

**DEVELOPMENT AND VALIDATION OF METHODS FOR THE
DIAGNOSIS OF LUNG CANCER VIA SEROLOGICAL BIOMARKERS**

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND
SCIENCES OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE
IN
MOLECULAR BIOLOGY AND GENETICS

By

Abbas Güven Akçay

February 2019

DEVELOPMENT AND VALIDATION OF METHODS FOR THE DIAGNOSIS OF LUNG CANCER VIA SEROLOGICAL BIOMARKERS

By Abbas Güven Akçay

2019

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science:

Ali Osmay Güre (Advisor)

Murat Alper Cevher

Zeliha Günnur Dikmen

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

DEVELOPMENT AND VALIDATION OF METHODS FOR THE DIAGNOSIS OF LUNG CANCER VIA SEROLOGICAL BIOMARKERS

Abbas Güven Akçay

M.Sc. in Molecular Biology and Genetics

Advisor: Ali Osmay Güre

February 2019

Over 10% of all new cancer cases are lung cancer. Moreover, estimates till 2030 indicate that already increasing lung cancer incidences will keep increasing, especially in developing countries like Turkey. Lung cancer, the leading cause of cancer deaths, has two large divisions: Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC). SCLC is the most aggressive subtype of lung cancer. And although, the treatment options and median survival time is more favorable in Limited Disease (LD), high tumor growth rate and metastatic tendency of SCLC even in the early stages, makes the diagnosis troublesome. Similarly, if NSCLC is diagnosed in early stages, surgery option is open and this increases the patient survival rate. However, current methods in screening and diagnosis, such as computed tomography (CT) and positron emission tomography (PET), are all limited by false positivity rates. Additionally, biopsy methods used in histological evaluations are both invasive and prone to false negativity. Therefore, new diagnostic tools which are cheap, accurate and non-invasive are in high demand. Autologous antibodies are abundantly elicited and stably exist in patient sera years before the clinical diagnosis of disease. Several such antibodies were reported by our group and other groups in lung cancer. Therefore, new diagnostic methods incorporating autologous antibodies can be a huge step forward in early diagnosis of lung cancer. Moreover, miRNAs, with their unique hormone like features such as circulation in serum and their regulatory effects in cell, are another good candidate for the early diagnosis of lung cancer. Therefore, in this study I aimed to develop a reliable, robust and automated evaluation method to re-evaluate custom Protein Array (cPA) screenings previously performed in our lab, and to determine the autologous antibodies with highest discriminatory power between SCLC patients & healthy controls. Moreover, I aimed to develop a Quartz Crystal Microbalance with Dissipation (QCM-D) based immunoassay to be incorporated later in the validation of cPA results. Lastly, in a parallel study I aimed to identify and validate novel miRNA biomarkers NSCLC. My results indicate that cPAs can have better sensitivity and specificity than ELISA and that QCM-D can be developed as an alternative to ELISA. miRNAs identified *in silico*, can also be validated *ex vivo*.

Previously, Protein Arrays (PAs) and cPAs were screened using 49 SCLC patient's and 50 healthy serums in our laboratory, incorporating visual and manual evaluations. Sensitivity and specificity values were calculated for individual autologous-antibodies and a number of autologous-antibody panels. Moreover, validations of cPA results were carried via ELISA. However, large discrepancies between cPA and ELISA results, as well as inconsistencies among ELISA results urged me to consider re-evaluation of cPA results with a more robust way, and to focus on developing a method superior to ELISA in autologous-antibody evaluations. Therefore, I incorporated AIDA to generate numeric values out of cPA screening images and filtered low quality data with optimized cut-off values. Several Receiver Operating Characteristic (ROC) curves were plotted using evaluated data. Improved results were evident by the increased Area Under Curve (AUC) values in both individual and combined ROC curves. Moreover, I developed a QCM based immunosensor for detection of anti-SOX2 antibody to be incorporated later in validation of cPA results. Binding interaction between anti-SOX2 antibody and SOX2 protein was modelled using 1:1 Langmuir Isothermal Binding and standard curves generated in QCM. In a parallel study, I also investigated miRNAs significantly upregulated in NSCLC when compared to high risk controls. For that purpose, miRNA expression datasets were gathered from GEO. Selected 2 datasets with the same sample type were analyzed for common significantly upregulated miRNAs among these two datasets. Significantly upregulated miRNAs were subjected to logistic regression analysis with LASSO regularization (error metrics: AUC and MSE) to select best panel of miRNAs that can distinguish NSCLC patients from healthy controls in given datasets. Moreover, selected miRNAs were analyzed with qRT-PCR to validate the panel.

I was able to re-evaluate cPA results by eliminating low quality data from numeric values generated via AIDA software from cPA images. I identified a panel of 4 autologous antibodies (FKBP8 – P53 – SOX2 – POLB) which resulted in 60% sensitivity at 100% specificity in discrimination of SCLC from controls. ROC of this autologous antibody panel had an AUC of 95.04%. Given panel surpassed diagnostic power of the only commercially available diagnostic kit of the same kind; EarlyCDT-Lung. Moreover, proof of concept for measurements of anti-protein antibodies were carried successfully in QCM, using anti-SOX2 antibody-SOX2 protein pair in PBS buffer as an example for it. Early results of anti-SOX2 mAb QCM indicate a linear assay range comparable to ELISA. Langmuir Isothermal Binding model revealed a strong interaction between antibody and protein in our QCM anti-SOX2 measurement experiments. Lastly, I was able to select 5 miRNAs using logistic regression and LASSO regularization that can best discriminate between NSCLC patients and high risk controls. However, validation experiments using qRT-PCR needs to be repeated as low Ct values and prominent hemolysis in serum samples prevented drawing meaningful conclusions.

Keywords: Lung cancer, SCLC, NSCLC, custom protein array, early diagnosis, serum, serological, autologous antibodies, miRNA, biomarker, ELISA, QCM, ROC, sensitivity, specificity, logistic regression, LASSO, qRT-PCR.

ÖZET

AKCİĞER KANSERİNİN SEROLOJİK BİYOBELİRTEÇLER İLE TEŞHİSİNE YÖNELİK YÖNTEM GELİŞTİRİLMESİ VE DOĞRULANMASI

Abbas Güven Akçay

Moleküler Biyoloji ve Genetik, Yüksek Lisans

Tez Danışmanı: Ali Osmay Güre

Şubat 2019

Yeni tespit edilen kanser vakalarının %10'u akciğer kanseridir. Dahası 2030 için yapılan öngörüler halihazırda artmakta olan akciğer kanseri vakalarının tüm dünyada ve özellikle Türkiye gibi gelişmekte olan ülkelerde daha da artacağını göstermektedir. Şu an en ölümcül kanser tipi olan akciğer kanserinin alt tipleri iki başlık altında toplanır: Küçük Hücreli Akciğer Kanseri (KHAK) ve Küçük Hücreli Dışı Akciğer Kanseri (KHDAK). KHAK akciğer kanserinin en agresif tipidir. Erken Aşama (EA) hastalıkta tedavi seçenekleri ve ortalama hayatta kalma süreleri daha çok olmasına karşın, KHAK'deki hızlı tümör büyüme hızı ve hastalığın çok hızlı metastaz yapma eğilimi erken teşhisi zorlaştırmaktadır. Benzer şekilde KHDAK'de da Erken Aşama hastalıkta cerrahi operasyon seçeneğinin mümkün olması dolayısıyla hasta hayatta kalma oranı artmaktadır. Fakat şu an tarama ve erken teşhiste kullanılan Bilgisayarlı Tomografi (BT) ve Pozitron Emisyon Tomografisi (PET) gibi yöntemler verdikleri sonuçlarda yüksek hatalı pozitiflik olması sebebiyle yetersizdir. Ek olarak tümör cinsini tespit ederken histolojik değerlendirmede kullanılan biyopsi yöntemleri hem hasta için invazivdir hem de örnekleme hataları dolayısıyla verdikleri sonuçlar yüksek hatalı negatiflik içerebilmektedir. Bu yüzden teşhiste kullanılmak üzere ucuz, hassas ve invaziv olmayan yöntemlere büyük ihtiyaç bulunmaktadır. Otolog antikorlar kanserin klinik teşhisinden yıllar önce hasta serumunda bolca ve yüksek stabilitede bulunmaktadır. Akciğer kanserinde bu tür antikorların varlığına hem bizim grubumuz hem de diğer bilimsel çevrelerce değinilmiştir. Dolayısıyla otolog antikorların ölçümüne dayalı teşhis yöntemleri akciğer kanserinin erken teşhisinde büyük gelişmeler sağlayabilecek potansiyel içermektedir. Bunun dışında miRNA'lar vücutta yaptıkları hormon benzeri etkiler, serumda sirkülasyonları ve

hücrelerdeki düzenleyici etkileriyle akciğer kanserinin erken teşhisinde kullanılabilecek bir diğer adaydır. Bu sebeplerle, çalışmamda labımızda yapılmış cPA çalışmalarının tekrar değerlendirilmesi için güvenilir ve otomatize bir değerlendirme yöntemi geliştirmeyi amaçladım. Ek olarak, daha sonra cPA sonuçlarının doğrulamasında kullanılmak üzere QCM temelli bir immünoşansör geliştirmeyi planladım. Son olarak, bunlara paralel bir çalışmada KHDAK'nin erken teşhisinde kullanılmak üzere miRNA temelli biyobelirteç tespit ve doğrulama çalışmaları yapmayı planladım. Aldığım sonuçlar cPA'ların ELISA'ya kıyasla daha iyi hassaslık ve özgünlük değerlerine sahip olabileceğine ve ELISA'ya alternatif olarak QCM-D temelli yöntemlerin geliştirilebileceğine işaret etmektedir. Ayrıca *in silico* tespit ettiğim miRNA'ların *ex vivo* validasyonu mümkündür.

Labımızda gerçekleştirilen öncül çalışmalarda cPA ile 49 KHAK hastasına ve 50 sağlıklı kontrole ait serumlar taranmış ve sonuçları manuel ve görsel şekilde değerlendirilmişti. Bu çalışmada tek tek otolog antikorların ve bazı otolog antikor panellerinin hassaslık ve özgüllük değerleri hesaplanmıştı. Daha sonra cPA çalışmaları ELISA yöntemi ile doğrulanmıştır. Fakat cPA ve ELISA sonuçları arasındaki bazı büyük uyumsuzluklar ve ELISA yöntemi ile alınan ölçümlerin kendi içinde çelişkiler barındırması bu çalışmayla cPA sonuçlarının daha stabil bir yöntemle tekrar değerlendirilmesinde ve otolog antikorları ölçmede ELISA'dan daha başarılı bir yöntem geliştirme çabamda teşvik edici sebep oldu. Bu yüzden AIDA ile cPA tarama görüntülerinden numerik ölçüm değerleri elde edip düşük kalite sinyallerin belli bir eşik değeri ile elimine edilmesi sağlandı. Elde edilen veriler kullanılarak çok sayıda ROC eğrisi çizildi. Çizilen tekli otolog antikor ve panel ROC eğrilerindeki artan AUC değerleri uyguladığımız yeni değerlendirme yönteminin başarısı gösterdi. Ayrıca bu çalışmada anti-SOX2 antikorunu PBS solüsyonundan ölçebilen QCM temelli bir biyosensör geliştirildi. Bu yöntemin daha sonra cPA sonuçlarını doğrulamada kullanılması planlanmakta. SOX2 ile anti-SOX2 arasındaki bağlanma etkileşimi QCM standart eğri değerlerinin 1:1 Langmuir İzotermal Bağlanma modeline oturtulmasıyla incelendi. Bunlara paralel ve ek olarak KHDAK'de yüksek riskli kontrollere kıyasla anlamlı şekilde yukarı yönlü değişen miRNA'lar incelendi. Bunun için miRNA ekspresyon datasetleri GEO sitesinden belirlendi. Bunların içinden aynı örnek tipine sahip 2 KHDAK dataseti seçilip bu iki datasette de ekspresyon değerleri anlamlı ve yukarı yönlü değişim gösteren miRNA'lar gerekli analizlerle tespit edildi. Tespit edilen miRNA'lar lojistik regresyon ve LASSO düzenleme yöntemleriyle (Hata metriği: AUC ve MSE) analiz edilip hastaları kontrollerden ayırmada en başarılı olan miRNA paneli seçildi. Seçilen miRNA paneli için qRT-PCR ile doğrulama çalışması yapıldı.

cPA görüntülerinden AIDA programıyla elde edilen sayısal değerlerdeki düşük kaliteli verileri başarılı bir şekilde eleyip cPA sonuçlarını tekrar değerlendirebildim. 4 otolog antikordan (FKBP8 – P53 – SOX2 – POLB) oluşan bir panel belirledim. Belirlenen panelin KHAK hastalarını kontrollerden ayırma gücü %100 özgüllükte %60 hassasiyet olarak belirlendi. Bu panel için çizilen ROC eğrilerinin AUC değeri %95.04 olarak hesaplandı. Bahsi geçen panelin teşhis gücü halihazırda piyasada bulunan benzer tek kit olan EarlyCDT-Lung kitinin teşhis gücünün üstünde yer aldı. Anti-SOX2 antikoru ve SOX2 proteini immünojenik çiftini kullanarak PBS solüsyonunda gerçekleştirilen QCM ölçümleriyle bu yöntemin otolog antikörlerin ölçülmesinde kullanılabilirliği gösterildi. Ön çalışmalar anti-SOX2 mAb QCM yönteminin lineer ölçüm aralığının ELISA ile kıyaslanabilir seviyelerde olduğunu gösterdi. QCM anti-SOX2 yöntemiyle alınan ölçümlerin 1:1 Langmuir İzotermal Bağlanma modeline oturtulmasıyla kullandığım antijen antikör çifti arasında güçlü bir bağlanma etkileşimi tespit edildi. Ayrıca lojistik regresyon ve LASSO düzenleme yöntemiyle KHDAK hastalarını yüksek risk kontrollerden en iyi ayıran 5 miRNA tespit edildi. Fakat, qRT-PCR doğrulama deneylerinde elde edilen düşük Ct değerleri ve serumlarda tespit edilen yüksek hemoliz değerleri dolayısıyla deney sonuçlarından anlamlı çıkarımlar yapmak mümkün olmadı. Dolayısıyla yapılan doğrulama deneylerinin tekrarlanması gerekmektedir.

Anahtar Kelimeler: Akciğer kanseri, KHAK, KHDAK, özel yapım protein mikrodizilim, serum, serolojik, otolog antikör, miRNA, biyobelirteç, ELISA, QCM, ROC, hassaslık, özgüllük, lojistik regresyon, LASSO, qRT-PCR

ACKNOWLEDGEMENTS

The solution to life's "absurdity" as suggested by Camus in Myth of Sisyphus is to create our own meanings in life. Actualizing through the personal goals in life, we can create our own meanings and have the fulfillment of "a" purpose. As an individual beginning to his journey through the academia, I already set goals for myself to achieve. However, the paths towards these goals are not always free of obstacles. Needless to say, I definitely felt the rough start with my 3.5 years long MSc journey, in which I had to overcome problems that most of my peers did not have to. This is not a bad thing! Because as stated by Molière: "the greater the obstacle, the more glory in overcoming it". Luckily, sometimes the exact "overcoming" part or the burdens that "overcoming" lay on our body are eased by the touch of people in our lives. I am honored to be in the minds of these great people. With the opportunity at hand, here I would like to share my gratitude towards them by at least mentioning their names.

First of all, I would like to express my deep gratitude to my supervisor Assoc. Prof. Dr. Ali Osmay Güre. I am grateful to have such a kind supervisor who created a space for me to learn by trial and error. I appreciate how he always bared with me on my failures. His strong support and everlasting belief in my abilities I felt, it was what gave me the willpower to push beyond myself and transcend through what I thought was a dead end. I learned so much through his supervision not only in subjects related to academia but also in my personal life.

I also owe my best regards to Prof. Dr. Zeliha Günnur Dikmen and Asst. Prof. Murat Alper Cevher for their valuable counseling and insights in preparation of this thesis and being a part of my thesis committee.

I am sincerely thankful to Dr. Özlem Ertekin and Dr. Mehmet Altay ÜNAL for their ongoing intellectual help and immense support in development of the QCM immune assay. Without their valuable experiences, guidance and optimistic attitude it would be quite difficult, if ever possible for me to continue to the currently ongoing autologous-antibody QCM project.

I also would like to thank Assist. Prof. Urartu Özgür Şafak Şeker for allowing me to use his lab area and the QCM-D device in QCM related parts of the project. I am also thankful to him for finding time demonstrating QCM and helping with the analysis within the tight time schedule he has.

I appreciate Prof. Dr. Burçak Vural's and Prof. Dr. Zeliha Günnur Dikmen's efforts in providing us the serum samples for this study.

I sincerely appreciate all the support Muhammad Waqas Akbar provided to me. This one is especially dear to me. He not only patiently thought me several important things in the field as a colleague, but also been a great friend. After enduring the troubles of graduate life,

sharing all the laughter and tea together on a daily basis, I now see him as a big brother to myself. And with all honesty, if I am able to finish this study, his help was undeniable. I hope we can keep these memories fresh and add new ones in the future.

I also would like to thank to Hamza Yusuf Altun, my partner in TUBITAK 1512 project, for his help in the evaluation of cPAs as well as his hard work in actualization of our TUBITAK 1512 project.

I would like to thank to Barış Küçükkaraduman for his support both inside and outside our lab, in addition to his kind and understanding nature.

I would like to thank to Farid Ahadli for his contributions in NSCLC part of this project as a senior student. He helped the project with conducting several *in silico* analysis, requiring great attention and precision in addition to intensive labor. With his diligent work expect a great future ahead of him.

I would like to repeat names of some new and old lab members and extend my warm feelings towards the rest of the AOG members, Muhammad Waqas Akbar, Barış Küçükkaraduman, Ege Dedeoğlu, Seçil Demirkol, Şükrü Atakan, Kerem Mert Şenses, Alper Poyraz, Mehdi Ghassemi, Seyhan Türk, Farid Ahadli, Hamza Yusuf Altun and others who might have slipped from my mind.

I also would like to thank to my friends who never left my side and with whom I share a great deal of memories. I especially want to mention Murat Yaman and Fatma Seyhun Üstün, but also the remaining part of the “Kalabalık Grup” Seniye, Suphi, Berhan, Zeynep, Naz, Arzu, Elif, Yasin, Hakan, İbrahim, Cem and the rest who I might have slipped from my mind.

Penultimately, I also want to mention my deep appreciation to Deniz Esen, my beautiful and loving girlfriend, for being someone with whom I can forget everything and have the ease of mind in the darkest hours. Her mental, emotional and actual workwise help throughout the sleepless nights was far beyond mere support. I, frankly, could not expect more. Thank you for all the times we shared together and will share in the future.

Last but not least, I would like to thank to my family, Yasemin Akçay, Kani Akçay and Eren Akçay. No matter where I am, I always felt their love and caring, which has been a major source of determination in whatever I achieved till now.

TABLE OF CONTENTS

List of Figures	xiii
List of Tables.....	xiv
1. INTRODUCTION	1
1.1. Lung Cancer	1
1.1.1. Statistics & Clinical Staging of Lung Cancer.....	1
1.1.2. Imaging and Screening Options in Lung Cancer.....	2
1.2. Role of Immune System in Cancer.....	3
1.2.1. Immunoediting	3
1.2.2. Tumor Associated Antigens and Autologous Anti-Tumor Antibodies in Cancer Diagnosis.....	4
1.2.3. Importance of Longitudinal Studies in Serological Biomarker Studies of Cancer.....	4
1.3. miRNAs in Cancer	6
2. AIMS OF THE STUDY	7
2.1. Developing a reliable, robust and automated evaluation method to re-evaluate cPA screenings of SCLC patients and healthy controls.	7
2.2. Developing a method in Quartz Crystal Microbalance to validate cPA screening results.	7
2.3. Identification and validation of novel miRNA biomarkers in Non-Small Cell Lung Cancer.	7
3. MATERIALS AND METHODS.....	8
3.1. RE-ANALYSIS AND RE-EVALUATION OF SCLC AND HEALTHY SERA CUSTOM PROTEIN ARRAY SCREENINGS.....	8
3.1.1. Protein Array	8
3.1.2. Processing of cPA images in Photoshop	9
3.1.3. AIDA Signal Analysis Software	10
3.1.4. Data Filtration & Generation of Receiver Operating Characteristic Curves	11
3.2. QUARTZ CRYSTAL MICROBALANCE (QCM) METHOD DEVELOPMENT	12
3.2.1. Serum Collection	12
3.2.2. ELISA Measurements	12
3.2.3. QCM Chip Modification	13

3.2.4.	QCM Measurements.....	13
3.2.5.	Surface Characterization via ATR FT-IR.....	14
3.2.6.	QCM and ELISA Comparison	14
3.3.	DISCOVER AND VALIDATION of miRNA BIOMARKERS in NON-SMALL CELL LUNG CANCER.....	14
3.3.1.	Dataset Selection	14
3.3.2.	miRNA Selection Incorporating Logistic Regression and LASSO Regularization	15
3.3.3.	Serum Samples	16
3.3.4.	Hemolysis Assessment	16
3.3.5.	RNA Extraction and Quantification	16
3.3.6.	cDNA Generation and Quantification	16
3.3.7.	Quantitative Real Time Polymerase Chain Reaction (qRT-PCR).....	17
3.3.8.	Data Analysis	17
4.	RESULTS.....	18
4.1.	RE-ANALYSIS AND RE-EVALUATION OF SCLC AND HEALTHY SERA CUSTOM PROTEIN ARRAY SCREENINGS.....	18
4.1.1.	Protein Array	18
4.1.2.	Processing of cPA images in Photoshop	18
4.1.3.	Data Filtration & Generation of Receiver Operating Characteristic Curves	19
4.2.	QUARTZ CRYSTAL MICROBALANCE METHOD DEVELOPMENT	26
4.2.1.	ELISA Measurement	26
4.2.2.	QCM Chip Modification	27
4.2.3.	QCM Measurement	28
4.2.4.	Surface Characterization via ATR FT-IR.....	29
4.2.5.	QCM and ELISA Comparison	35
4.3.	DISCOVER AND VALIDATION of miRNA BIOMARKERS in NON-SMALL CELL LUNG CANCER.....	36
4.3.1.	Dataset Selection	36
4.3.2.	miRNA Selection Incorporating Logistic Regression and LASSO Regularization	37
4.3.3.	Hemolysis Assessment	45

4.3.4.	RNA Extraction and Quantification	45
4.3.5.	Quantitative Real Time Polymerase Chain Reaction (qRT-PCR).....	45
5.	DISCUSSION & FUTURE PERSPECTIVES.....	47
5.1.	PA Screening Evaluations	47
5.2.	Literature Search for the Autoantibodies in the Selected Panel	48
5.2.1.	FKBP8.....	48
5.2.2.	P53.....	48
5.2.3.	SOX2.....	48
5.2.4.	POLB.....	48
5.3.	QCM Development for Auto-Antibody Measurements	49
5.4.	miRNA Biomarkers in NSCLC.....	50
5.5.	Literature Search for the miRNAs in Selected Panel	51
5.5.1.	hsa-miR-130b-5p.....	51
5.5.2.	hsa-miR-132-5p.....	51
5.5.3.	hsa-let-7d-3p.....	51
5.5.4.	hsa-miR-1227-3p.....	51
5.5.5.	hsa-miR-199a-5p	52
6.	BIBLIOGRAPHY	53
7.	APPENDIX	58

