author's reputation or integrity may be impugned).

- Where content in the article is identified as belonging to a third party, it is the obligation of the user to ensure that any reuse complies with the copyright policies of the owner of that content.

- If article content is copied, downloaded or otherwise reused for non-commercial research and education purposes, a link to the appropriate bibliographic citation (authors, journal, article title, volume, issue, page numbers, DOI and the link to the definitive published version on **Wiley Online Library**) should be maintained. Copyright notices and disclaimers must not be deleted.

- Any translations, for which a prior translation agreement with Wiley has not been agreed, must prominently display the statement: "This is an unofficial translation of an article that appeared in a Wiley publication. The publisher has not endorsed this translation."

Use by commercial "for-profit" organisations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee. Commercial purposes include:

- Copying or downloading of articles, or linking to such articles for further redistribution, sale or licensing;

- Copying, downloading or posting by a site or service that incorporates advertising with such content;

- The inclusion or incorporation of article content in other works or services (other than normal quotations with an appropriate citation) that is then available for sale or licensing, for a fee (for example, a compilation produced for marketing purposes, inclusion in a sales pack)

- Use of article content (other than normal quotations with appropriate citation) by for-profit organisations for promotional purposes

- Linking to article content in e-mails redistributed for promotional, marketing or educational purposes;

- Use for the purposes of monetary reward by means of sale, resale, licence, loan,

transfer or other form of commercial exploitation such as marketing products

- Print reprints of Wiley Open Access articles can be purchased from:corporatesales@wiley.com

Further details can be found on Wiley Online Libraryhttp://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.9

APPENDIX C

PUBLICATIONS AND SUBMISSIONS

CHAPTER 2

This work was published in the journal PROTEOMICS-Clinical Applications on May 13th, 2013

Wang, J., Barker, K., Steel, J., Park, J., Saul, J., Festa, F., Wallstrom, G., Yu, X., Bian, X., Anderson, K. S., Figueroa, J. D., LaBaer, J. and Qiu, J. (2013), A versatile protein microarray platform enabling antibody profiling against denatured proteins. Prot. Clin. Appl., 7: 378–383. doi: 10.1002/prca.201200062

CHAPTER 3

This work was submitted to the journal Cancer Epidemiology, Biomarkers & Prevention, and is currently under revision.

Jie Wang[a], Jonine D. Figueroa[a], Garrick Wallstrom, Kristi Barker, Jin G. Park, Gokhan Demirkan, Jolanta Lissowska, Karen S. Anderson, Ji Qiu*, Joshua LaBaer*. (2015), Plasma autoantibodies associated with basal-like breast cancers.

[a]Co-first authors *Corresponding authors

CHAPTER 4

This work is currently under preparation for publication.==---