List of Figures

Figure 1.1: Schematic illustration of immunosurveillance theory.....	3
Figure 1.2: Anti-p53 autologous antibody seropositivity in relation to development of cancer....	5
Figure 3.1: 12 distinct protein spotting patterns and a reference InkDot on a PA unit.....	8
Figure 3.2: Selection of Individual cPA Sections in Photoshop via Rectengular Marquee Tool and an Example of Images Employed in Generation of Unified Images	10
Figure 4.1: All cPA strips used in screening of SCLC patients and Healthy Controls.....	18
Figure 4.2: An Example of Signal Harboring Unit Elimination Using 1.5x StDev Cut-off.....	19
Figure 4.3: Pattern of discarded and analyzed signal harboring units in each cPA strip.....	20
Figure 4.4: Individual antibody ROC curves with highest AUC values.....	22-23-24
Figure 4.5: Combination ROC curve of proteins selected for QCM analysis.....	25
Figure 4.6: ELISA anti-SOX2 Standard curves prepared in different serum dilutions and PBS buffer & Pearson R correlation of each standard curve.....	26
Figure 4.7: QCM chip measurements throughout the surface modification.....	27
Figure 4.8: QCM measurements of anti-SOX2 mAb binding on SOX2 protein immobilized on AT-cut 5MHz QCM chip.....	28
Figure 4.9: Illustration of chip surface modification	29
Figure 4.10: Absorbance spectrum of 11-MUA polymerized QCM chip surface measured via ATR FT-IR.....	32
Figure 4.11: Absorbance spectrum of SOX2 immobilized QCM chip surface measured via ATR FT-IR.....	33
Figure 4.12: Absorbance spectrum of QCM chip surface after anti-SOX2 measurement, analyzed via ATR FT-IR.....	34
Figure 4.13: Comparison of ELISA and QCM standard curve measurements for anti-SOX2 mAb in PBS Buffer.....	35
Figure 4.14: Common miRNAs between all datasets.....	37
Figure 4.15: Error metric for Logistic Regression training in Training Group of GSE61741.....	39
Figure 4.16: ROC curve plotted from Training group of GSE61741 using LASSO generated model.....	40
Figure 4.17: ROC curve plotted from Validation group of GSE61741 using LASSO generated model with 5miRNAs.....	41
Figure 4.18: Error metric for Logistic Regression training in Training Group of GSE24709 ...	42
Figure 4.19: ROC curve plotted from GSE24709 using LASSO generated model with 4 miRNAs.	43
Figure 4.20: Mean Expression of LASSO Selected miRNAs in GSE61741 and GSE24709.....	44
Figure 4.21: Relative miRNA expressions prepared for 5 miRNAs in 7 NSCLC patients and their matched high risk controls.	46

List of Tables

Table 4.1: 11 Best ROC curve results according to AUC values.....	21
Table 4.2: Signal peaks obtained after chip surface characterization with ATR FT-IR and their attributed chemical structure.....	30-31
Table 4.3: List and details of the datasets selected for further analysis.....	36
Table 4.4: List of 18 significantly upregulated miRNAs common between GSE61741 and GSE24709.....	38
Table 4.5: List of 5 miRNAs selected by LASSO training in GSE61741 Training Group.....	39
Table 4.6: List of 4 miRNAs selected by LASSO training in GSE24709 dataset	44
Table 4.7: OD 414 measurements of samples for hemolysis assessment.....	45



Abbreviations

NSCLC	Non-Small Cell Lung Cancer
SCLC	Small Cell Lung Cancer
NE	Neuro-endocrine
TNM	Tumor-Node-Metastasis
VALSG	Veterans Administration Lung Study Group
LD	Limited Disease
ED	Extensive Disease
PET	Positron Emission Tomography
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
LDCT	Low Dose Computed Tomography
USPSTF	United States Preventive Services Task Force
HIV	Human Immunodeficiency Virus
TAA	Tumor Associated Antigens
miRNA	micro Ribonucleic Acid
PCR	Polymerase Chain Reaction
RT-PCR	Reverse Transcription Polymerase Chain Reaction
qPCR	Quantitative Polymerase Chain Reaction
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
cPA	custom Protein Array
QCM	Quartz Crystal Microbalance
PBS	Phosphate Buffered Saline
ELISA	Enzyme Linked Immunosorbent Assay
LASSO	Least Absolute Shrinkage and Selection Operator
FB	Fetal Brain
cDNA	Complementary Deoxyribonucleic Acid
IPTG	Isopropyl β -D-1-thiogalactopyranoside
<i>E. coli</i>	Escherichia coli
PVDF	Polyvinylidene fluoride
SAS	Signal Analysis Software
ROC	Receiver Operating Characteristic
TIF	Tagged Image File Format
JPEG	Joint Photographic Experts Group
StDev	Standard Deviation
AUC	Area Under Curve
LDH	Lactate dehydrogenase
AP	Alkaline Phosphatase
DHFR	Dihydrofolate Reductase
IgG	Immunoglobulin G

mAb	monoclonal Antibody
11-MUA	11-Mercaptoundecanoic Acid
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
NHS	N-Hydroxysuccinimide
QCM-D	Quartz Crystal Microbalance with Dissipation Monitoring
BSA	Bovine Serum Albumin
ATR FT-IR	Attenuated Total Reflection Fourier Transform Infrared Spectroscopy
GEO	Gene Expression Omnibus
TCGA-LUAD	The Cancer Genome Atlas Lung Adenocarcinoma
TCGA-LUSC	The Cancer Genome Atlas Lung Squamous Cell Carcinoma
NIH	National Institutes of Health
MPEA	Microfluidic Primer Extension Assay
COPD	Chronic obstructive pulmonary disease
MSE	Mean Squared Error
HRG	High Risk Control
SAM	Self-Assembled Monolayer
MHz	Megahertz

1. INTRODUCTION

1.1. Lung Cancer

1.1.1. Statistics & Clinical Staging of Lung Cancer

Estimates point out to above 18 million cancer diagnosis and almost 10 million cancer related deaths in 2018 [3]. This high cancer incidences were expected to continuously increase till 2030 in all cancers but colon and prostate, and the rate of increase is expected to be more dramatic in developing countries like Turkey, when compared to developed countries [4; 5]. Lung cancer, consisting of 10% of all cancer cases, is not only one of the most prevalent cancer types but also the deadliest with being the leading cause of cancer deaths [6]. Moreover, lung cancer is expected to hold this position as the cancer type with highest fatality until the end of estimation period at 2030 [4].

Diagnosis and treatment of lung cancer is still an ongoing problem, as several morphological phenotypes and genetic abnormalities seen throughout this cancer reminds the profile of a heterogeneous disease [6]. Therefore, the disease is divided into many sub-types. The two largest divisions encompassing many subtypes are Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC). NSCLC has a variety of subtypes such as adenocarcinoma, squamous cell carcinoma and large cell carcinoma [6], whereas SCLC is differentiated by its neuro endocrine (NE) characteristics [7]. SCLC comprises 15% of all cancer cases and is the most aggressive subtype of this cancer with its less than 7% 5-year survival rate [6; 8].

There are 2 systems currently used in clinical staging of SCLC: 4 stage Tumor-Node-Metastasis (TNM) classification system and Veterans Administration Lung Study Group (VALSG) Limited (LD) and Extensive Disease (ED) system. The LD – ED system does its categorization based on metastatic status of the cancer within or outside of thorax. Therefore, it is more informative about the response to treatment and clinically more useful [9]. Conversion between the two systems can be done easily. The first three stages (I-II-III) in 4 stage TNM system corresponds to Limited Disease in VALSG and TNM stage 4 (IV) is equivalent to Extensive Disease in VALSG system [7].

Several researches conducted using VALSG system shows that prognosis of patients in Limited Disease SCLC is better than those in Extensive Disease [10-13]. While the median survival time in Limited Disease is 18 months, it drops to 9 months in patients with Extensive Disease and to 2-4 months for untreated patients [14]. Although, prognosis is better and treatment options are more abundant in Limited Disease [9; 14], high tumor growth rate and tendency of distant metastasis in SCLC starting from the early stages of the disease makes the early diagnosis of SCLC quite challenging [15; 16].

Situation is similar with NSCLC. If the cancer diagnosed in early stages, radical surgery is the gold standard treatment and has good results [17]. Moreover, despite the developments seen in NSCLC through targeted therapies, such developments are lacking in SCLC. Furthermore, even though there has been improvements in targeted therapies in NSCLC, there are no such advancements in SCLC [18]. According to Byers et al. (2015), one of the hindrances in the development of targeted therapies against SCLC is the lacking of diagnostic methods [18]. Therefore, diagnosing of the disease at early stages before spreading to adherent and distant tissues is important for both patient prognosis and development of treatment options in the future.

1.1.2. Imaging and Screening Options in Lung Cancer

Imaging options currently available in screening of SCLC are chest radiography, positron emission tomography (PET), magnetic resonance imaging (MRI) and computed tomography (CT) [19-21]. However, due to quick progression of the disease, two thirds of the patients are diagnosed in the late stages of the disease after onset of symptoms [15; 22; 23]. After patients are symptomatic, these screening methods require further histological evaluation and validation of the samples taken via bronchoscopic biopsy techniques and transthoracic biopsy techniques for the proper diagnosis of the disease [24-27]. Moreover, fine needle aspiration based methods are prone to false negativity due to sampling errors [26; 28].

A large longitudinal study conducted using low dose computed tomography (LDCT) with over 31500 asymptomatic people for a duration of 11 years were concluded in year 2006 with diagnosis of 484 cancer patients. Initially, the fact that 85% of the patients were diagnosed in 1st Stage and they had a 5 year survival rate of 92% gave the impression that LDCT screenings was successful [29]. However, later studies revealed a group of patients within these early diagnosed ones which had cancer growing in such a slow rate that they would not require any treatment [30]. Operating such people who would not require treatment increased morbidity rates, thus raising skepticism towards screening studies among clinicians. Although, US Preventive Services Taskforce (USPSTF) advice LDCT screenings for high risk group people, unnecessary radiation exposure of people without cancer and most importantly high false negativity rates, keeps the debate on utility of suggesting LDCT screenings going [31-33]. However, with standardizations in screenings and analysis, it is still highly expected that LDCT screenings will be widespread [34]. In the case of widespread LDCT screenings, one of the most important subjects will be cost/benefit relationship. This can be improved with development of cheap non-invasive diagnostic methods that can be used in conjunction with LDCT.

1.2. Role of Immune System in Cancer

1.2.1. Immunoediting

Protein expression patterns change during tumorigenesis and this triggers the release of specific antibodies by the immune system to counteract the tumor differentiation [35]. The immunosurveillance theory formulated by Frank MacFarlane Burnet and Lewis Thomas suggests an interaction between the tumorigenic cells and the immune system as the cancer develops, which pressures the tumorigenic cells to adapt to escape detection by the immune system to survive in three phases: elimination, equilibrium and escape (Figure 1.1) [35]. In a healthy individual, the immune system is constantly monitoring the body for onset of cancer to be able to generate an early innate and humoral immunogenic response before the tumorigenic cells can evolve to avoid recognition. This is supported by the increased occurrence of cancer in immunosuppressed individuals, such as those infected with human immunodeficiency virus (HIV) [36]. In addition to that, immunosuppressed people have an increase in cancer sub-types when compared to immunocompetent people [37]. Therefore, with the involvement of humoral immune system, tumor-targeting autoantibodies can be found in higher concentration in serum than the tumor-associated antigens themselves, which presents them as a valuable target for early detection of cancer [37; 38]. However, the capability of the immune system to detect tumor cells decreases throughout tumorigenesis due to cancer immunoediting. The selective pressure on the tumor to escape the immune system's response to it results in adaptation of its cells specifically to become increasingly undetectable throughout the equilibrium and escape phases of immunosurveillance and establish their immunosuppressive tumor microenvironment [35; 39].

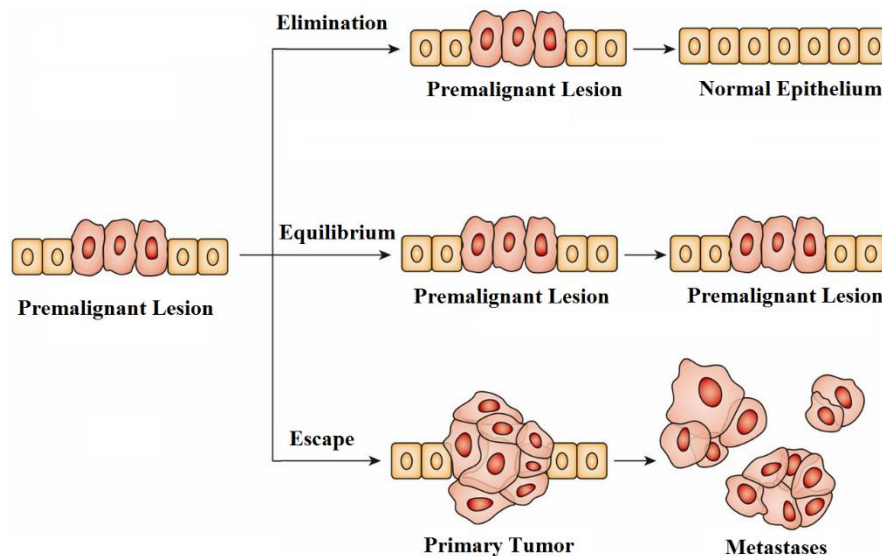


Figure 1.1: Schematic illustration of immunosurveillance theory [1]

1.2.2. Tumor Associated Antigens and Autologous Anti-Tumor Antibodies in Cancer Diagnosis

Immune system has the ability to differentiate between self and foreign. However, this system is not perfect and in certain conditions can react to self-antigens [38]. Cancer patients, being an example of the phenomena, many times respond to self-antigens most of which is expressed by the tumors. These self-antigens expressed in tumor microenvironment, later released into extracellular space and bodily fluids are called tumor associated antigens (TAAs) [38]. These antigens can be from proteins that are mutated like p53, overexpressed like NY-ESO-1, abnormally glycosylated like MUC-1, ectopically present or more [38; 40]. Utilizing tumor associated antigens (TAAs) and especially their complementary autologous antibodies for diagnosis and immunotherapy purposes has clinical advantages. From a therapeutic standpoint, tumor-targeting vaccines can be developed using TAAs commonly found in cancer patients, as well as in a personalized manner [41]. Moreover, working with TAAs is not ideal for prognostic purposes as their retrieval from the tumorigenic tissue is often only possible with an invasive biopsy procedure, which is costly and uncomfortable for the patient. Moreover, TAAs, present in low concentrations only, are expected to differ from patient to patient and are not consistently expressed by the same tumor due to cancer immune-editing. On the other hand, autologous anti-tumor antibodies are abundant in the blood serum even in early stages of cancer and are more stable than TAAs, which makes them easier to obtain from the patient in large volumes to be used in diagnostic experiments [38]. Indeed, breast and lung studies show that, autoantibody measurements against TAAs can be made 5 year before the symptomatic disease [42]. With amplifying the aberrant signals from TAAs and having the potential for compensating heterogeneity in tumor when checked in antibody panels, the autoantibodies might be the robust biomarkers for early detection of cancer [42-53]. Therefore, rapid and accurate identification of such autologous antibodies is essential for the reliable development of immuno-oncogenic diagnostic methods.

1.2.3. Importance of Longitudinal Studies in Serological Biomarker Studies of Cancer

It is important to design studies in a way that allow monitoring of biomarkers from samples of the same patients in following years. There are many studies showing importance of measuring biomarkers in following years for early diagnosis. In 2013, it was shown that incorporating change over time in a longitudinal screening of serological CA125 biomarker can diagnose more ovarian cancer patients than non-longitudinal single threshold methods. In this study, longitudinal CA125 screening was able to detect aberrant biomarker levels 10 months earlier and at 42% lower concentrations than

single threshold measurements [54]. Importance of longitudinal biomarker studies were also shown for mesothelioma early diagnosis [55-57]. In our previous studies autologous antibodies seem to be very effective in lung cancer diagnosis [58; 59]. Even though there are not many longitudinal studies carried with autologous antibodies, in one of the most important studies it has been shown that longitudinal serum positivity is a clear biomarker for cancer development in patient while sporadic serum positivity is an evidence for cancer to be not progressing (Figure 1.2) [2]. There are many other biomarkers detected that has possible roles in early diagnosis of other cancers, and those biomarkers can be used to detect people require application of preventative therapies [60]. If large longitudinal cohort studies can be designed to detect power of those autologous biomarkers in early diagnosis and detection of cancer recurrence, they can be of use for early diagnosis of lung cancer in conjunction with screening methods.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	N1	N2	N3	N4
1981	1		1		1			1	1	1					
1982	C		1					1	1						
1983			C		C			1							
1984		0				1	1	C	1			1		0	0
1985		1		0		1	1		1			0	1	0	0
1986		C		0		1	1		1			0		1	1
1987				1		C	1		C		1	0	0	0	0
1988							C					0			
1989				C											
1990												0			
1991										C					
1992															
1993														0	
1994															
1995															
1996															
1997															
1998															
1999											C				
2000															
2001													0		0

Figure 1.2: Anti-p53 autologous antibody seropositivity in relation to development of cancer. Anti-p53 autologous-antibody seropositivity followed between 1981 and 2001 in people with asbestos exposure. 11 people with positive serum anti-p53 antibody levels were later diagnosed with cancer (C) (C1,2,4-8,10,11: lung cancer; C3: mesothelioma; C9, lymphoma); but 4 people (N) were not diagnosed with cancer. In 11 cases where antibody positivity (“1”) was observed 1 to 12 years before the cancer diagnosis. Although, measurements were not taken for each year in all of the cases, the positivity (dark color) and negativity (light color) status can be assumed unchanged between two consecutive measurements with the same result. However the cases where negative results were interrupted by sporadic seropositivity were not diagnosed with cancer development (N1-4) [2].

1.3. miRNAs in Cancer

Cancer does not occur and develop purely based on its ability to thrive with genetic mutations, but require epigenetically altering of its proteins to suit the needs [61]. MicroRNAs (miRNAs) are small noncoding RNAs that interfere with translating proteins from mRNAs, thus playing a role in shaping the unique tumor microenvironment [61]. miRNAs and other non-coding RNAs do not stay confined within the cell only, but travel intracellularly and circulate in body fluids [62]. Some miRNAs are shown to have tumor suppressing or oncogenic activities and are important in many regulatory systems [63]. Uptake of miRNAs by other cells have the capability of changing their gene expression pattern, therefore these circulating miRNAs are informative about the condition of their origin like hormones and cytokines are [62]. This makes miRNAs, which can be easily obtained from biological fluids such as blood serum, a good molecule to assess the condition of a patient, a powerful diagnostic and prognostic tool [62; 64].

miRNA signatures, which are defined as a multitude of miRNAs associated with a specific condition when present together, can be used for identifying different types of cancers at their different stages throughout their progression and treatment, especially if new miRNA polymorphisms can be detected [64]. miRNAs can be profiled with more than one method, such as RT-PCR assays and tissue slide-based assays [64]. It is suggested that biomarkers such as miRNAs are necessary tools to be able to provide personalized, targeted therapies towards cancer cases as they can contribute to prognostic assessments and therapeutic targets [65]. Therefore, there have been many extensive studies to characterize and associate miRNA signatures to distinct cancers over the last decade [66], as well as attempts to create systematical tests which can potentially have clinical usage [67].

2. AIMS OF THE STUDY

2.1. Developing a reliable, robust and automated evaluation method to re-evaluate cPA screenings of SCLC patients and healthy controls.

- a) Reliably generating numeric values out of cPA screening images by filtering out low quality data.
- b) Determining autologous antibodies with the highest discriminatory power between SCLC patients and healthy controls.
- c) Determining autologous antibody combinations (panels) with the highest discriminatory power between SCLC patients and healthy controls.

2.2. Developing a method in Quartz Crystal Microbalance to validate cPA screening results.

- a) Demonstrating proof of concept via measuring spike-in anti-SOX2 Antibodies in PBS.
- b) Cross-validating QCM anti-SOX2 Ab measurements in PBS, against ELISA.

2.3. Identification and validation of novel miRNA biomarkers in Non-Small Cell Lung Cancer.

- a) Identifying datasets to be employed in for logistic regression analysis.
- b) Identifying miRNAs with highest discriminatory power between NSCLC patients and high risk controls via logistic regression analysis and LASSO regularization method.
- c) Validating identified miRNAs *in silico*.
- d) Validating identified miRNAs via qRT-PCR.

3. MATERIALS AND METHODS

3.1. RE-ANALYSIS AND RE-EVALUATION OF SCLC AND HEALTHY SERA CUSTOM PROTEIN ARRAY SCREENINGS

3.1.1. Protein Array

Previously in our lab, as a part of the PhD thesis of Sukru Atakan, novel autologous antibody identification experiments were carried in Protein Arrays (PAs). In that study, commercially obtained Testis PA, Fetal Brain (FB) PA and custom PAs (cPAs) were used for serum screening. Production of Testis and FB PAs were carried employing corresponding cDNA libraries. cDNAs from those libraries were inserted in IPTG inducible vectors (pQE80LSN, pQE30NST) which were then expressed in *E. coli* to obtain certain protein producing colonies. These colonies were spotted on the PVDF membrane to form PAs. Testis PA had 26880 clones spotted on 22 x 22 cm PVDF membranes and Fetal Brain PA had 55296 clones spotted on two 22 x 22 cm membranes. Spotting on the membranes were done in 2304 units, where in each unit was 25 circular spots arranged in a 5 by 5 square. Spot in the very middle is called an Inkdot (reference) and the 24 spots around it were containing 12 duplicate clones spotted in an intricate pattern (Figure 3.1). Each spot on a unit is lysates of *E. coli* that inducibly expressed a certain cDNA clone from given cDNA libraries.

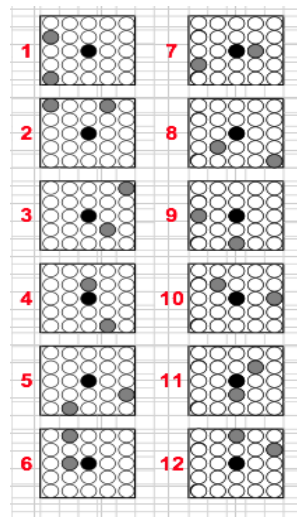


Figure 3.1. 12 distinct protein spotting patterns and a reference InkDot on a PA unit. Pattern info of how 12 different protein clones are spotted around a reference

49 SCLC patient sera and 50 health sera were both separately pooled within their groups. Pooled sera were screened with Testis (26880 clones) and Fetal Brain PAs (55296 clones). Evaluating the results using Manual Selection Software and Signal Analysis Software, clones that were most capable to

distinguish SCLC from healthy controls were identified to be used in generation of custom Protein Arrays (cPAs). Later, all of 49 SCLC patient sera and 50 healthy sera were screened individually using cPAs. Numeric values generated via SAS were used in Monte Carlo analysis to select best panels with the highest sensitivity values at certain specificity values. The best results were 31% sensitivity at 100% specificity incorporating 2 antigens and 46% sensitivity at 90% specificity incorporating 6 antigens[59].

However, the points to improve in that previous study was that the evaluations were lacking elimination of low quality data which may be present due to inadequate serum coverage, wrinkles on array strips and manual recording of cPA screening images or a combination of these issues. Resulting in a possible bias added into further ROC plotting which was done based on that data. Moreover, the generation of ROC curves were made manually and did cover a limited number of antibody combination panels among all possible combinations. With this rationale, we developed a low quality signal elimination method. Array data eliminated with the new method re-evaluated and used to generated individual antibody & combinatory biomarker panels, which then used to plot several ROC curves.

3.1.2. Processing of cPA images in Photoshop

First of all, custom Protein Arrays strips were too long to have screening images taken in a single photograph with high enough resolution. Therefore, each custom Protein Array strip were recorded as 3 separate non-overlapping images (A, B, C sections). As further analysis required them to be joined together, Photoshop was used to merge these 3 sections of a cPA strip to create a single image of each cPA strip. Initially, a canvas is created in Photoshop using shortcuts (CTRL+N). The canvas had following features: Width 152 pixels, Height 1753 pixels, Resolution 72 pixel/inch, Mode: Grayscale / 16 bit, Background Contents: Special, Color Profile: Dot Gain 15%, Pixel Aspect Ratio: Square Pixels. Rectangular Marquee Tool were used to select the individual strip parts from 3 sections of the cPAs (Figure 3.2).

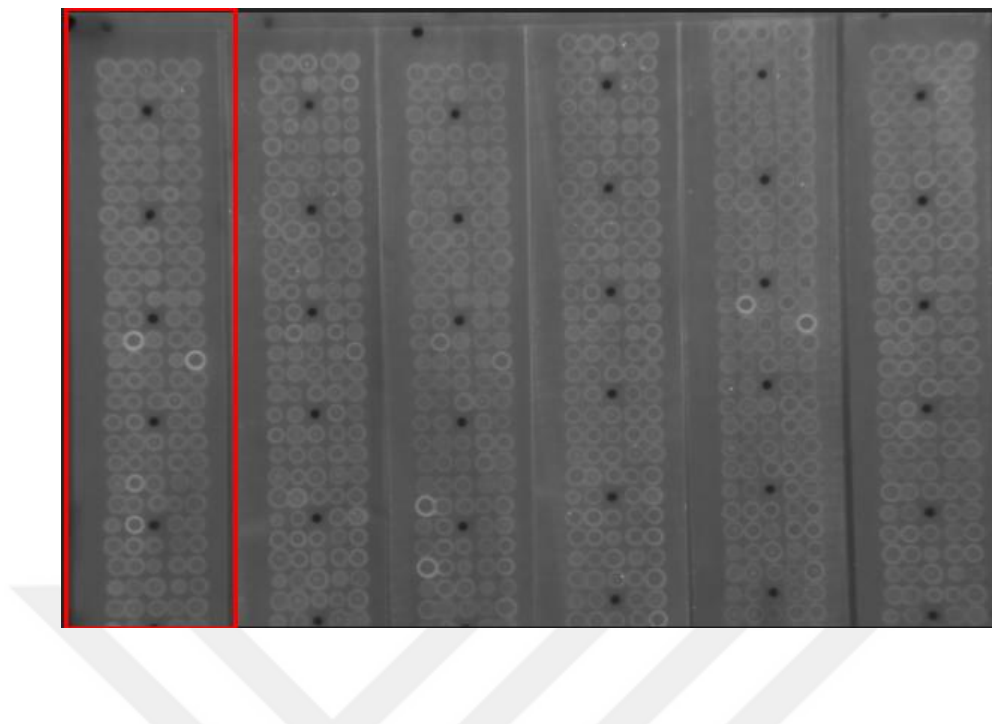


Figure 3.2. Selection of Individual cPA Sections in Photoshop via Rectangular Marquee Tool and an Example of Images Employed in Generation of Unified Images. Using Rectangular Marquee Tool in Photoshop, individual cPA strip sections were copied from A, B, C images. Each of the A, B and C image corresponds to a segment of cPA strips photographed adjacently. A part is shown here in this image.

Selected array piece was copied onto the canvas. This step was repeated for each of the 3 sections (A, B, C) considering their vertical alignment in respective order. Before permanently merging the images in canvas via shortcuts (CTRL + Shift + E), number of Inkdots in each generated image were checked to be 15 to verify correct alignment of 3 sections. After the image is merged, Level Adjustment is carried on the strip image to define the peak signal, amplify the signal intensity and standardize it. This procedure is necessary to make sure signals are comparable between different strips. Following standardization via Level Adjustment, single piece cPA strip image generated on Photoshop was saved in TIF or JPEG format. Given steps in generation of single cPA image from 3 sections were repeated for each and every cPA strip. After each cPA strip has a unified image in Photoshop, all array strips were arranged in a single canvas.

3.1.3. AIDA Signal Analysis Software

Level Adjusted and saved images were opened in AIDA with given display control settings: Quantum Level, Range 0 to 255, Gamma at lowest setting, Show Underexposed (checked). Array area was defined using Grid Definition within the Array Definition section of AIDA tools. With 15 signal

harboring units (with an Inkdot in the middle of each unit) existing on our cPA strips, Grid Definition settings were as given: <no pattern>, Rows: 15, Columns: 1, X Spacing: 0pixel, Y Spacing: 0pixel. With settings adjusted as needed, array area and individual inkdots were defined with avoiding the white background. Within each of 15 signal harboring units another Grid Definition was done to define 24 protein spots + Inkdot in the middle. 5 x 5 ring formation were used with 23 pixels set as the diameter of each ring. After initial settings were done, grids were further adjusted manually to better fit each of the 24 protein spots + Inkdot. This procedure was carried out for all protein spots existing in signal bearing units of every cPA strip. After grid definition was complete, Background Definition was set to Lowest Grid Dots from the drop down menu. Background definition with Lowest Grid Dots allow us to define the background by averaging certain number of lowest signal dots where the number of dots used is defined by our choice. Lastly in settings, integral-background was used for normalization. Finally, pixel intensity levels were measured using Final Calculation in Array Measurement section and saved as excel file to be used in further analysis.

3.1.4. Data Filtration & Generation of Receiver Operating Characteristic Curves

There were low quality signals possibly caused by several reasons such as inadequate sera coverage on cPAs, wrinkling on cPA strips and manual image recording. The low quality sera, named here as “bad data”, were eliminated by a cut-off value filtration. Mean was calculated as average of 360 spots on the cPA strip with 15 signal harboring units. Following standard deviation (StDev) was calculated individually for each of the 15 signal harboring units on the cPA strip. The mean of signal intensities of all spots on the cPA strip was taken as a baseline level in each signal harboring unit. Then mean minus 1.5 StDev of the specific signal harboring unit were used to set the cut-off in that particular signal harboring unit. Applied cut-off was represented by the formula: $(\text{Mean}) - (1.5 \times \text{StDev})$. If a signal harboring unit had spots that fail to pass the set cut-off, all signals in that signal harboring unit were discarded. Set cut-off's efficiency was verified by checking cPAs with eye to see if the the dark areas on the strips were in match with cut-off eliminated areas.

Data collection and filtration were followed by generation of Receiver Operating Characteristic (ROC) curves to test diagnostic capabilities of each antibody. ROC curves for individual antibodies were plotted via R program's “pROC” package. Plotted individual ROC curves were sorted based on their AUC values. 11 antibodies with highest AUC values were selected for plotting combination ROC curves. In the plotting of ROC curves for antibody combinations “protComb” package was used in R. Code for this custom “protComb” package was written by Murat Isbilen. “protComb” package equally weighed all antibodies while generating combination ROC curves for a given antibody combination panel.

3.2. QUARTZ CRYSTAL MICROBALANCE (QCM) METHOD DEVELOPMENT

3.2.1. Serum Collection

For this study sera of 10 SCLC patients and 10 age matched healthy controls were used. For SCLC patients, ethics committee approval was previously obtained from Atatürk Chest Disease and Chest Surgery Education and Research Hospital, Ankara, Turkey. For healthy controls, ethics committee approval was previously obtained from Çapa Chest Diseases and Chest Surgery Education and Research Hospital of İstanbul, Turkey. The previous ethics committee approvals were re-issued and approved by Bilkent Ethics Committee. The diagnosis of SCLC patients was confirmed pathologically. Clinical data of patients, such as gender, age, LDH levels, tumor stage and AP levels, were collected at the time of diagnosis. All of the SCLC patients were on chemotherapy treatment. Moreover, follow up of patients and generation of survival data is carried by clinicians. Serum collection was handled at time of diagnosis for the patients. Controls were assessed to be healthy at the time of serum collection. All sera were mainly aliquoted at 100 mL volumes and at smaller volumes for frequent use. All sera was stored at -80 °C.

3.2.2. ELISA Measurements

SOX2, p18 and DHFR proteins used in ELISA were produced using prokaryotic pQE expression system (Qiagen Inc., Valencia, CA, USA). DHFR served as positive control and p18 served as negative control in the experiments.

For coating, 96 well Immulon 4 HBX Flat Bottom Microtiter Plates (Thermo Scientific, Rochester, NY, USA) were coated with 100 uL of 2 ug/mL purified protein, corresponding to 200 ng protein per well, at +4 °C for 24 hours. Followed by a manual wash step where each well is washed for 6 times using 300 uL of wash buffer (PBS/T with 0.005% Tween20) in each wash. Using 200 uL of blocking buffer (PBS/T with 5% non-fat dried milk and 0.005% Tween20) per well, each well was blocked at +4 °C for 16 hours. Plates were re-washed as mentioned above. Sera diluted to either 1:400 or 1:1600 were used for serum incubation at room temperature for 2 hours.

In the previous studies in our lab, ELISA measurements of individual antibodies with highest AUC (in individual ROCs) were done with in house protocol. However, the determination of positive and negative signal were made based on Absorbance cut-off and no anti-protein IgG concentrations were calculated. Therefore, we generated standard curves of spiked-in anti-SOX2 mAb in varying serum concentrations and PBS. 6 ELISA standard curves were generated by spiking-in anti-SOX2 mAb in

1:400, 1:1600, 1:3200, 1:6400, 1:12800 serum dilutions and PBS buffer. Standard curve dilutions were covering between 200 ng/mL and 0.195 ng/mL. 5 dilution steps were prepared by serially diluting 200 ng/mL with 1/4 dilution steps. Standard curves were compared to each other to see the effects of non-specific binding in each serum dilution. Using unknown dilutions of patient sera, serum anti-SOX2 autologous antibody concentrations of patients with highest sera-positivity were back calculated. Calculated values were used in determination of quartz crystal chip modification parameters.

3.2.3. QCM Chip Modification

AT cut 5 MHz Quartz Chip surfaces were cleaned by pulse sonicating in degassed double distilled water. Procedure was repeated 3 times for 3 minutes (replacing water each time). Sonication process was repeated degassed in analytical grade Ethanol (Sigma-Aldrich) again 3 times for 3 minutes (replacing ethanol each time). After sonication, chips were dried under Nitrogen gas and dipped in Piranha Solution (3:1 mixture of sulfuric acid and 30% hydrogen peroxide) for 10 seconds to etch the surface. Followed by three times degassed double distilled water rinse and 3 times Ethanol rinse. Cleaned chip was immersed in 2mM ethanol solution of 11-Mercaptoundecanoic Acid (11-MUA) for an overnight incubation in the dark at room temperature. After 11-MUA incubation, chips were primed with PBS buffer in QSense Explorer QCM-D device (Biolin Scientific). Flow rate was 27 μ L/sec. Surface is treated with 1:1 mixture of freshly prepared 400mM EDC and 100mM NHS for a duration of 8 min with 27 μ L/sec flow rate (total volume 216 μ L). EDC-NHS treatment immediately followed by 50 μ g/mL SOX2 protein injection (in PBS with pH: 7.4 or 5.5, filter sterilized in 0.2 μ m filter) for 8 min (total volume 216 μ L). Surface was blocked with 50 μ g/mL BSA (in pH: 7.4 PBS) for 8 min (total volume 216 μ L). Unreacted NHS esters were capped with 1M ethanolamine solution (pH: 8.5) treatment for a duration of 8 min. Chips were primed with PBS (pH: 7.4) for a short while before the measurements were carried.

3.2.4. QCM Measurements

Samples for QCM measurement were prepared in PBS solution by spiking-in anti SOX2 mAb (MAB2018) (R&D Systems, Minneapolis, MN, USA). Measurements were taken one after the other with cycles of 8 min sample injection and 15 min PBS injection. Flow rate was set to 27 μ L/min in pump. Signals were recorded and analysis were carried in recorded data using GraphPad Prism 6 (GraphPad Software Inc., La Jolla, CA, USA)

Anti-SOX2 mAb binding (8 min at 27uL/min) to SOX2 protein immobilized Au surface in a concentration range 32 ng/mL and 512 ng/mL were carried and frequency changes on the AT cut 5 Mhz crystal were recorded. Recorded changes are plotted to generate a calibration curve. Measurements were then fitted to 1:1 Langmuir binding model to determine the binding association and dissociation rates. Pearson R correlation analysis were also carried to determine correlation levels between the model and the experimental results.

3.2.5. Surface Characterization via ATR FT-IR

QCM Chip was subjected to several surface modification steps and Attenuated Total Reflection Fourier Transform Infrared Spectroscopy (ATR FT-IR) measurements were taken for two steps of the surface modification: following 11-MUA incubation and following SOX2 protein conjugation to the surface via EDC/NHS amine coupling chemistry. Moreover, ATR FT-IR measurements were also taken after sample binding to the modified chip surface (anti-SOX2 mAb binding).

3.2.6. QCM and ELISA Comparison

Several samples prepared for and analyzed in both ELISA and QCM. Generated data were used to plot standard curves for both methods. Standard curves of QCM were subjected to comparison to ELISA data to evaluate assay performance.

3.3. DISCOVER AND VALIDATION of miRNA BIOMARKERS in NON-SMALL CELL LUNG CANCER

3.3.1. Dataset Selection

All serum, plasma and whole blood GEO Datasets for Non-Small Cell Lung Cancer (NSCLC) between 2008 and 2018 were searched and sorted based on the sample type and platform. 6 datasets were detected. However, else than 2 datasets, GSE24709 and GSE61741, rest of the datasets were all from different platforms from each other. Therefore, for further analysis GSE24709 and GSE61741 were selected and downloaded from GEO. Moreover, all of the tumor tissue sample miRNA sequencing data from the Cancer Genome Atlas Lung Squamous Cell Carcinoma (TCGA-LUSC) and Lung Adenocarcinoma (TCGA-LUAD) were downloaded from National Institutes of Health (NIH). GEO dataset GSE24709 contains febit *Homo sapiens* miRBase 13.0 Microfluidic Primer Extension Assay (MPEA) data of peripheral blood samples drawn from 19 healthy controls, 28 lung cancer patients and

24 chronic obstructive pulmonary disease (COPD) patients as high risk group. GEO dataset GSE61741 contains miRNA microarray data of serum samples drawn from 94 healthy controls, 15 long-lived individuals, and 940 patients affected from various diseases.

3.3.2. miRNA Selection Incorporating Logistic Regression and LASSO Regularization

Differential miRNA expression analysis was conducted between the 19 normal controls of GEO24709, and 17 non-small cell cancer lung cancer (NSCLC) patients of the same dataset. miRNAs with the Benjamini-Hochberg corrected p-value less than 0.05, and log2 fold change over 1.0 were considered to be differentially expressed. The same analysis were conducted between 73 Lung cancer samples, and 94 healthy control samples of GSE61741 dataset. Filters and procedures were identical to the previous analysis. 18 miRNAs were found to be upregulated in cancer group in both of the datasets.

Later, LASSO regression model were used to select best miRNA panel in terms of differentiation power between NSCLC and Healthy Control groups. For this, samples from GEO61741 were assigned into Training and Validation groups. Training group were used to train the logistic regression model. In the training of logistic regression model, LASSO regularization using glmnet() package of R program were used to select miRNAs. miRNAs selected among 18 significantly upregulated miRNAs, and the selection was done on the basis of ability to separate NSCLC patients from Healthy Controls out of 18 significantly upregulated miRNAs. 10 fold cross validation with various λ (lambda) values was conducted, and Area Under Curve (AUC) was used as the error metric. Model with the lowest λ value which lies with the 1 Standard Error of the model with highest mean AUC were chosen. Selected model had 5 miRNAs in the panel. 5 miRNA panel was then used to draw Receiver Operating Characteristics (ROC) curve using the both training and validation datasets generated from GEO61741.

5 miRNAs obtained in the previous analysis were used to train a new model with GSE24709 dataset. 10-fold cross validation over the range of different λ values was conducted on the whole dataset. Mean Squared Error (MSE) was used as the error metric. Model with the lowest λ value which lies within the 1 Standard Error of the model with the lowest MSE was chosen. Selected model had 4 of the miRNAs from previous panel. These 4 miRNAs were then used to draw Receiver Operating Characteristics (ROC) curve by using GSE24709 expression data.

3.3.3. Serum Samples

For this study sera of 7 Non-Small Cell Lung Cancer (NSCLC) patients and age, gender and smoking status (package day x years) matched High Risk Group Controls (HRG) were used. Mean age was 64.86 for NSCLC patients and 63.14 for HRG group. Mean smoking level (package day x years) was 38.00 for NSCLC patients and 41.83 for HRG group. 6 of the NSCLC patients and their matching HRG controls were male, while 1 NSCLC patient and her matching HRG control was female. Moreover, all patients, except one, had lung lobectomy operation (excision of the lung part containing limited tumor) prior to serum collection. For the one patient with no surgery, serum collection was done prior to bronchoscopy. Diagnosis of the patients were confirmed pathologically and patients were identified as either adenocarcinoma or squamous cell carcinoma. Controls were assessed to be healthy at the time of serum collection. All sera were stored in bulk and were stored at -80 °C.

3.3.4. Hemolysis Assessment

Hemolysis status of sera were assessed via oxy-hemoglobin peaks at 414 nm. Absorbance < 0.072 filter at 414 nm were used for low hemolysis risk, $0.072 \leq \text{Absorbance} \leq 0.3$ filter at 414 nm were used for moderate hemolysis risk and Absorbance > 0.3 at 414 nm were used for high hemolysis risk [68]. Measurements in 414 nm wavelength were taken with NanoDrop One/One C Microvolume UV–Vis Spectrophotometer (Thermo Fisher Scientific, Fitchburg, WI).

3.3.5. RNA Extraction and Quantification

For miRNA isolation from sera miRNeasy Serum/Plasma Advanced Kit (Qiagen Inc., Valencia, CA, USA) (Cat. No: 217204) were utilized. miRNA isolations were carried according to the manufacturer's recommendations. For each isolation 200 uL serum were used. Isolated RNA is re-suspended in 20uL Nuclease Free Water (Ambion, Carlsbad, CA, USA) and stored at -20 °C for a day before cDNA synthesis. To determine quantity and quality of the RNA isolates, measurements were taken with NanoDrop One/One C Microvolume UV–Vis Spectrophotometer (Thermo Fisher Scientific, Fitchburg, WI) before they are stored at -20 °C.

3.3.6. cDNA Generation and Quantification

cDNA synthesis reaction from extracted RNA were done using miRCURY LNA miRNA RT Kit (Qiagen Inc., Valencia, CA, USA) (Cat. No: 339340) according to the manufacturer's

recommendations. To determine quantity and quality of the synthesized cDNA measurements were taken with NanoDrop One/One C Microvolume UV–Vis Spectrophotometer (Thermo Fisher Scientific, Fitchburg, WI) before they were used in qPCR analysis. (-) Reverse Transcription (RT) and (+) RT controls were used. (-) RT controls are generated using hsa-miR-103a-3p internal reference primer and isolated miRNAs from H1, H2, N1, N2 samples. (+) RT control were lacking isolated miRNA.

3.3.7. Quantitative Real Time Polymerase Chain Reaction (qRT-PCR)

Primers for target miRNAs (hsa-mir-130b-5p, hsa-miR-132-5p, hsa-let-7d-3p, hsa-miR-1227-3p, hsa-miR-199a-5p) and primers for internal controls (hsa-miR-16-5p and hsa-miR-103a-3p) were re-suspended in 220 uL of Nuclease Free Water (Ambion, Carlsbad, CA, USA) and stored at -20 °C until use.

miRCURY LNA SYBR Green PCR Kit (4000) (Qiagen Inc., Valencia, CA, USA) (Cat. No: 339347), clear/white low profile, 96 well plates (Bio-Rad, USA) (Cat. No: HSR9905) and LightCycler 480 II (Roche, USA) were used for performing qRT-PCR experiments. SYBR Green was used as the reporter. Every miRNA were run in triplicate for each sample for a total of 14 samples. 7 Samples were NSCLC patients with serums collected before thoracic surgery, and 7 samples were their matched high risk controls. Each reaction consisted of 5 uL miRCURY LNA SYBR Green Master Mix, 3 uL of 1:60 diluted cDNA, 1 uL miRNA primer, 1uL Nuclease Free Water.

Reaction steps consisted of first incubation at 95 °C for 10 min, followed by 50 cycles of incubation at 95 °C for 15 seconds and 56 °C for 1 min. Melting curve analysis were carried between 56 °C and 90 °C. In every PCR plate, for every miRNA, one template control with three replicates was also run.

Hsa-miR-16-5p was used as endogenous reference control for normalization by manually calculating Ct values out of $\Delta\Delta C_t$ Method.

3.3.8. Data Analysis

Data analysis were done using hsa-miR-16-5p as endogenous reference control for normalization. Ct values calculated with $\Delta\Delta C_t$ Method. Graphpad Prism 6 (GraphPad Software Inc., La Jolla, CA, USA) program were used for calculations and all plotting.

4. RESULTS

4.1. RE-ANALYSIS AND RE-EVALUATION OF SCLC AND HEALTHY SERA CUSTOM PROTEIN ARRAY SCREENINGS

4.1.1. Protein Array

In this part of the study, we tried to re-analyze a previously conducted experiment in pursuit of improving evaluation quality and resulting ROC curves. For this we used images of cPAs used in screening of 49 SCLC patients and 50 healthy sera in that experiment conducted by PhD. Sukru Atakan.

4.1.2. Processing of cPA images in Photoshop

cPA images used in screening of 49 SCLC patients and 50 healthy controls were processed in Photoshop to generate single piece images. These images later on grouped in a single Photoshop canvas. Both the single cPA images and grouped cPA image, seen in Figure 4.1, used in AIDA software analysis to generate numerical readings out of cPA screening images.

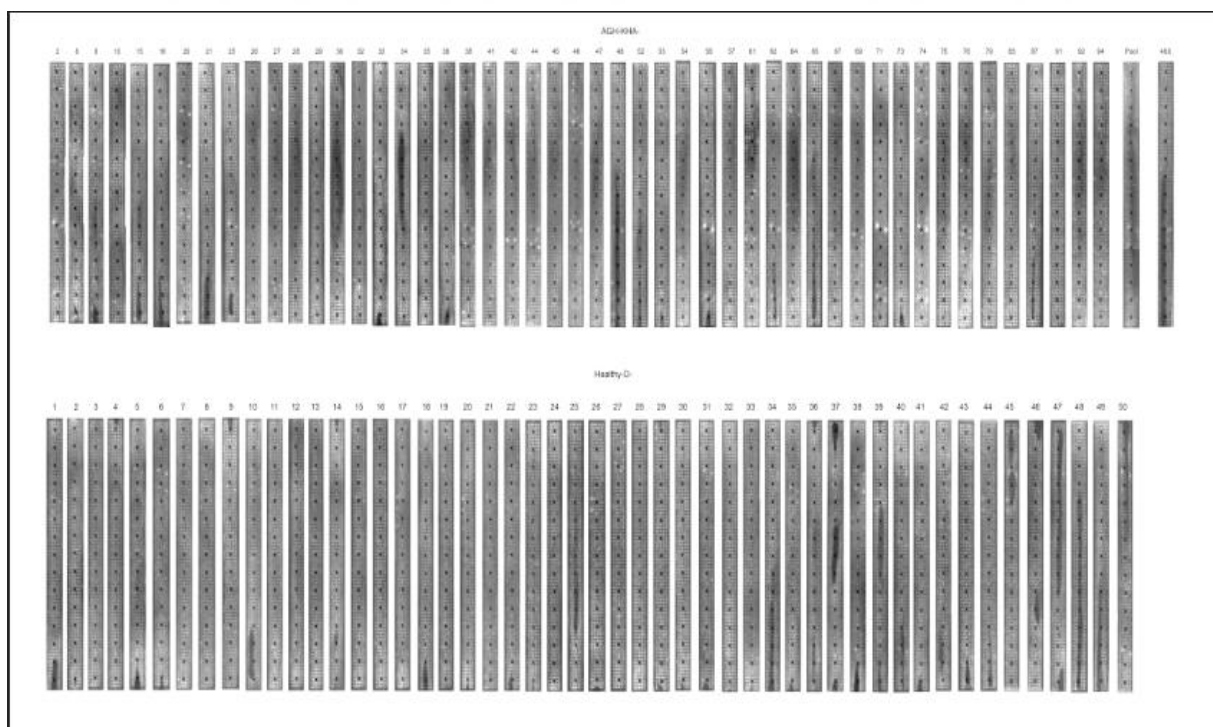


Figure 4.1: All cPA strips used in screening of SCLC patients and Healthy Controls. 3 sections of each cPA strip is joined together and lined up side by side via Photoshop.

4.1.3. Data Filtration & Generation of Receiver Operating Characteristic Curves

Bad data was a potential problem in evaluating the signals. Because lowest intensity spots were used to set background (noise) levels. Also, in control PAs the lower background values lead to higher false positivity, verifying the problem with low intensity signals. Elimination of bad data were done through cut-off value filtration. In the decision of keeping or eliminating a certain signal harboring unit, overall mean of that cPA strip and StDev of the corresponding signal harboring unit were used. $(\text{Mean}) - (2.5 \times \text{StDev})$ was the first cut-off to be tried, however, this approach resulted in elimination of positive signals together with bad signals (shady areas). Therefore, $(\text{Mean}) - (1.5 \times \text{StDev})$ was tried which successfully eliminated signal harboring units with bad data (Figure 4.2). The remainder of sections used in ROC curve generation as well as eliminated signal harboring units were shown in Figure 4.3.

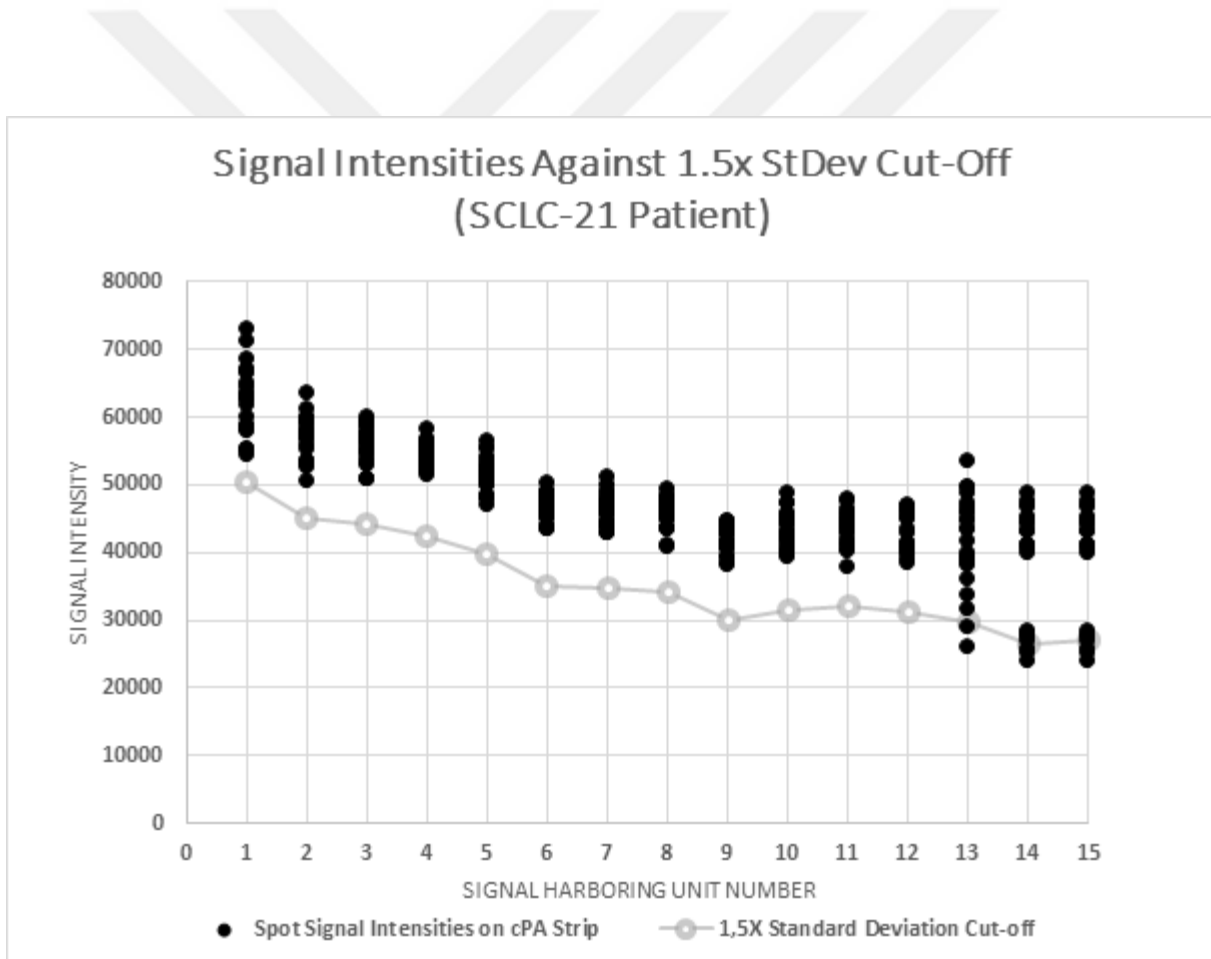


Figure 4.2 An Example of Signal Harboring Unit Elimination Using 1.5x StDev Cut-off. Gray line indicates the 1.5x StDev Cut-off calculated for every signal harboring cell. If a signal harboring cell had signal readings below the decided cut-off, as in the cases of 13th, 14th and 15th cells in here, the entire set of data in that signal harboring unit was discarded as “bad data”.

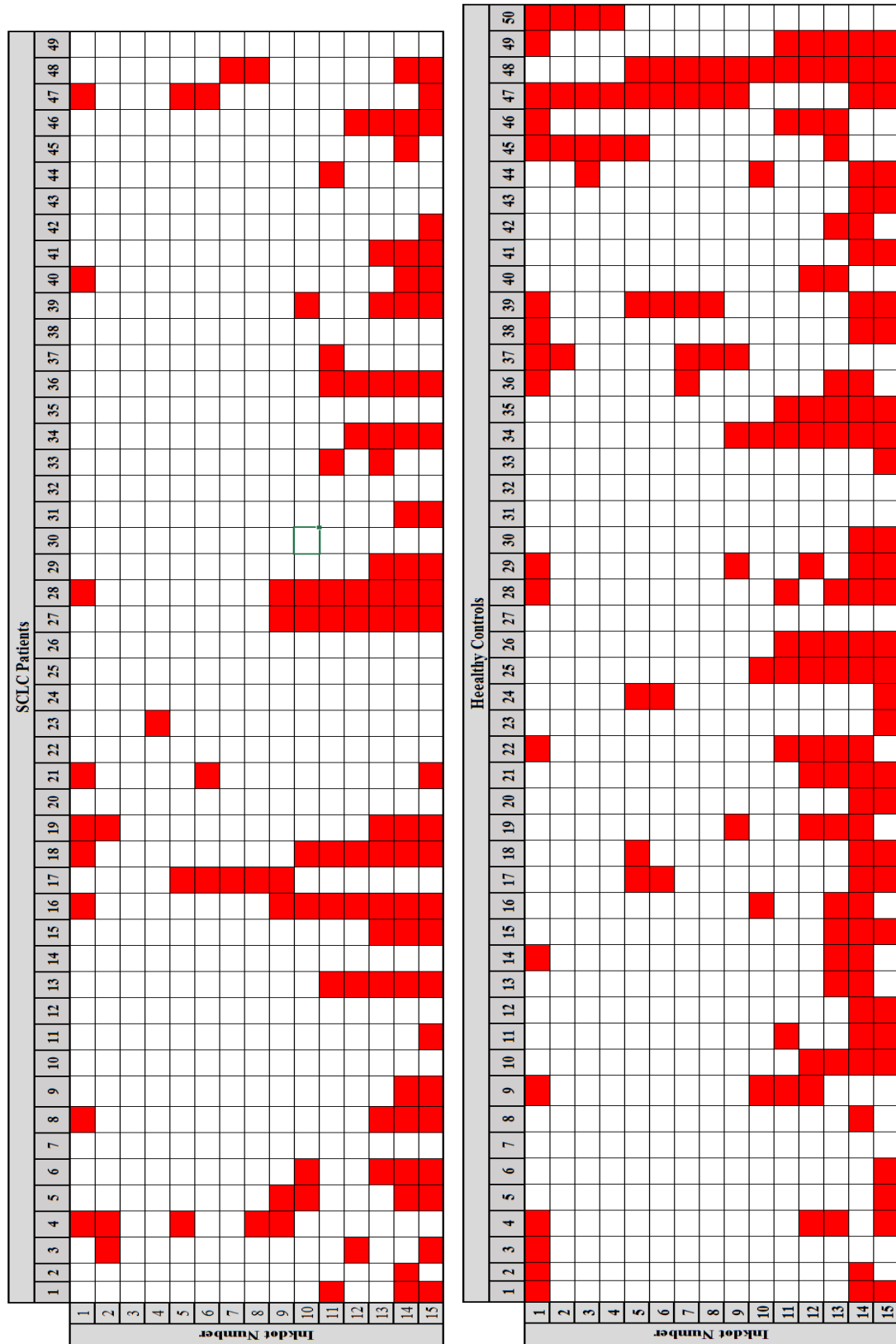


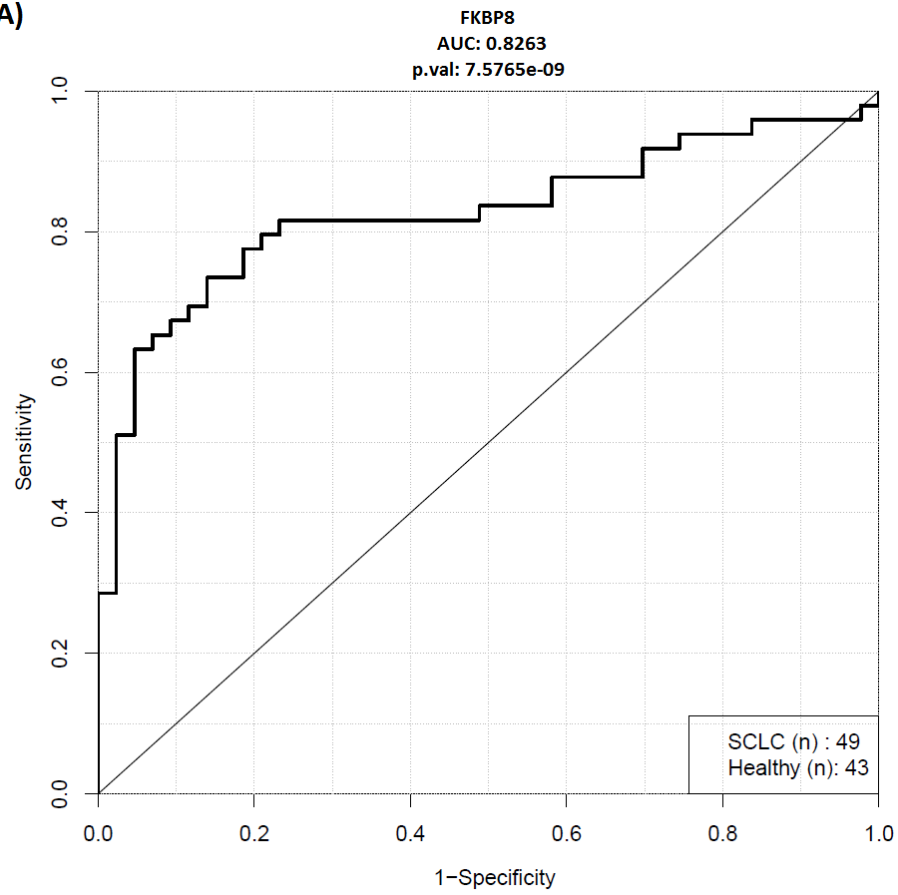
Figure 4.3: Pattern of discarded and analyzed signal harboring units in each cPA strip. Horizontal lines rows represents individual cPA strips and vertical lines represents signal harboring unit number (named as “Inkdot Number”). Darker boxes indicate the signal harboring cells which were discarded after 1.5x StDev Cut-off. The layout at the left hand side represents SCLC patients and the layout at right hand side represents Healthy Controls.

After cut-off filtration remaining data were analyzed in “pROC” package of R program. 180 ROC curves were generated based on signal intensities belonging to each of the 180 proteins spotted on cPAs. ROC results were ranked based on their AUC values. Best 11 ROC curve results based on their AUC are shown in Table 4.1. All 11 of highest AUC individual antibody ROC curves had significant p-values and top 8 individual antibody ROC curves had highly significant p-values. Best AUC was seen in individual ROC of FKBP8 with 82.6%. Variation of sample size among antibodies was a result of “bad signal” elimination process which eliminated varying number of samples for each antibody. Among 180 individual ROC curves, plots of 4 with the highest AUC values can be seen in Figure 4.4.

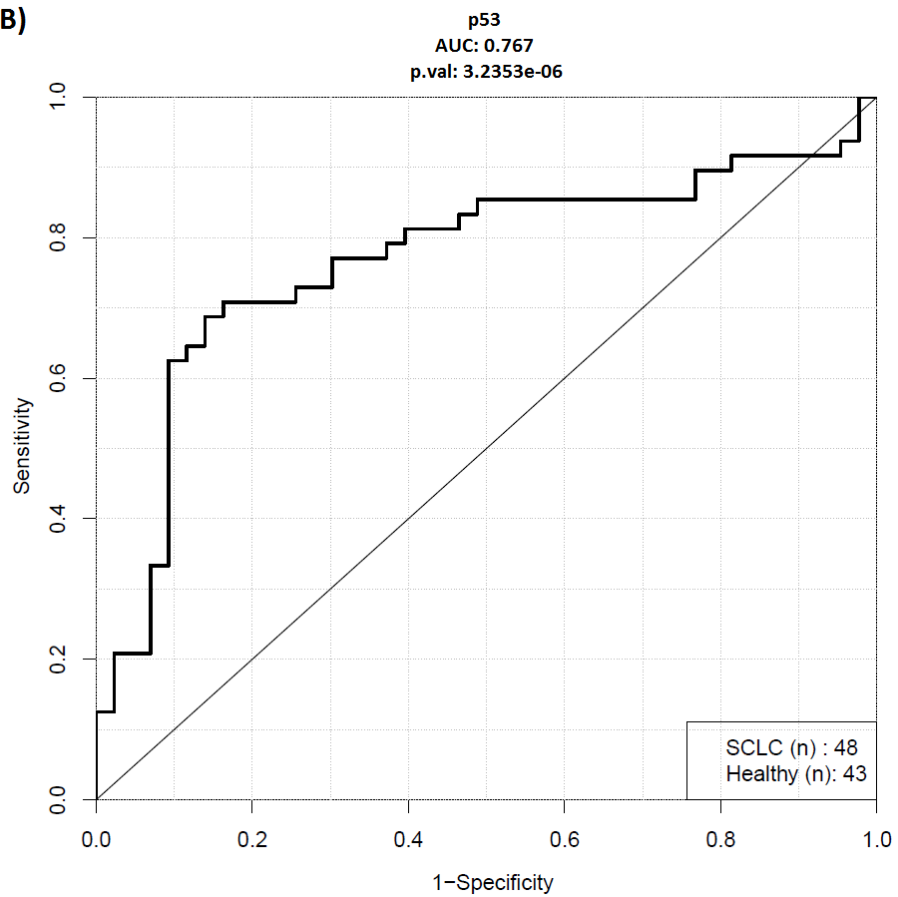
Table 4.1: 11 Best individual autologous-antibody ROC curve results according to AUC values.

Antibody Name	AUC	p-value	Total Sample Size	SCLC (n)	Healthy (n)
FKBP8	0,826	7,58E-09	92	49	43
SOX2	0,818	1,25E-07	81	42	39
POLB	0,780	9,59E-06	74	39	35
p53	0,767	3,24E-06	91	48	43
GON4L	0,760	7,90E-06	88	46	42
ENSA	0,760	9,82E-06	87	47	40
TUBB4	0,711	0,000271499	88	46	42
TIMM44	0,704	0,000539027	85	46	39
PPIA	0,650	0,009871125	81	42	39
HuD	0,643	0,020907917	70	40	30
DUSP8	0,636	0,012359584	92	49	43

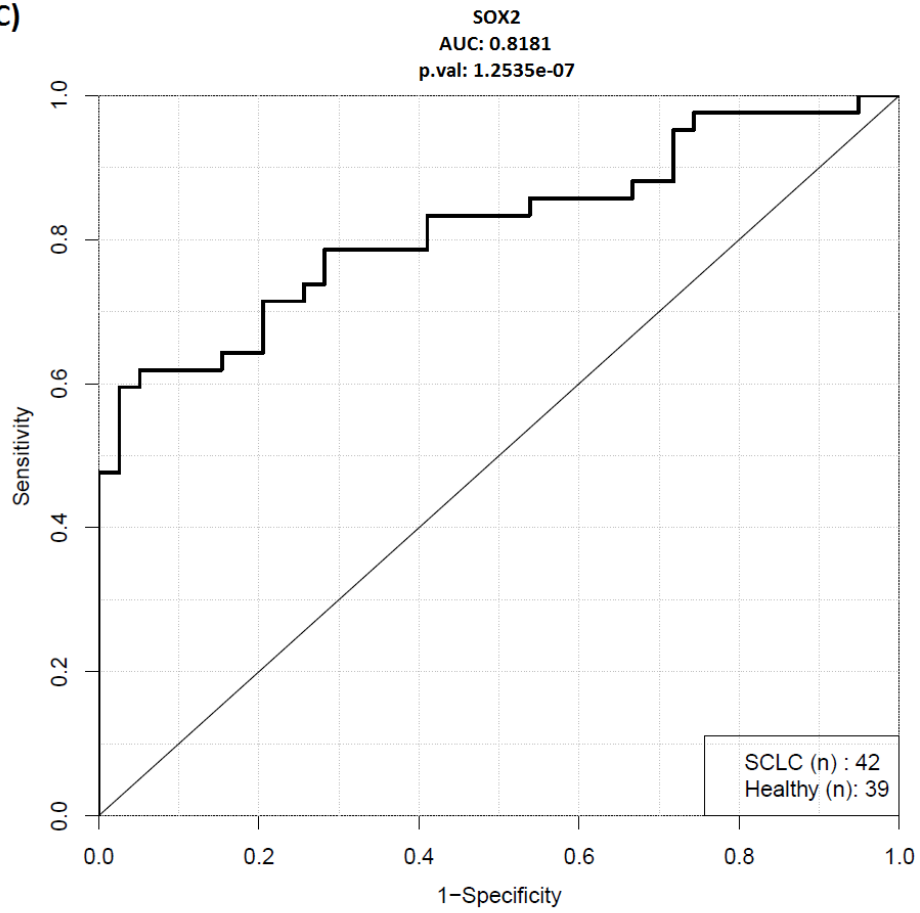
A)



B)



c)



d)

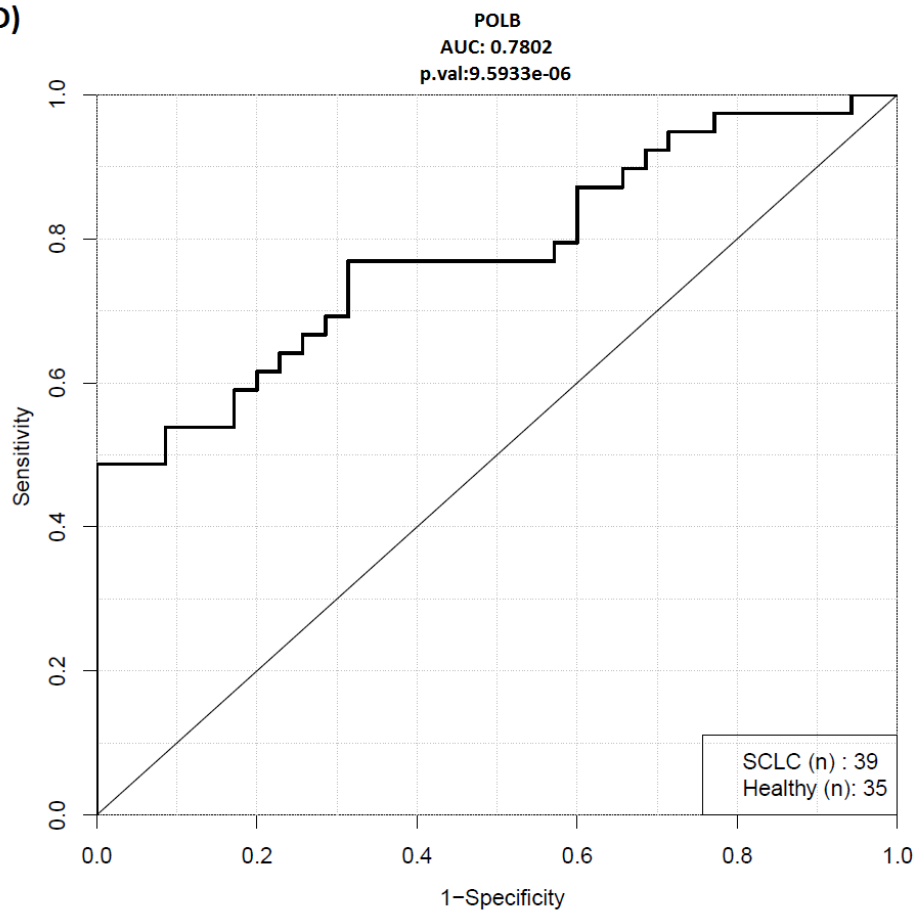


Figure 4.4: Individual antibody ROC curves with highest AUC values. Signal intensities obtained from custom Protein Array results via AIDA software were used to plot ROC curves of each antibody. Antibodies with highest AUC were selected to be used in plotting of combination ROC curves. Antibodies with highest AUC values in their ROCs were against proteins: **A)** FKBP8, **B)** p53, **C)** SOX2 and **D)** POLB.

Following individual ROCs, antibodies with highest AUC in their ROC curves (Table 4.1) were used in combination ROC curves. In plotting of ROC curves for antibody combinations, all antibodies were equally weighted. 2036 combination ROC curves were plotted for combined antibody panels. Plotted ROCs were sorted based on their AUC and a filtration of $AUC \geq 90\%$ were applied. 1016 ROC curves had AUC values higher than 90%. 1016 ROC curves that passed AUC filtration were further sorted on total sample size and $n < 60$ were discarded. This filtration process left 517 combination ROC curves, top 100 of these ROC curve results can be seen as a list in Supplementary Table 1.

For further analysis and validation with QCM in the future, a panel of 4 antibodies (Figure 4.5) were selected on the basis of, novelty, and individual antibody ROC results (AUC). Selected antibody panel had autologous antibodies against: SOX2, FKBP8, POLB and P53 proteins.

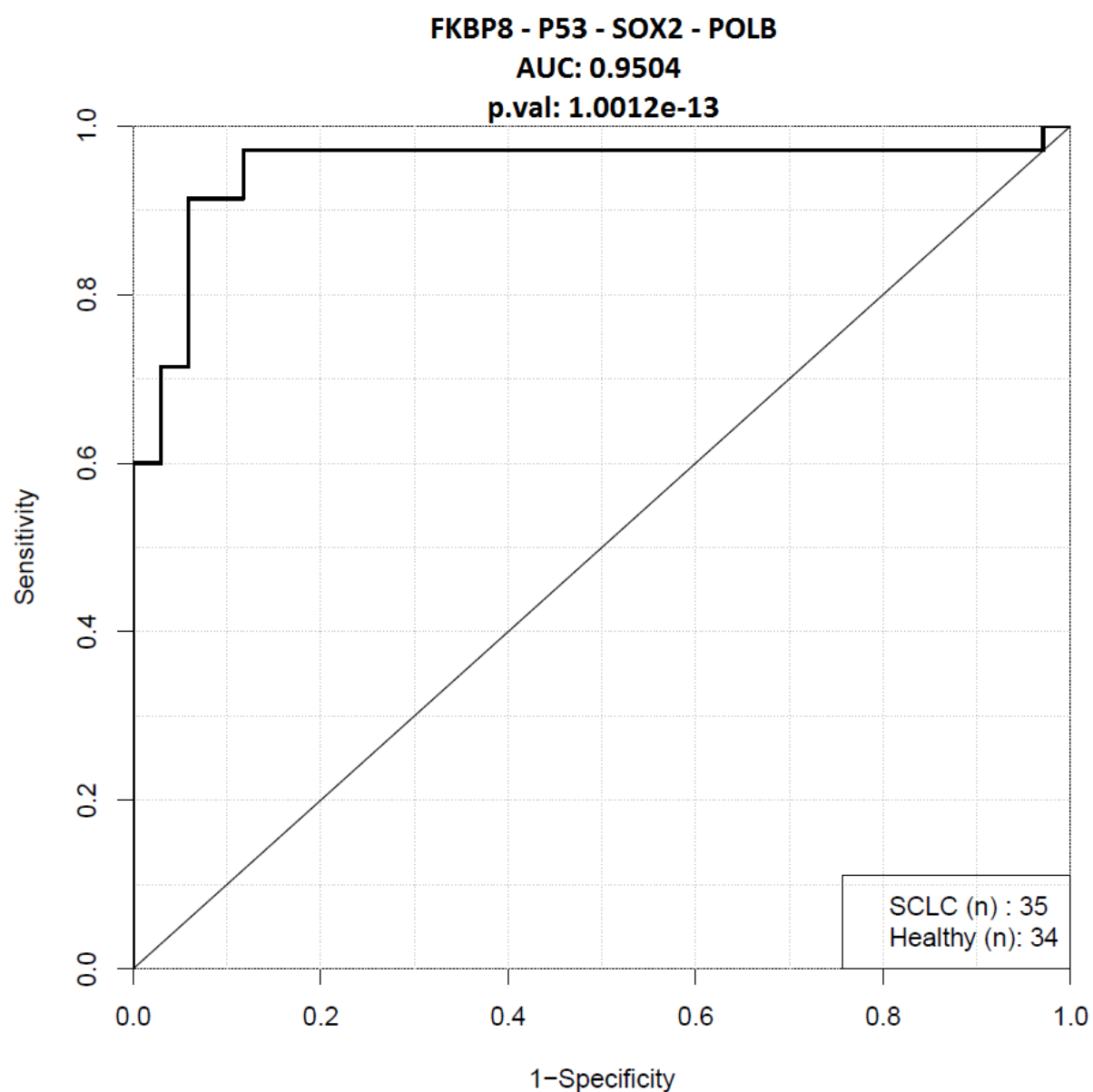


Figure 4.5: Combination ROC curve of proteins selected for QCM analysis. Signal intensities obtained from custom Protein Array results via AIDA software were used to plot ROC curves of protein combinations.

4.2. QUARTZ CRYSTAL MICROBALANCE METHOD DEVELOPMENT

4.2.1. ELISA Measurement

To evaluate effect of non-specific binding on standard curves, 6 ELISA standard curves were generated by spiking-in anti-SOX2 mAb in 1:400, 1:1600, 1:3200, 1:6400, 1:12800 serum dilutions and PBS buffer (Figure 4.6). Standard curve dilutions were covering a range between 200 ng/mL and 0.195 ng/mL anti-SOX2. 5 dilution steps were prepared by serially diluting 200 ng/mL anti-SOX2 with 1/4 dilution steps. All of the standard curves generated in serum dilutions and PBS were highly correlating with each other.

Pearson R values						
x	1/400 serum dilution	1/1600 serum dilution	1/3200 serum dilution	1/6400 serum dilution	1/12800 serum dilution	PBS
1/400 serum dilution		0.9998609	0.9996663	0.9997456	0.9987338	0.9960713
1/1600 serum dilution	0.9998609		0.9999375	0.9999725	0.9993914	0.9973169
1/3200 serum dilution	0.9996663	0.9999375		0.9999413	0.9995915	0.9977686
1/6400 serum dilution	0.9997456	0.9999725	0.9999413		0.9995671	0.9976128
1/12800 serum dilution	0.9987338	0.9993914	0.9995915	0.9995671		0.9991266
PBS	0.9960713	0.9973169	0.9977686	0.9976128	0.9991266	

P Values						
x	1/400 serum dilution	1/1600 serum dilution	1/3200 serum dilution	1/6400 serum dilution	1/12800 serum dilution	PBS
1/400 serum dilution		2.90099E-08	1.67018E-07	9.70862E-08	2.40373E-06	2.31214E-05
1/1600 serum dilution	2.90099E-08		5.8498E-09	1.13484E-09	5.55427E-07	1.07889E-05
1/3200 serum dilution	1.67018E-07	5.8498E-09		5.16594E-09	2.50209E-07	7.46344E-06
1/6400 serum dilution	9.70862E-08	1.13484E-09	5.16594E-09		2.8109E-07	8.54166E-06
1/12800 serum dilution	2.40373E-06	5.55427E-07	2.50209E-07	2.8109E-07		1.14397E-06
PBS	2.31214E-05	1.07889E-05	7.46344E-06	8.54166E-06	1.14397E-06	

ELISA Standard Curves

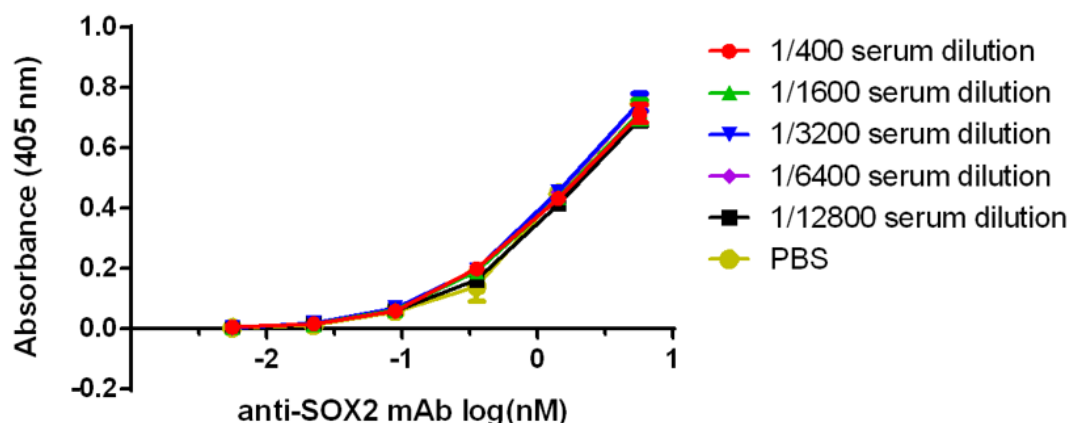


Figure 4.6 ELISA anti-SOX2 Standard curves prepared in different serum dilutions and PBS buffer & Pearson R correlation of each standard curve. All of the standard curves generated for anti-SOX2 antibody were highly correlating with each other, indicated out by Pearson R values close to 1 and highly significant p. Values. This indicates standard curves prepared in 1:400 serum dilution, where the validation experiments were done previously, can be compared to standard curves prepared in lower serum dilutions and pure PBS.

4.2.2. QCM Chip Modification

Chip modification were started via immersion of clean QCM chip in 2mM Ethanol solution of 11-MUA for SAM formation. QCM chip with SAM was primed in buffer treated with EDC/NHS, SOX2 protein, blocking buffer and ethanolamine, respectively. Signal changes throughout the procedure were recorded (Figure 4.7).

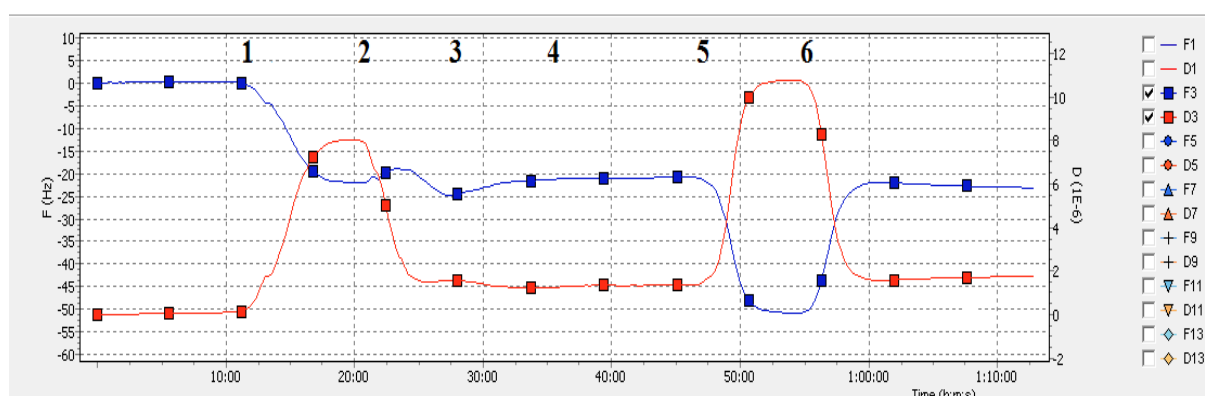


Figure 4.7 QCM chip measurements throughout the surface modification. Major events are number coded vertically on top of the timepoint. **1)** EDC:NHS injection, **2)** SOX2 protein (50 ug/mL in pH 5.5 PBS) injection, **3)** PBS (pH: 7.4), **4)** Blocking Buffer (50ug/mL BSA in PBS), **5)** Ethanolamine, **6)** PBS (pH: 7.4)

In SOX2 conjugation to the surface two pHs (7.4 and 5.5) were tried for SOX2 protein for optimization of amine coupling chemistry. Conjugation reaction carried using SOX2 protein re-suspended in pH 7.4 resulted in insufficient frequency change on the surface indicating inadequate protein conjugation. Therefore, no significant anti-SOX2 binding been observed on the chip surface modified with SOX2 at pH 7.4. However, SOX2 protein re-suspended in pH 5.5 generated significant frequency change during surface modification, indicating substantial SOX2 conjugation onto the surface, which is later proven by anti-SOX2 bindings.

4.2.3. QCM Measurement

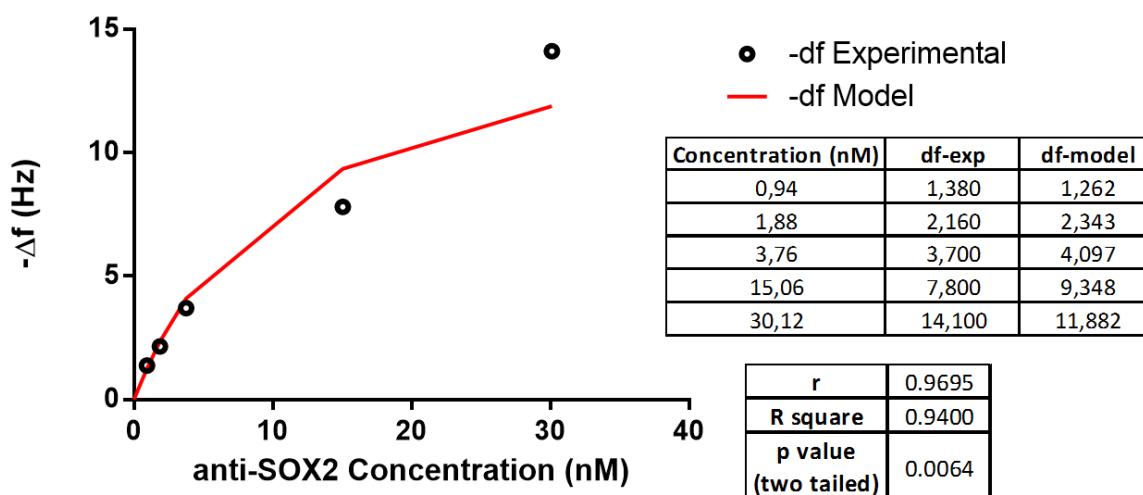


Figure 4.8: QCM measurements of anti-SOX2 mAb binding on SOX2 protein immobilized on AT-cut 5MHz QCM chip. Anti-SOX2 mAb (8 min at 27 uL/min) binding measurements on SOX2 protein immobilized surface in a concentration range 32 ng/mL and 512 ng/mL fitted on 1:1 Langmuir isothermal interaction model.

Anti-SOX2 mAb binding (8 min at 27 uL/min) to SOX2 protein immobilized Au surface in a concentration range 32 ng/mL and 512 ng/mL were carried and frequency changes on the AT cut 5 Mhz crystal were recorded. Recorded changes are plotted to generate a calibration curve. Measurements were then fitted to 1:1 Langmuir isothermal binding model to determine the binding K_D value. K_D was calculated as 11.2 nM. Moreover, Pearson R correlation analysis between the measurements and fitted 1:1 Langmuir isothermal binding model curve were carried. Pearson correlation analysis resulted in a R square value of 0.94 with a significant two-tailed p value 0.0064. QCM measurements, 1:1 Langmuir isothermal binding model curve, Pearson correlation analysis results and raw experimental data can be seen in Figure 4.8.

4.2.4. Surface Characterization via ATR FT-IR

Although, the binding chemistry during the surface modification is widely used in the literature, surface characterization was done via ATR FT-IR to check chemical changes on the chip surface throughout the protein immobilization steps, illustrated in Figure 4.9, as well as anti-SOX2 binding to SOX2 immobilized QCM chips.

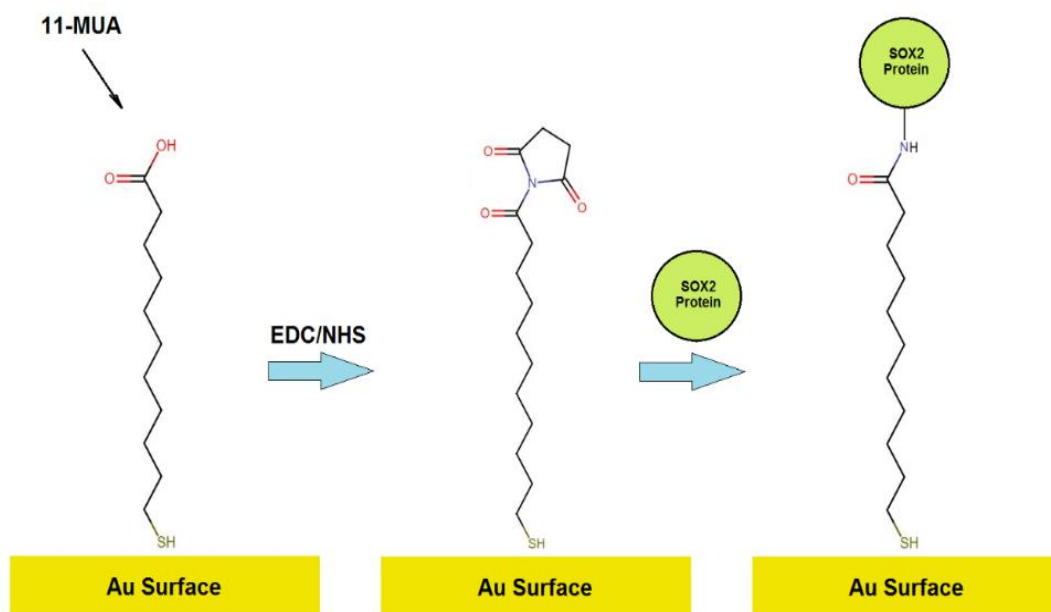


Figure 4.9: Illustration of chip surface modification. Chip surface modification can be simplified into 3 steps: Self-assembled Monolayer (SAM) formation via polymerization of 11-MUA on chip surface, EDC/NHS treatment to create reactive groups on carboxyl end of 11-MUA and lastly protein (SOX2) covalent conjugation to surface through primary amine groups on the protein.

Peaks observed in ATR FT-IR analysis and their associated chemical background is given in Table 4.2. For 11-MUA self-assembled monolayer (SAM) modified chip surface, fingerprints of CH₂, COOH, C=O groups were predominantly observed as listed in Table 4.2, for which the full spectrum graph can be seen in Figure 4.10.

Table 4.2: Signal peaks obtained after chip surface characterization with ATR FT-IR and their attributed chemical structure.

11-MUA Modified Chip Surface	
Wavelength (nm) Peak Observed	Associated Chemical Bonds
2921	CH ₂ asymmetric stretching [69]
2850	CH ₂ symmetric stretching [69]
1724	-COOH groups of free or non-hydrogen-bonded and side-by-side dimeric hydrogen bonded modes [70; 71]
1700-1600	C=O stretching [72; 73]
1466	(C(=O)-O ⁻) carboxylate stretching [73]
1264	CH ₂ rocking and wagging [74]
1120	CO stretching
1074	C-C/C-O stretching vibration
SOX2 Conjugated Chip Surface	
Wavelength (nm) Peak Observed	Associated Chemical Bonds
2957	CH ₃ asymmetric stretching
2920	CH ₂ asymmetric stretching [69]
2870	CH ₃ symmetric stretching
2850	CH ₂ symmetric stretching [69]
1727	-COOH groups of free or non-hydrogen-bonded and side-by-side dimeric hydrogen bonded modes [70; 71]
1700-1600	C=O stretching [72; 73; 75]
1550	N-H bending and C-H stretching [75]
1465	(C(=O)-O ⁻) carboxylate stretching [73]
1379	-OH bending [75] C-H deformation
1262	CH ₂ rocking and wagging [74] and PO ₂ asymmetric stretching

1121	C-O stretching
1075	PO ₂ symmetric stretching
SOX2 Conjugated Chip Surface After Anti-SOX2 Binding	
Wavelength (nm) Peak Observed	Associated Chemical Bonds
2956	CH ₃ asymmetric stretching
2918	CH ₂ asymmetric stretching [69]
2871	CH ₃ symmetric stretching
2849	CH ₂ symmetric stretching [69]
1728	-COOH groups of free or non-hydrogen-bonded and side-by-side dimeric hydrogen bonded modes [70; 71]
1700-1600	C=O stretching [72; 73; 75]
1546	N-H bending and C-H stretching [75]
1466	(C(=O)-O-) [73]
1378	-OH bending [75] C-H deformation
1263	CH ₂ rocking and wagging [74] and PO ₂ asymmetric stretching
1121	C-O stretching
1074	PO ₂ symmetric stretching

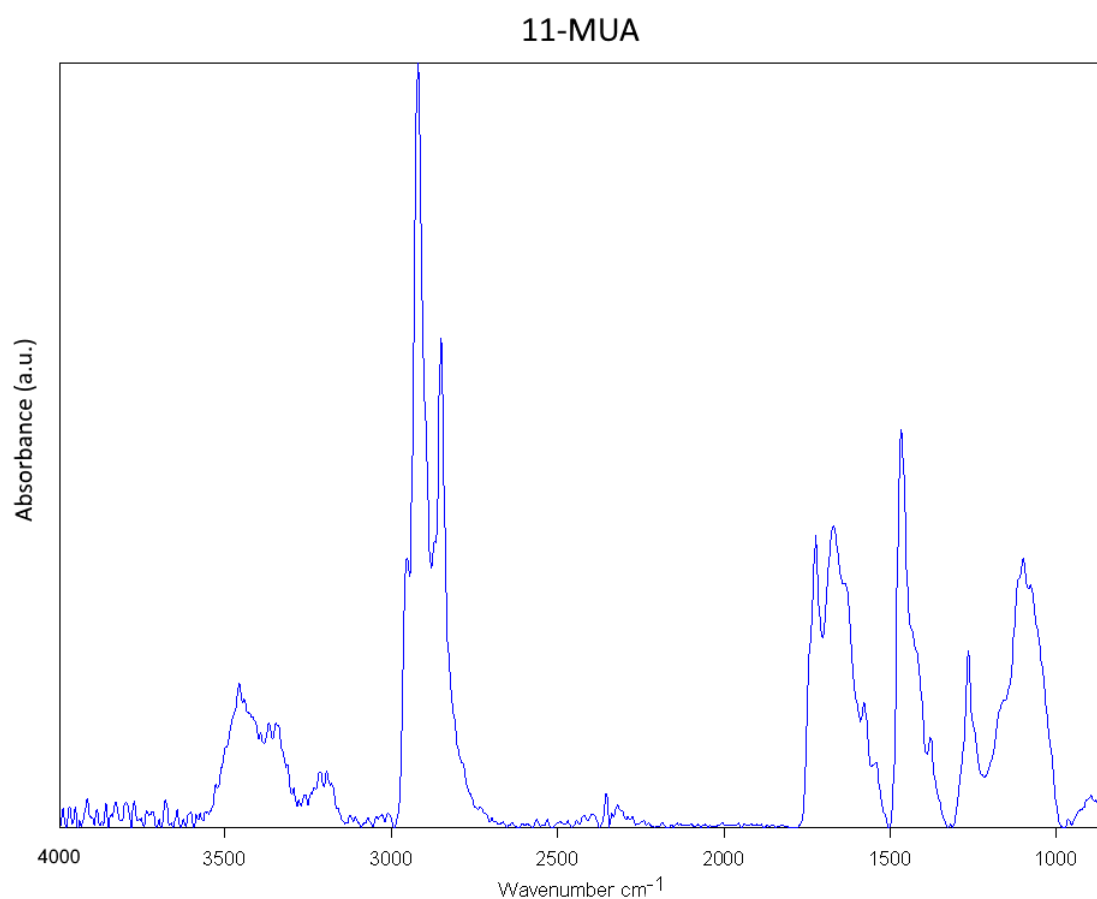


Figure 4.10: Absorbance spectrum of 11-MUA polymerized QCM chip surface measured via ATR FT-IR. Observed peak(s) at 2921 and 2850 indicate CH₂, 1724 indicate COOH group, 1700-1600 range indicate C=O bond.

For SOX2 immobilized chip surface, fingerprints of CH₃, CH₂, COOH, C=O and N-H groups were predominantly observed as listed in Table 4.2, for which the full spectrum graph can be seen in Figure 4.11.

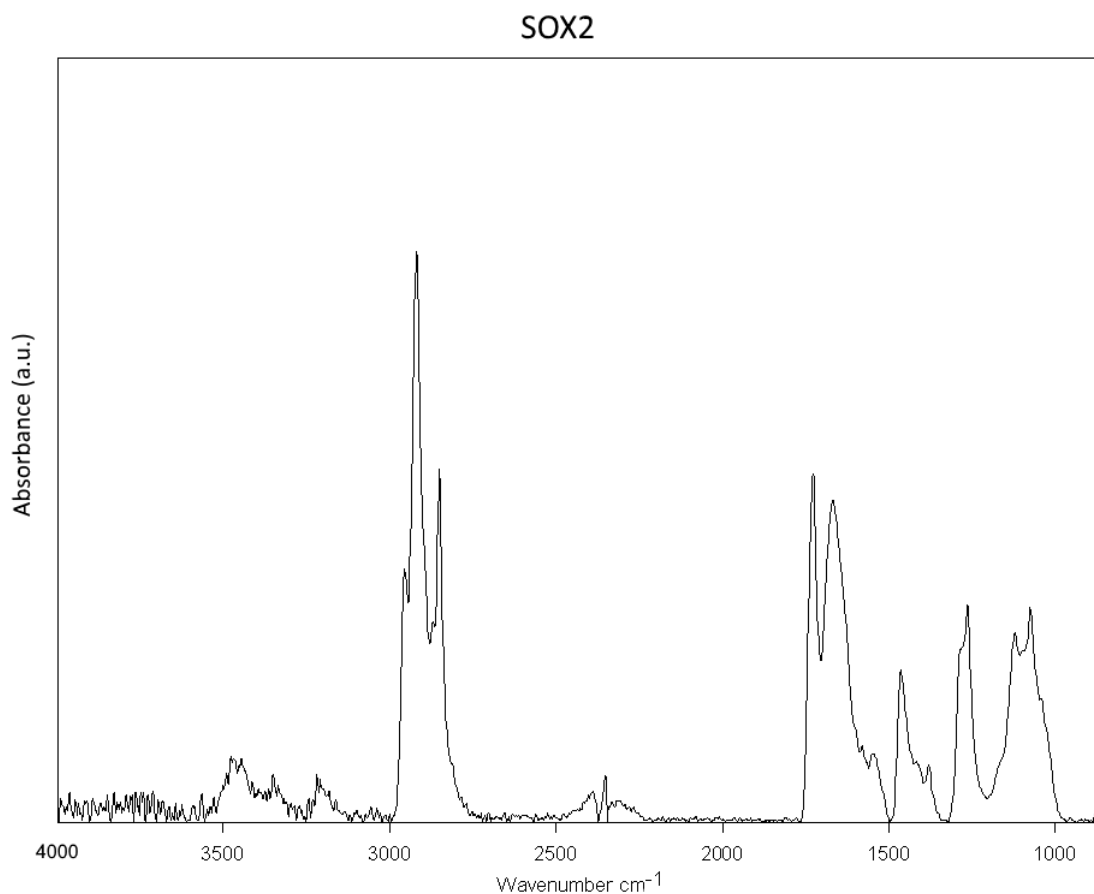


Figure 4.11: Absorbance spectrum of SOX2 immobilized QCM chip surface measured via ATR FT-IR. Observed peak(s) at 2957 indicate existence of CH₃, 2920 and 2870 indicate CH₂, 1727 indicate COOH, 1700-1600 range indicate C=O and 1550 indicate NH groups.

For anti-SOX2 sample bound chip surface, fingerprints of again CH₃, CH₂, COOH, C=O and N-H groups were predominantly observed as listed in Table 4.2, for which the full spectrum graph can be seen in Figure 4.12.

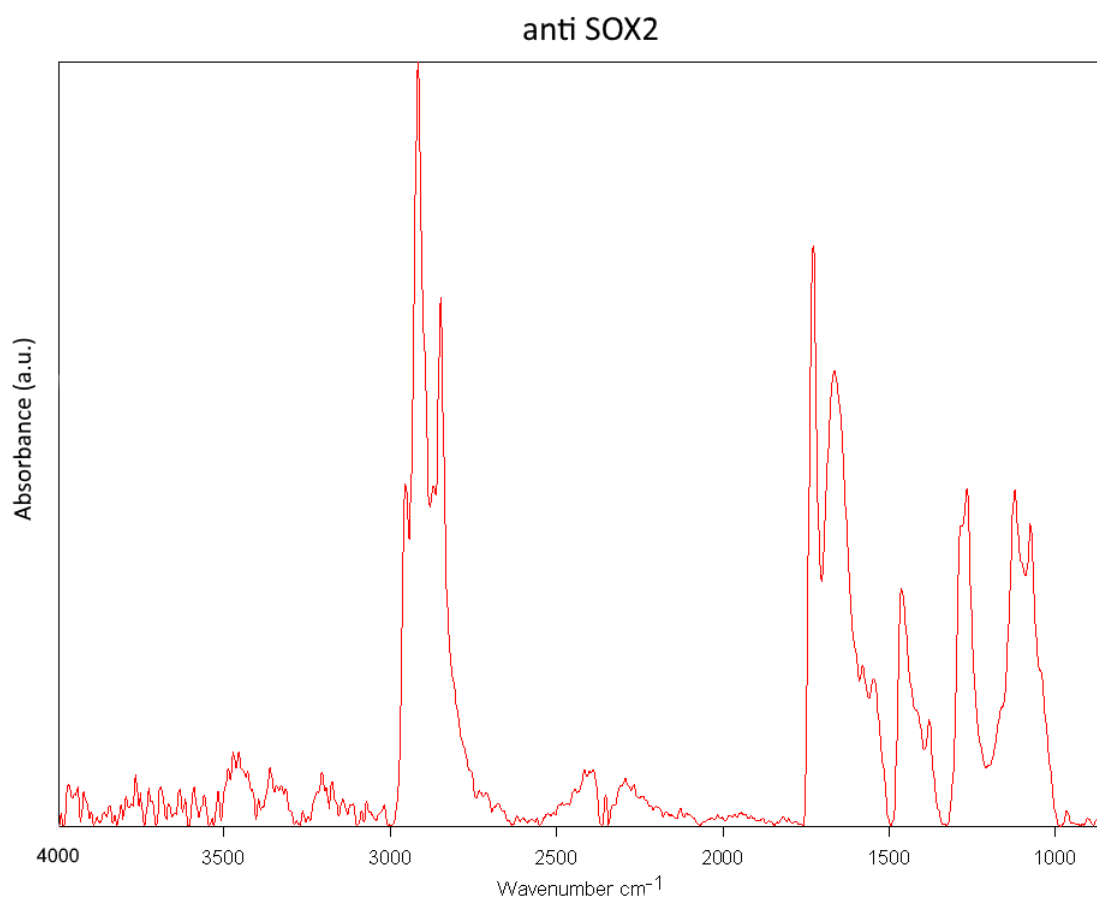


Figure 4.12: Absorbance spectrum of QCM chip surface after anti-SOX2 measurement, analyzed via ATR FT-IR. Observed peak(s) at 2956 indicate existence of CH₃, 2918 indicate CH₂, 1728 indicate COOH, 1700-1600 range indicate C=O and 1546 indicate NH groups.

4.2.5. QCM and ELISA Comparison

Calibration curve obtained via anti-SOX2 mAb binding (8 min at 27uL/min) to SOX2 protein immobilized Au surface in a concentration range 32 ng/mL and 512 ng/mL is shown together with ELISA measurements of anti-SOX2 mAb in SOX2 protein immobilized ELISA plates in a concentration range between 0.20 ng/mL and 200 ng/mL in Figure 4.13. Both methods had comparable detection ranges.

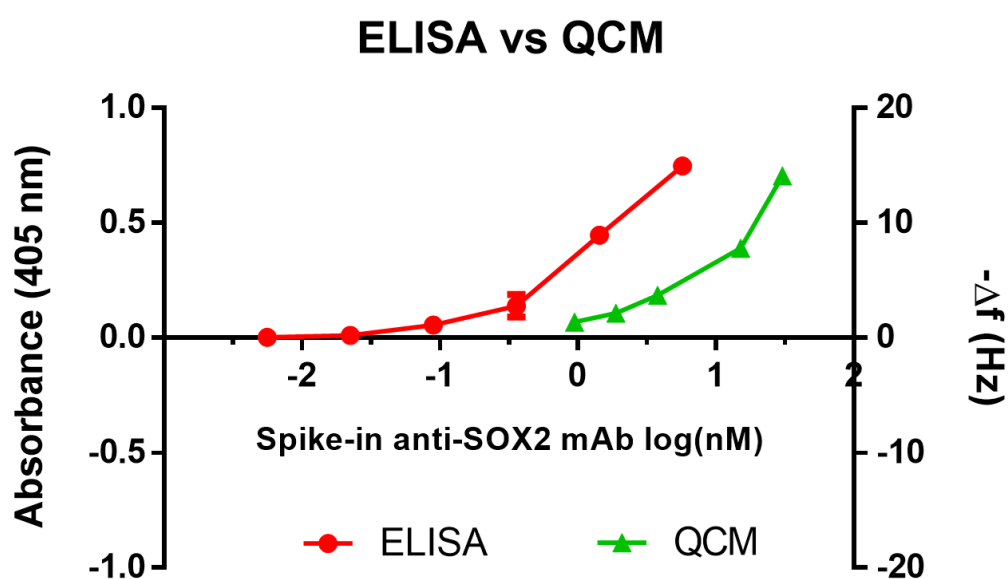


Figure 4.13: Comparison of ELISA and QCM standard curve measurements for anti-SOX2 mAb in PBS Buffer. On the left hand side with circular data points ELISA measurements between 0.2 ng/ml and 200 ng/mL are displayed. On the righthand side with triangular data points QCM measurements between 32 ng/ml and 512 ng/mL are displayed. Both assays had comparable detection ranges. QCM had linear range between 12.5 ng/mL and 200 ng/mL and QCM had linear range between 32 ng/mL and 256 ng/mL.

4.3.DISCOVER AND VALIDATION of miRNA BIOMARKERS in NON-SMALL CELL LUNG CANCER

4.3.1. Dataset Selection

NSCLC datasets of serum, plasma and whole blood were searched from GEO Datasets for the time period between 2008 and 2018. Among 6 datasets that were found, 4 of them had varying experiment platforms, which caused an issue in normalization. Also varying sample types made it difficult to make comparison between datasets. Therefore, only 2 of the datasets GSE24709 and GSE61741 with the same platform and sample type were utilized in further analysis. Normalized data is downloaded from GEO. Detailed information of the datasets are given in Table 4.3. Moreover, miRNA sequencing data from TCGA-LUAD and TCGA-LUSC of NIH were used to check miRNA profile in tumor tissue.

Table 4.3: List and details of the datasets selected for further analysis.

Dataset Number	Accession Number	Author(s)	Sample Type	Publication Year	Description of Comparison Groups	Platform
D1	GSE61741	Keller A., Leidinger P., Vogel B., Backes C. et al.[76]	Peripheral blood	2014	NSCLC patients vs Normal control	febit <i>Homo Sapiens</i> miRBase 13.0
D2	GSE24709	Keller A, Leidinger P.[77]	Peripheral Blood	2011	NSCLC patients vs Normal control	febit <i>Homo Sapiens</i> miRBase 13.0

GSE61741 have peripheral miRNA blood profile analysis of lung cancer patients. Expression profiling were done for 94 healthy controls, 15 long-lived individuals, and 940 patients affected from varying number of diseases.

GSE24709 have peripheral miRNA blood profile analysis of lung cancer patients. Expression profiling containing 863 miRNAs were done for lung cancer patients, normal controls and chronic obstructive pulmonary disease (COPD) patients.

4.3.2. miRNA Selection Incorporating Logistic Regression and LASSO Regularization

Initially, differential expression analysis were conducted in GSE61741 and GSE24709 whole blood datasets. Benjamini-Hochberg corrected p-value less than 0.05, and log2 fold change over 1.0 were used in selection of significantly upregulated miRNAs. 18 miRNAs were commonly upregulated in both datasets (Table 4.4). Moreover, miRNAs within the tumor tissue were checked using miRNA sequencing data from TCGA-LUSC and TCGA-LUAD. None of the 18 miRNAs which are significantly upregulated in both whole blood datasets (GSE61741 and GSE24709) were common with the miRNAs upregulated in tumor tissue. (Figure 4.14).

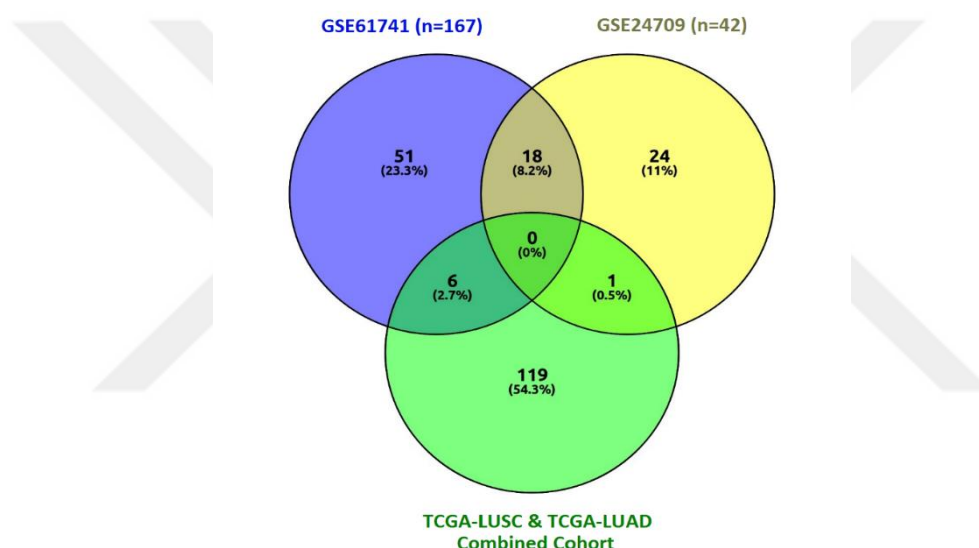


Figure 4.14: Common miRNAs between all datasets. Blood datasets GSE61741 and GSE24709 had 18 significantly upregulated miRNAs in common. However, none of the 18 miRNAs were common with unified TCGA-LUSC & TCGA-LUAD dataset. Both blood datasets, individually, had significantly upregulated common miRNAs with unified TCGA-LUSC & TCGA-LUAD dataset.

Following detection of significantly upregulated miRNAs, 18 selected miRNAs were subjected to logistic regression analysis via LASSO. Samples and their controls from GEO61741 dataset were assigned into two groups: Training and Validation. In the Training group logistic regression model was trained in R using “glmnet()” LASSO regularization package. 10 fold cross validation with various λ (lambda) values was conducted, and Area Under Curve (AUC) was used as the error metric. Model with the lowest λ value which lies with the 1 Standard Error of the model with highest mean AUC were chosen. The λ value with highest AUC, chosen λ value and their corresponding miRNA numbers can be seen in Figure 4.15.

Table 4.4: List of 18 significantly upregulated miRNAs common between GSE61741 and GSE24709

miRNA Annotation	GSE 61741			GSE 24709		
	p.value	BH corrected p.value	logFC	p.value	BH corrected p.value	logFC
hsa-miR-432-5p	0.000464	0.002772	1.330118	0.000003	0.000419	3.506204
hsa-miR-942-5p	0.017899	0.048493	1.001192	0.000024	0.001753	3.307177
hsa-miR-154-5p	0.000838	0.004557	1.295248	0.000227	0.006986	2.794426
hsa-miR-1261	0.000008	0.000126	1.720620	0.000085	0.003661	2.718142
hsa-miR-130b-5p	0.000000	0.000003	2.095646	0.000552	0.011929	2.623108
hsa-miR-769-3p	0.001161	0.005894	1.269865	0.000325	0.009056	2.613471
hsa-miR-132-5p	0.000001	0.000022	1.914010	0.000509	0.011929	2.576689
hsa-miR-373-3p	0.000167	0.001261	1.499702	0.000625	0.012126	2.491918
hsa-miR-1321	0.000334	0.002179	1.362636	0.000626	0.012126	2.467577
hsa-miR-1246	0.002245	0.009818	1.233802	0.001313	0.019565	2.416414
hsa-miR-328-3p	0.000077	0.000691	1.271431	0.001718	0.021136	2.268317
hsa-miR-206	0.000415	0.002595	1.386943	0.004173	0.036951	2.250639
hsa-let-7d-3p	0.000000	0.000000	1.679559	0.000114	0.004132	2.166025
hsa-miR-135a-5p	0.000373	0.002395	1.348858	0.000542	0.011929	2.110843
hsa-miR-140-5p	0.007905	0.025490	1.040529	0.004152	0.036951	1.975060
hsa-miR-624-5p	0.002391	0.010343	1.074925	0.004351	0.037545	1.524002
hsa-miR-1227-3p	0.000016	0.000214	1.024278	0.001400	0.019968	1.136471
hsa-miR-199a-5p	0.000000	0.000000	1.311275	0.000068	0.003661	1.120959

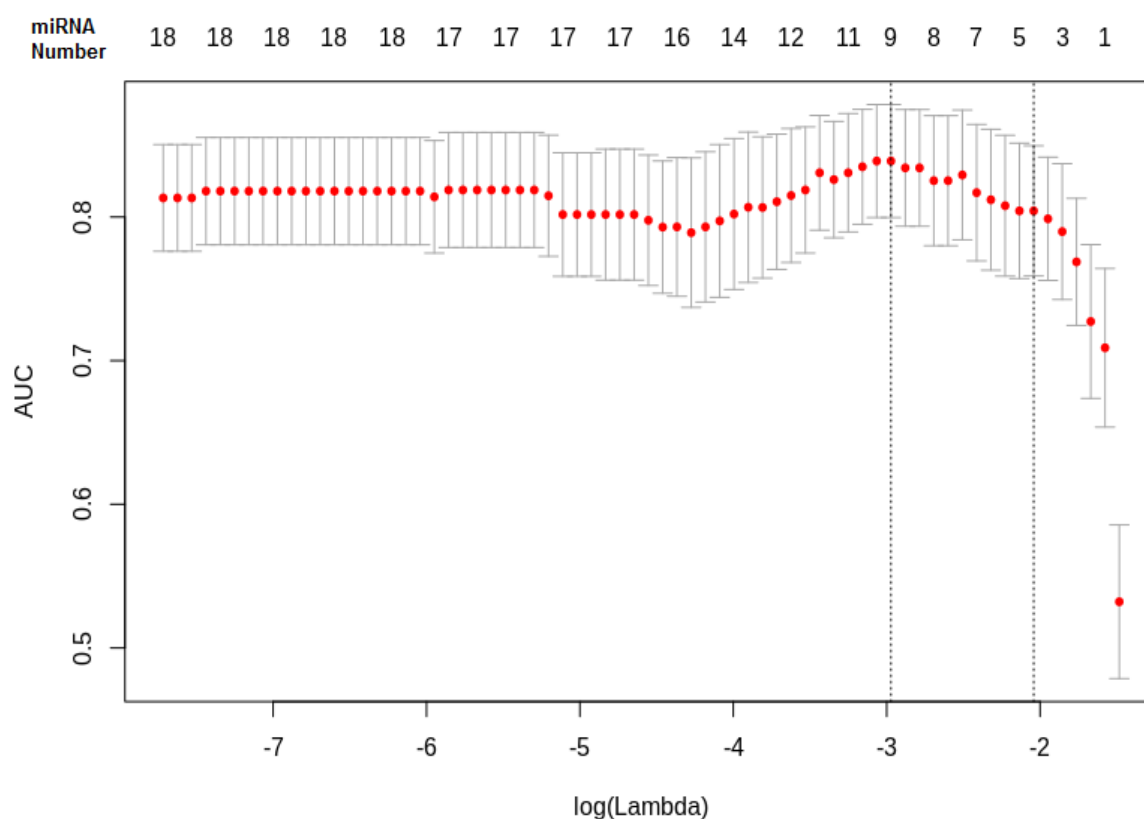


Figure 4.15: Error metric for Logistic Regression training in Training Group of GSE61741. Red dots are AUC values for corresponding $\log(\text{Lambda})$ values used for determination of Logistic Regression formula. Vertical dashed line at the left indicates the $\log(\text{Lambda})$ value with highest AUC. Vertical dashed line at the right indicates the $\log(\text{Lambda})$ value that lies within the 1 Standard Error of highest AUC model.

With the selected lambda values, Logistic Regression training generated a list of 5 miRNAs. Selected miRNAs had different weights which is incorporated in a logistic regression formula to differentiate NSCLC from Healthy controls. This miRNA list and their LASSO coefficients (weights) which incorporate the model can be seen in Table 4.5.

Table 4.5: List of 5 miRNAs selected by LASSO training in GSE61741 Training Group.

miRNA Number	miRNA Annotation	Lasso Coefficients
1	hsa-miR-130b-5p	0.10570584
2	hsa-miR-132-5p	0.08791144
3	hsa-let-7d-3p	0.03595553
4	hsa-miR-1227-3p	0.13960747
5	hsa-miR-199a-5p	0.02399983

Generated model with 5 miRNAs were used to plot ROC curves in Training and Validation groups of GSE61741. ROC curves plotted in Training group of GSE61741 had an AUC of 86.13% (Figure 4.16).

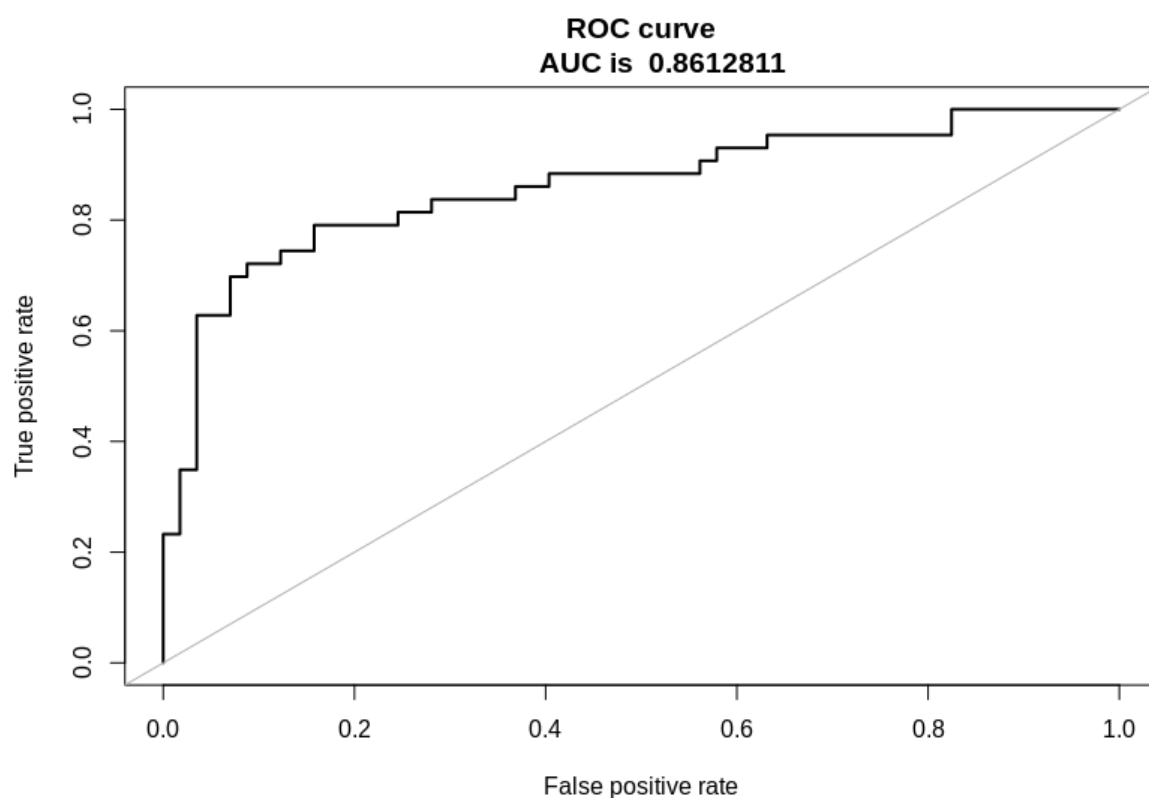


Figure 4.16: ROC curve plotted from Training group of GSE61741 using LASSO generated model. ROC curve plotted with 5 miRNAs which are selected with LASSO from Training group of GSE61741 had an AUC of 0.86 in the same group.

Generated LASSO model with 5 miRNAs were validated in Validation group of GSE61741. ROC curve (Figure 4.17) plotted for Validation group of GSE61741 with 5 miRNA LASSO model had an AUC value of 0.81.

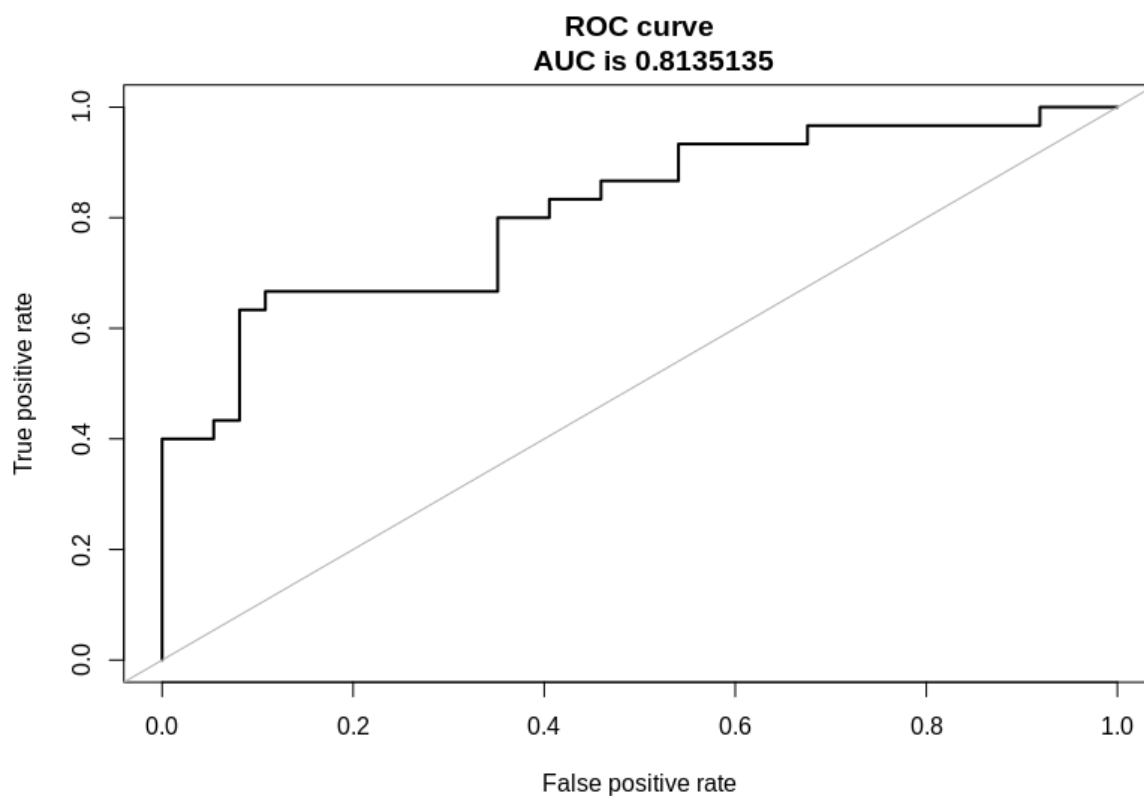


Figure 4.17: ROC curve plotted from Validation group of GSE61741 using LASSO generated model with 5miRNAs. ROC curve plotted with 5 miRNAs which are selected with LASSO from Training Group of GSE61741 had an AUC of 0.81 in Validation group of GSE61741 .

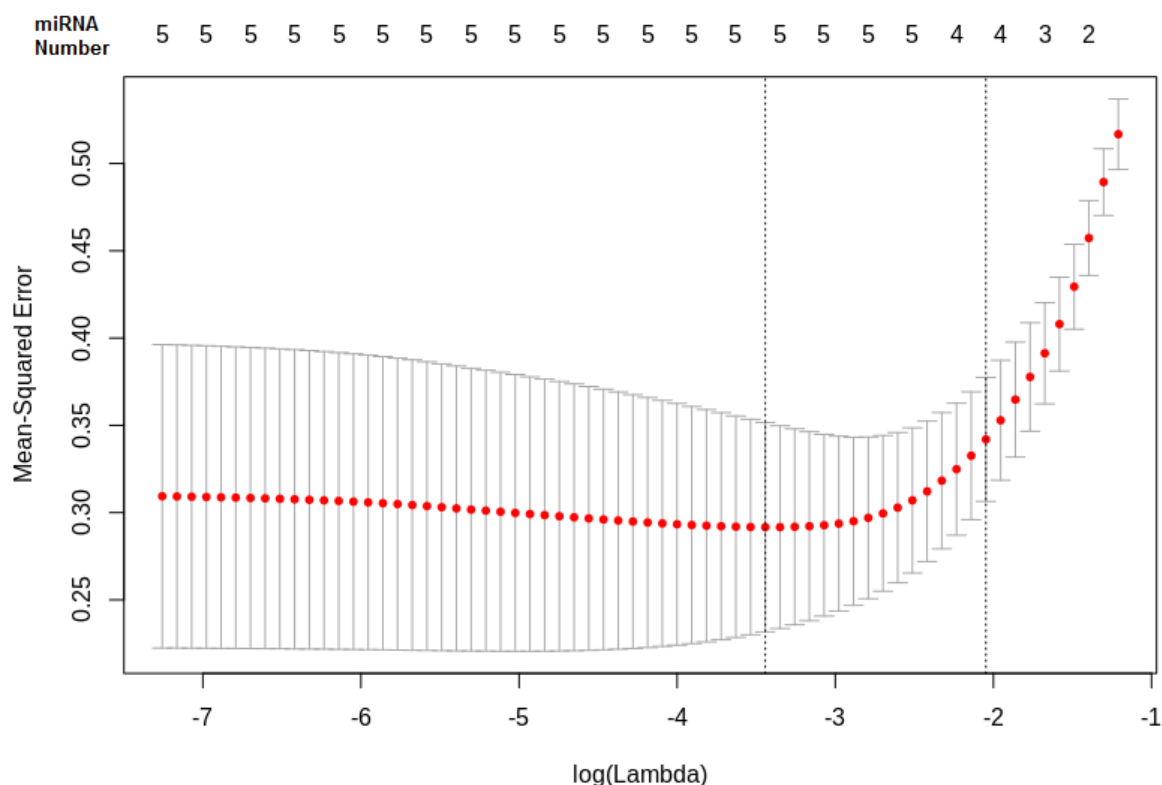


Figure 4.18: Error metric for Logistic Regression training in Training Group of GSE24709. Red dots are MSE values for corresponding $\log(\text{Lambda})$ values used for determination of Logistic Regression formula. Vertical dashed line at the left indicates the $\log(\text{Lambda})$ value with lowest MSE value. Vertical dashed line at the right indicates the $\log(\text{Lambda})$ value that lies within the 1 Standard Error of the model with lowest MSE and the corresponding miRNA number.

5 miRNAs selected in GSE61741 using LASSO were used to train another model to differentiate NSCLC from Healthy controls in GSE24709. Sample number of the dataset were not enough to generate Training and Validation groups. Therefore, training and validation of logistic regression model were done in the whole dataset using LASSO regularization package. Mean Squared Error (MSE) was used as the error metric. Model with the lowest λ value which lies within the 1 Standard Error of the model with the lowest MSE was chosen (Figure 4.18). Selected model had 4 of the miRNAs from previous panel (Table 4.6). These 4 miRNAs were then used to draw an ROC curve by using GSE24709 expression data, with an AUC of 93% (Figure 4.19).

Table 4.6: List of 4 miRNAs selected by LASSO training in GSE24709 dataset.

miRNA Number	miRNA Annotation	Lasso Coefficients
1	hsa-miR-130b-5p	0.02807713
2	hsa-miR-132-5p	0.06595173
3	hsa-let-7d-3p	0.28314593
5	hsa-miR-199a-5p	0.44272157

5 miRNAs selected from a subset of 18 miRNAs (significantly upregulated and common between GSE61741 and GSE24709) via logistic regression were subjected to a validation experiment using qRT-PCR. Serum of NSCLC patients and High Risk Group (HRG) controls with matched age, sex, smoking status were used. Prior to that, serum expression levels were checked in GSE61741 and GSE24709 (Figure 4.20), pointing out to high mean expressions in 3 of 5 miRNAs when compared to other 2.

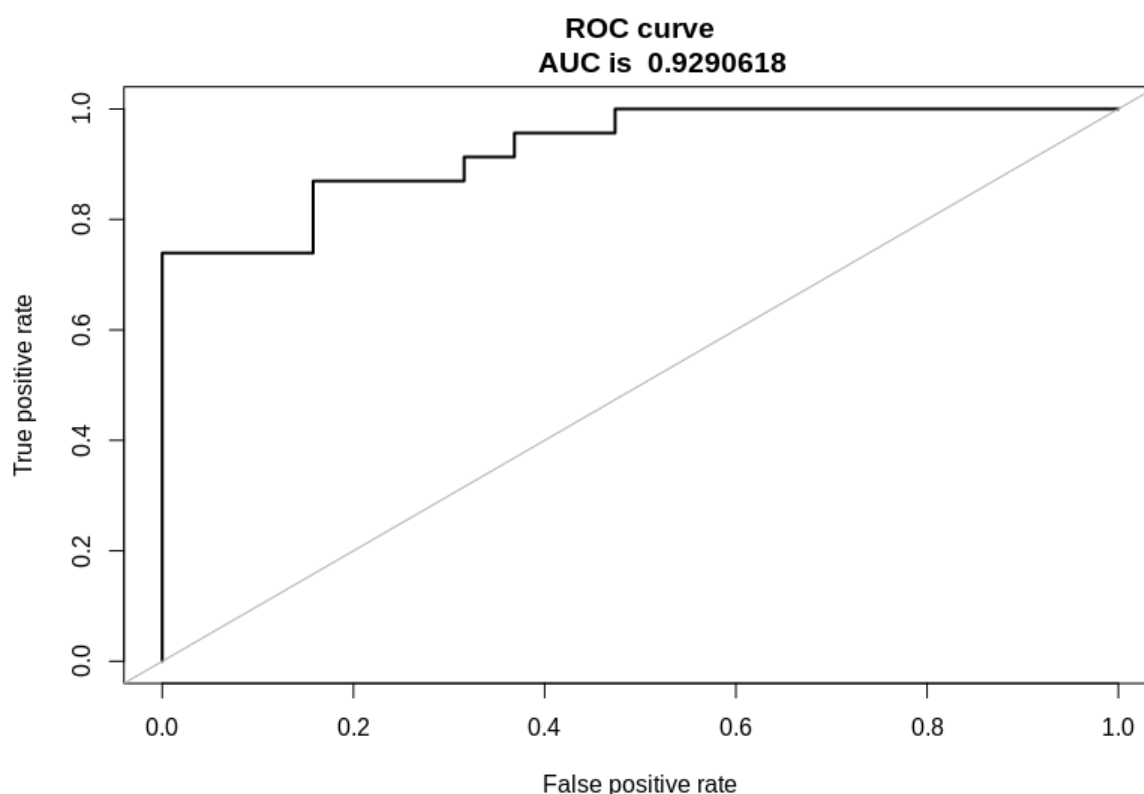
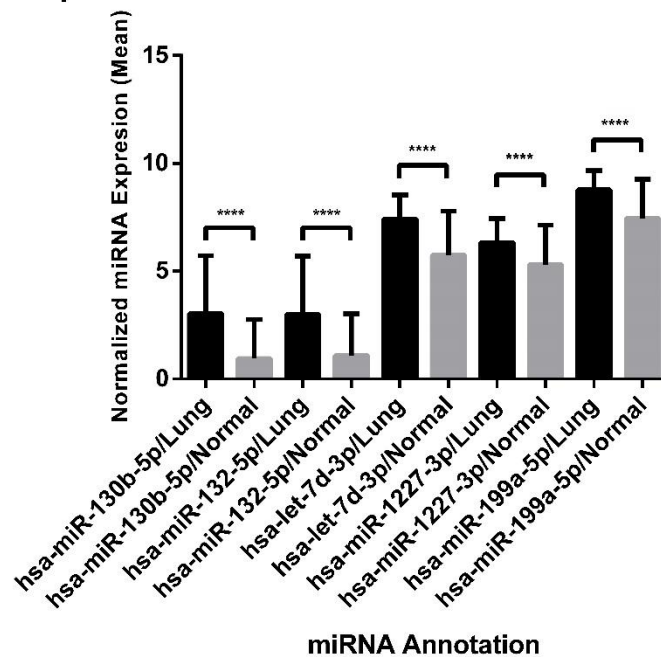


Figure 4.19: ROC curve plotted from GSE24709 using LASSO generated model with 4 miRNAs. ROC curve plotted with 4 miRNAs which are selected with LASSO from GSE24709 had an AUC of 0.929 in the same dataset.

A)

Mean Expression of LASSO Selected miRNAs in GSE61741 dataset



B)

Mean Expression of LASSO Selected miRNAs in GSE24709 dataset

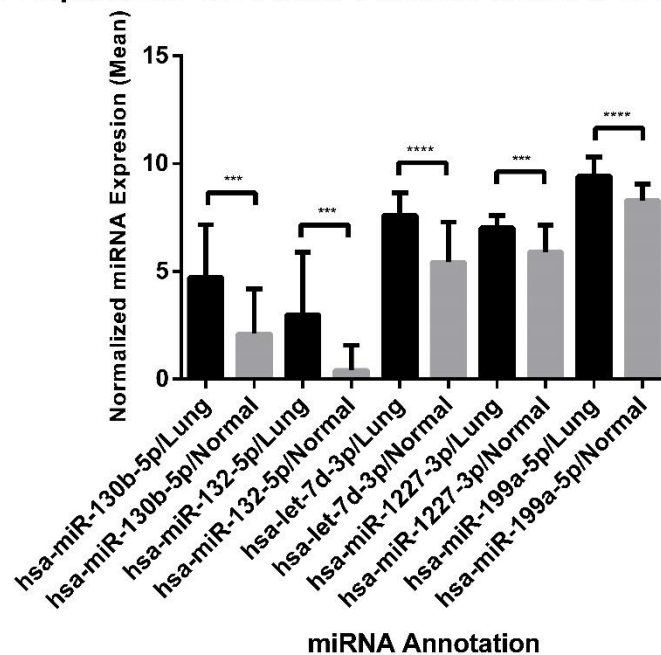


Figure 4.20: Mean Expression of LASSO Selected miRNAs in GSE61741 and GSE24709. In both datasets selected 5 miRNAs had significant upregulation in NSCLC group compared to Normal Controls. **A)** Expression levels of 5 miRNAs in GSE61741 and **B)** Expression levels of 5 miRNAs in GSE24709.

4.3.3. Hemolysis Assessment

Hemolysis status of sera were assessed via oxy-hemoglobin peaks at 414 nm. Absorbance<0.072 filter at 414 nm were used for low hemolysis risk, $0.072 \leq \text{Absorbance} \leq 0.3$ filter at 414 nm were used for moderate hemolysis risk and Absorbance>0.3 at 414 nm were used for high hemolysis risk [68]. Based on given cut-offs all of the serum measurements indicate medium to high hemolysis risk (Table 4.7).

Table 4.7: OD 414 measurements of samples for hemolysis assessment.

Sample Code (High Risk Group)	Absorption (414nm) Nanodrop
H1	2.68
H2	0.77
H3	2.72
H4	1.28
H5	2.53
H6	0.66
H7	3.18
Sample Code (NSCLC)	Absorption (414nm) Nanodrop
N1	1.97
N2	3.77
N3	0.64
N4	15.00
N5	4.53
N6	1.30
N7	2.01

4.3.4. RNA Extraction and Quantification

To determine quantity and quality of the RNA isolates, measurements were taken with NanoDrop One/One C Microvolume UV–Vis Spectrophotometer (Thermo Fisher Scientific, Fitchburg, WI) using RNA measurement protocol. RNA isolates calculated by nanodrop were in a quite big range from 1.5 ng/uL to 33.5 ng/uL.

4.3.5. Quantitative Real Time Polymerase Chain Reaction (qRT-PCR)

To validate our results we employed qRT-PCR. However, we were unable to successfully generate Ct values for all samples and all miRNA cases (Figure 4.21), resulting in a low quality qPCR data. This issue was probably due to low cDNA template existence in qPCR reactions. Because, miRNAs for which the most sample Ct values are calculated in our qPCR experiment are also miRNAs

with highest mean expressions in GSE61741 and GSE24709. Experiments needs to be repeated with patient serum that has low hemolysis before a significant analysis on the qPCR data is done.

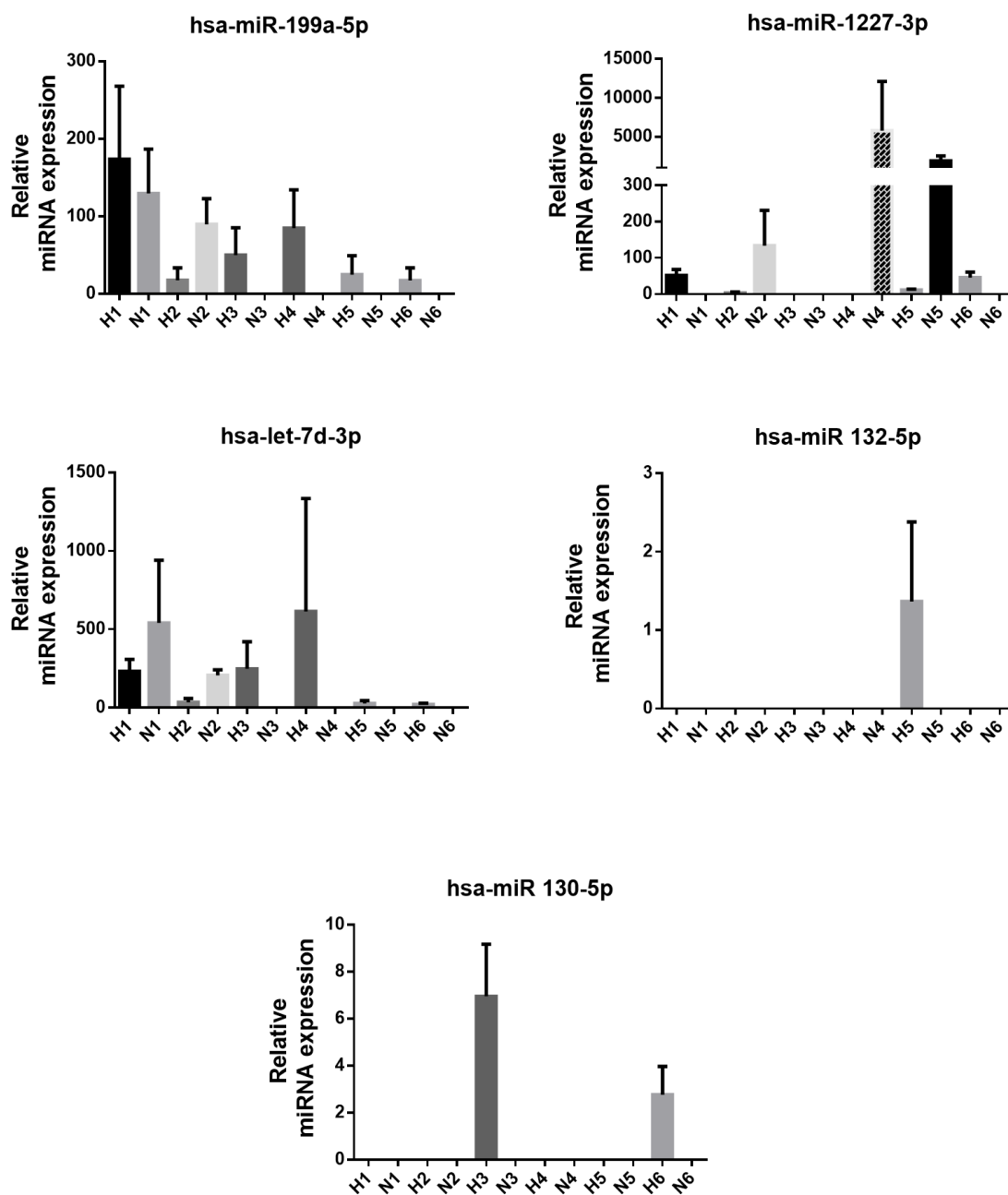


Figure 4.21: Relative miRNA expressions prepared for 5 miRNAs in 7 NSCLC patients and their matched high risk controls.

5. DISCUSSION & FUTURE PERSPECTIVES

5.1. PA Screening Evaluations

In order to improve on previous evaluations of custom Protein Array results generated in our lab, we have applied a new approach to filter out low quality signal exist throughout the screening results, which might be due to inadequate serum coverage and wrinkles on custom Protein Array strips and image recording. We used AIDA software to generate numerical values out of cPA screening images and applied a cut-off based filtration on the signal intensity results. Two filters were tried. (Mean)-(2.5xStDev) and (Mean)-(1.5xStDev). If a signal harboring unit had spots that fail to pass the calculated cut-off, all signals in that signal harboring unit were discarded. Efficiency of set cut-off was verified by checking cPAs with eye. We were able to successfully eliminate low intensity signals via (Mean)-(1.5xStDev) cut-off supported by improvements in AUC values in individual and combination ROCs compared to ROCs from previous cPA evaluations in our lab.

We were able to obtain several antibody panels with $AUC \geq 95\%$. We selected a panel of autologous antibodies against FKBP8 - P53 - SOX2 – POLB which had an AUC 95.04%, with 60% sensitivity at 100% specificity. This is significant, as the only commercial kit based on autologous anti-tumor antibodies (EarlyCDT-Lung Kit) employs 7 antibodies (p53, NY-ESO-1, CAGE, GBU4-5, SOX2, MAGE A4, HuD) and reaches 28% sensitivity at 98% specificity or 41% sensitivity at 93% specificity [78; 79]. Moreover, EarlyCDT kit's improved sensitivity and specificity values throughout the years with incorporation of new auto-antibodies [78-80] is a major support on the widely expected potential performance of auto-antibodies when they are used in panels [42-53]. However, auto-antibodies are not without limitations as diagnostic biomarkers. Some patients show little to no serum reactivity against TAAs, indicating an immune-suppressed phenotype which is in concordance with the current understanding of immunosurveillance and immunoediting theories [35-39]. But these problems can be overcome by incorporating different types of biomarkers such as aberrancies in proteins (tumor associated antigens, post translational modifications, mutations) and miRNAs [38-40].

Although similar results were obtained in our lab with previous evaluations and validated with ELISA and Western Blot. Discrepancies between cPA and ELISA results were existent. However, superior reported sensitivities of protein array based systems against ELISA with up to 3 orders of magnitude sensitivity [81] suggest the issue to be the cause of insufficient limit of detection in ELISA. Therefore, in need of a new system to validate our results we expect to develop a QCM based immune-assay for the measurement of these 4 auto-antibodies (FKBP8 – P53 –SOX2 – POLB). This system, if proven its value in high-throughput, can be used to develop a diagnostic method for cheap, quick and easy detection of serum autologous antibodies.

5.2. Literature Search for the Autoantibodies in the Selected Panel

A detailed literature search is done individually for the autoantibodies in the selected autologous antibody panel (FKBP8 - P53 - SOX2 – POLB).

5.2.1. FKBP8

FKBP8, aka FKBP38, is shown to prevent apoptosis by interacting and anchoring Bcl-2 to mitochondrial membrane [82].

5.2.2. P53

With high immunogenicity of p53 protein, p53 Antibodies are studied in the sera of more than 900 patients with a myriad of cancer types [83]. In breast, colon, oral, and gastric cancers this autoantibody is consistently associated with high-grade tumors and increased mortality [83]. A meta-analysis on esophageal cancer has shown that p53 auto-antibodies are the most assessed between 35 different others [84]. Also, more and more evidence indicate inhibitory effects of p53 on autoimmunity through induction of reg-T cells [85], as well as inflammation caused by p53 dysregulation [86].

5.2.3. SOX2

Auto antibodies against this protein first detected in our group, showing serological reactivity against SOX B group proteins in up to 40% of the SCLC patients [87]. Later, shown that epitopes of SOX2 protein are sera reactive in other cancers such as prostate, breast, glioblastoma and myeloma [88-90].

5.2.4. POLB

POLB is one of the widely studied proteins in conjunction with cancer due to its role as DNA polymerase with the enzymatic activity in excision repair. Almost one third of all tumor sequencing info indicate mutations in the gene compared to matched healthy tissues [91]. Aberrancies such as SNPs and overexpressions in POLB gene are widely reported for gastric, bladder, esophageal cancers [91-93].

5.3. QCM Development for Auto-Antibody Measurements

In order to develop a QCM based method for measurement of serum autoantibodies, setting it as an example we tried to develop a method to measure anti-SOX2 antibodies. Initially, we tried to modify AT cut 5MHz quartz crystal chips to immobilize. We were able to successfully generate a self-assembled monolayer of 11-MUA on the gold surface of the QCM chip. SOX2 (pH: 5.5) protein was successfully immobilized on the surface using EDC/NHS amine coupling chemistry. Modified chip was able to measure anti-SOX2 antibodies in PBS (pH 7.4) in concentrations as low as 4ng/mL and as high as 4096 ng/mL (not shown), here we reported a range of 32 ng/mL to 1024 ng/mL.

To our knowledge this method is first of its kind to try to measure anti-tumor antibodies in a QCM immunoassay format. Also, this study is the first to show the interaction of the anti-SOX2 antibody on SOX2 that is immobilized on surface in PBS (pH 7.4) buffer. Interaction between anti-SOX2 antibody and protein is modelled using 1:1 Langmuir Isothermal Binding model, revealing a K_D value of 11.2 nM between the antigen and antibody.

Surface characterization of QCM chip throughout the experiments with ATR FT-IR indicated organic bonds such as CH₃, CH₂, C=O and COOH throughout the experiments, as well as N-H bond in SOX2 conjugation and anti-SOX2 binding events. These results support the success of the surface modification and antibody binding events.

Also comparisons made between measurements of anti-SOX2 antibody in ELISA and QCM show comparable detection ranges in both assays. However, with further development and optimization of the QCM immunoassay, it has the potential to surpass the ELISA. Also, one of the advantages of QCM against ELISA in terms of detection limit is the serum dilution, 1:400 or 1:1600 serum dilutions are needed for ELISA measurement. However, QCM doesn't require such dilutions and has been reported to work in high serum concentrations [94]. With these at hand, our QCM is already superior to ELISA in terms of Limit of Detection.

It is important to design studies in a way that allow monitoring of biomarkers from samples of the same patient in following years. There are many studies showing importance of measuring biomarkers in following years for early diagnosis. In 2013 it has been shown that incorporating change over time by carrying longitudinal screening of Cancer Antigen 125 (CA125) biomarker in serum for ovarian cancer is able to detect 20% of cases earlier than non-longitudinal methods at 99% specificity. In 6 to 12 month period before the symptomatic clinical diagnosis, it is possible to diagnose over cancer with 17% sensitivity at 99% specificity using single-threshold methods, while employing longitudinal methods increase the sensitivity to 50% at the same specificity level [54]. Moreover, measurements of serum PSA levels is one of the most widely used biomarker screenings which initially incorporated in clinical decision making of prostate cancer by single cut point measurements [95]. Early use of the PSA

measurements indicated insufficient sensitivity and specificity of the assay [96]. This was due to the fact that elevated PSA at a single time point might be due to stable PSA levels or slowly increasing levels and more likely to reflect a benign prostate disease than an aggressive cancer [97]. However, a study conducted on serum samples collected from patients, 7 to 25 years before clinical diagnosis indicated that measurements of PSA can be a sensitive and specific early diagnostic marker [98]. There are studies that show longitudinal biomarker studies will also be important for diagnosis in other cancers such as mesothelioma [55-57]. Moreover, with longitudinal measurements a concept known as “delta check”, which is a quality control method that compares present and past test results of a patient, compares the difference to a pre-set criteria to eliminate faulty measurements, can be incorporated [99]. In our study autologous antibodies seem to be very effective in small cell lung cancer diagnosis, with their promising 60% sensitivity at 100% specificity. Even though there are not many longitudinal studies carried with autologous antibodies, in one of the most important studies in that subject it has been shown that longitudinal serum positivity is a clear biomarker for cancer development in patient while sporadic serum positivity is an evidence for cancer to be not progressing [2]. If we can fully develop our QCM based assay, we can later design large longitudinal cohort studies to detect the true power of those biomarkers in early diagnosis of cancer and detection of recurrence.

5.4. miRNA Biomarkers in NSCLC

We have identified 5 miRNAs (hsa-mir-130b-5p, hsa-miR-132-5p, hsa-let-7d-3p, hsa-miR-1227-3p, hsa-miR-199a-5p) *in silico* which are able to differentiate between NSCLC patients and healthy controls in given datasets. Datasets were trained via logistic training algorithm with LASSO regularization to decide on best possible cut-offs. To validate our results we employed qRT-PCR. However, we were unable to successfully generate Ct values for all samples and all miRNA cases. Resulting in a low quality qPCR data. This issue was probably due to low cDNA template existence in qPCR reactions. Because, miRNAs for which the most sample Ct values are calculated in our qPCR experiment are also miRNAs with highest mean expressions in GSE61741 and GSE24709. Experiments needs to be repeated with patient serum that has low hemolysis before a significant analysis on the qPCR data is done.

5.5. Literature Search for the miRNAs in Selected Panel

5.5.1. hsa-miR-130b-5p

The mature hsa-miR-130b-5p is the less stable and less investigated variant of miR-130b first identified in mice [100]. It was shown that miR-130b is upregulated in tumor initiating cells [101]. In NSCLC, a p53-initiating pathway mediating cell cycle arrest in G2/M phase by decreasing the stability of miR-130b was suggested as a tumor suppressing mechanism [102]. Therefore upregulation of it might be related to bad prognosis.

5.5.2. hsa-miR-132-5p

miR-132 has been suggested as a tumor suppressor in multiple studies concerning NSCLC, as it is found to be downregulated both in patient and tissue culture samples [103-105]. It was shown that miR-132 inhibits ZEB2 gene, which belongs to a family of zinc-finger protein coding genes and facilitates metastasis in many cancer types [105].

5.5.3. hsa-let-7d-3p

let-7d is another miRNA which functions as a tumor suppressor in healthy lung cells [106]. It was reported that the low expression level of let-7 in NSCLC patients is also associated with bad prognosis of the disease [107]. Let-7 is suspected to regulate proto-oncogenes such as c-MYC and RAS and therefore act as a cell cycle checkpoint control at G1/S phase transition [108]. LIN28B has been identified as a negative regulator of let-7, which can be inhibited post-transcriptionally to increase apoptosis in lung cancer cells [104].

5.5.4. hsa-miR-1227-3p

miR-1227 is a miRNA processed from an alternative precursor called mirtron [109] and has been studied less than the other miRNAs in our study. In digestive and excretory system cancers, an overall decline in expression of miR-1227 was observed [110]. There are no systematic studies linking miR-1227 with specific tumorigenic pathways yet, but it has been shown that exposure to the environmental pollutant particulate matter, which is linked to lung cancer, downregulates expression of miR-1227 [111].

5.5.5. hsa-miR-199a-5p

miR-199a is coded by two identical sequences in the genome, one present in chromosome 1 and other in chromosome 19. In colorectal cancer, NSCLC and arsenic-induced cancer, miR-199a-5p downregulation has been associated with tumorigenesis [112-114]. Moreover, it has been shown that miR-199a-5p levels can be used as a predictor for cancer relapse in squamous cell carcinoma patients, making it a promising prognostic target [115]. Also, as miR-199a-5p is found upregulated in COPD patients, it can also be utilized to distinguish between the two types of disease [116].



6. BIBLIOGRAPHY

1. Finn, O. J., *A Believer's Overview of Cancer Immunosurveillance and Immunotherapy*. J Immunol, 2018. **200**(2), 385-391.
2. Li, Y., Karjalainen A., Koskinen H., et al., *p53 autoantibodies predict subsequent development of cancer*. International Journal of Cancer, 2005. **114**(1), 157-160.
3. Bray, F., Ferlay J., Soerjomataram I., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA: A Cancer Journal for Clinicians, 2018. **68**(6), 394-424.
4. Rahib, L., Smith B. D., Aizenberg R., et al., *Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States*. Cancer Res, 2014. **74**(11), 2913-2921.
5. Thun, M. J., Delancey J. O., Center M. M., et al., *The global burden of cancer: priorities for prevention*. Carcinogenesis, 2010. **31**(1), 100-110.
6. Semenova, E. A., Nagel R., Berns A., *Origins, genetic landscape, and emerging therapies of small cell lung cancer*. Genes & Development, 2015. **29**(14), 1447-1462.
7. Van Meerbeeck, J. P., Fennell D. A., De Ruyscher D. K., *Small-cell lung cancer*. Lancet, 2011. **378**(9804), 1741-1755.
8. Califano, R., Kerr K., Morgan R. D., et al., *Immune Checkpoint Blockade: A New Era for Non-Small Cell Lung Cancer*. Curr Oncol Rep, 2016. **18**(9), 59.
9. Ganti, A. K., Zhen W. K., A., *Limited-Stage Small-Cell Lung Cancer: Therapeutic Options*. Oncology, 2007. **21**(3).
10. Lim, E., Belcher E., Yap Y. K., et al., *The role of surgery in the treatment of limited disease small cell lung cancer: time to reevaluate*. J Thorac Oncol, 2008. **3**(11), 1267-1271.
11. Schreiber, D., Rineer J., Weedon J., et al., *Survival outcomes with the use of surgery in limited-stage small cell lung cancer: should its role be re-evaluated?* Cancer, 2010. **116**(5), 1350-1357.
12. Koletsis, E. N., Prokakis C., Karanikolas M., et al., *Current role of surgery in small cell lung carcinoma*. J Cardiothorac Surg, 2009. **4**, 30.
13. Brock, M. V., Hooker C. M., Syphard J. E., et al., *Surgical resection of limited disease small cell lung cancer in the new era of platinum chemotherapy: Its time has come*. J Thorac Cardiovasc Surg, 2005. **129**(1), 64-72.
14. Simon, G. R., Wagner H., American College of Chest P., *Small cell lung cancer*. Chest, 2003. **123**(1 Suppl), 259S-271S.
15. Jackman, D. M., Johnson B. E., *Small-cell lung cancer*. Lancet, 2005. **366**(9494), 1385-1396.
16. Kato, Y., Ferguson T. B., Bennett D. E., Burford T. H., *Oat cell carcinoma of the lung. A review of 138 cases*. Cancer, 1969. **23**(3), 517-524.
17. Zarogoulidis, K., Zarogoulidis P., Darwiche K., et al., *Treatment of non-small cell lung cancer (NSCLC)*. J Thorac Dis, 2013. **5 Suppl 4**, S389-396.
18. Byers, L. a. R., C. M., *Small cell lung cancer: Where do we go from here?* Cancer, 2015. **121**(5), 664-672.
19. Biederer, J., Beer M., Hirsch W., et al., *MRI of the lung (2/3). Why ... when ... how?* Insights Imaging, 2012. **3**(4), 355-371.
20. Lowe, V. J., Fletcher J. W., Gobar L., et al., *Prospective investigation of positron emission tomography in lung nodules*. J Clin Oncol, 1998. **16**(3), 1075-1084.
21. National Lung Screening Trial Research, T., Aberle D. R., Adams A. M., et al., *Reduced lung-cancer mortality with low-dose computed tomographic screening*. N Engl J Med, 2011. **365**(5), 395-409.
22. Chapman, C. J., Thorpe A. J., Murray A., et al., *Immunobiomarkers in small cell lung cancer: potential early cancer signals*. Clin Cancer Res, 2011. **17**(6), 1474-1480.
23. Kalemkerian, G. P., Akerley W., Bogner P., et al., *Small cell lung cancer*. J Natl Compr Canc Netw, 2013. **11**(1), 78-98.

24. Annema, J. T., Versteegh M. I., Veselic M., et al., *Endoscopic ultrasound-guided fine-needle aspiration in the diagnosis and staging of lung cancer and its impact on surgical staging*. J Clin Oncol, 2005. **23**(33), 8357-8361.
25. Chin, R., Jr., Cappellari J. O., McCain T. W., et al., *Increasing use of bronchoscopic needle aspiration to diagnose small cell lung cancer*. Mayo Clin Proc, 2000. **75**(8), 796-801.
26. Delgado, P. I., Jorda M., Ganjei-Azar P., *Small cell carcinoma versus other lung malignancies: diagnosis by fine-needle aspiration cytology*. Cancer, 2000. **90**(5), 279-285.
27. Tondini, M., Rizzi A., *Small-cell lung cancer: importance of fiberoptic bronchoscopy in the evaluation of complete remission*. Tumori, 1989. **75**(3), 266-268.
28. Cristallini, E. G., Ascani S., Farabi R., et al., *Fine needle aspiration biopsy in the diagnosis of intrathoracic masses*. Acta Cytol, 1992. **36**(3), 416-422.
29. International Early Lung Cancer Action Program, I., Henschke C. I., Yankelevitz D. F., et al., *Survival of patients with stage I lung cancer detected on CT screening*. N Engl J Med, 2006. **355**(17), 1763-1771.
30. Bach, P. B., Jett J. R., Pastorino U., et al., *Computed tomography screening and lung cancer outcomes*. JAMA, 2007. **297**(9), 953-961.
31. Zulueta, J. J., De-Torres J. P., *Lung Cancer Screening: The Balance between Harm and Benefit*. Am J Respir Crit Care Med, 2015. **191**(10), 1209.
32. Tanoue, L. T., Tanner N. T., Gould M. K., Silvestri G. A., *Reply: Lung Cancer Screening: The Balance between Harm and Benefit*. Am J Respir Crit Care Med, 2015. **191**(10), 1209-1210.
33. Tanoue, L. T., Tanner N. T., Gould M. K., Silvestri G. A., *Lung cancer screening*. Am J Respir Crit Care Med, 2015. **191**(1), 19-33.
34. Xu, D. M., Lee I. J., Zhao S., et al., *CT screening for lung cancer: value of expert review of initial baseline screenings*. AJR Am J Roentgenol, 2015. **204**(2), 281-286.
35. Mittal, D., Gubin M. M., Schreiber R. D., Smyth M. J., *New insights into cancer immunoediting and its three component phases--elimination, equilibrium and escape*. Curr Opin Immunol, 2014. **27**, 16-25.
36. Ebogo-Belobo, J. T., Kagoue Simeni L. A., Mbassa Nnouma G., et al., *Incidence of cancer in people living with HIV and prognostic value of current CD4()*. Bull Cancer, 2018.
37. Heo, C. K., Bahk Y. Y., Cho E. W., *Tumor-associated autoantibodies as diagnostic and prognostic biomarkers*. BMB Rep, 2012. **45**(12), 677-685.
38. Anderson, K. S., Labaer J., *The sentinel within: exploiting the immune system for cancer biomarkers*. J Proteome Res, 2005. **4**(4), 1123-1133.
39. Dunn, G. P., Old L. J., Schreiber R. D., *The immunobiology of cancer immunosurveillance and immunoediting*. Immunity, 2004. **21**(2), 137-148.
40. Trudgen, K., Khattar N. H., Bensadoun E., et al., *Autoantibody profiling for lung cancer screening longitudinal retrospective analysis of CT screening cohorts*. PLOS ONE, 2014. **9**(2), e87947.
41. Hammerich, L., Binder A., Brody J. D., *In situ vaccination: Cancer immunotherapy both personalized and off-the-shelf*. Mol Oncol, 2015. **9**(10), 1966-1981.
42. Van't Westeinde, S. C., Van Klaveren R. J., *Screening and early detection of lung cancer*. Cancer J, 2011. **17**(1), 3-10.
43. Trivers, G. E., De Benedetti V. M., Cawley H. L., et al., *Anti-p53 antibodies in sera from patients with chronic obstructive pulmonary disease can predate a diagnosis of cancer*. Clin Cancer Res, 1996. **2**(10), 1767-1775.
44. Doyle, H. A., Mamula M. J., *Post-translational protein modifications in antigen recognition and autoimmunity*. Trends Immunol, 2001. **22**(8), 443-449.
45. Lam, S., Boyle P., Healey G. F., et al., *EarlyCDT-Lung: an immunobiomarker test as an aid to early detection of lung cancer*. Cancer Prev Res (Phila), 2011. **4**(7), 1126-1134.
46. Qiu, J., Choi G., Li L., et al., *Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in prediagnostic lung cancer sera*. J Clin Oncol, 2008. **26**(31), 5060-5066.
47. Backes, C., Ludwig N., Leidinger P., et al., *Immunogenicity of autoantigens*. BMC Genomics, 2011. **12**, 340.
48. Khattar, N. H., Coe-Atkinson S. P., Stromberg A. J., et al., *Lung cancer-associated auto-antibodies measured using seven amino acid peptides in a diagnostic blood test for lung cancer*. Cancer Biol Ther, 2010. **10**(3), 267-272.

49. Zhong, L., Coe S. P., Stromberg A. J., et al., *Profiling tumor-associated antibodies for early detection of non-small cell lung cancer*. J Thorac Oncol, 2006. **1**(6), 513-519.
50. Zhong, L., Hidalgo G. E., Stromberg A. J., et al., *Using protein microarray as a diagnostic assay for non-small cell lung cancer*. Am J Respir Crit Care Med, 2005. **172**(10), 1308-1314.
51. Chapman, C. J., Murray A., Mcelveen J. E., et al., *Autoantibodies in lung cancer: possibilities for early detection and subsequent cure*. Thorax, 2008. **63**(3), 228-233.
52. Desmetz, C., Mange A., Maudelonde T., Solassol J., *Autoantibody signatures: progress and perspectives for early cancer detection*. J Cell Mol Med, 2011. **15**(10), 2013-2024.
53. Marshall, E., *Cancer screening. The promise and pitfalls of a cancer breakthrough*. Science, 2010. **330**(6006), 900-901.
54. Drescher, C. W., Shah C., Thorpe J., et al., *Longitudinal screening algorithm that incorporates change over time in CA125 levels identifies ovarian cancer earlier than a single-threshold rule*. J Clin Oncol, 2013. **31**(3), 387-392.
55. Roe, O. D., Creaney J., Lundgren S., et al., *Mesothelin-related predictive and prognostic factors in malignant mesothelioma: a nested case-control study*. Lung Cancer, 2008. **61**(2), 235-243.
56. Creaney, J., Olsen N. J., Brims F., et al., *Serum mesothelin for early detection of asbestos-induced cancer malignant mesothelioma*. Cancer Epidemiol Biomarkers Prev, 2010. **19**(9), 2238-2246.
57. Gube, M., Taeger D., Weber D. G., et al., *Performance of biomarkers SMRP, CA125, and CYFRA 21-1 as potential tumor markers for malignant mesothelioma and lung cancer in a cohort of workers formerly exposed to asbestos*. Arch Toxicol, 2011. **85**(3), 185-192.
58. Vural, B., Chen L. C., Saip P., et al., *Frequency of SOX Group B (SOX1, 2, 3) and ZIC2 antibodies in Turkish patients with small cell lung carcinoma and their correlation with clinical parameters*. Cancer, 2005. **103**(12), 2575-2583.
59. Atakan, S., *Identification And Utilization of Autologous Anti-Tumor Antibodies For The Diagnosis And Prognosis Of Cancer (Doctoral Dissertation)*. 2015.
60. Patel, H. D., Chalfin H. J., Carter H. B., *Improving Prostate Cancer Screening and Diagnosis: Health Policy and Biomarkers Beyond PSA*. JAMA Oncol, 2016. **2**(7), 867-868.
61. Ju, J., Jiang J., Fesler A., *miRNA: the new frontier in cancer medicine*. Future Med Chem, 2013. **5**(9), 983-985.
62. Trzybulska, D., Vergadi E., Tsatsanis C., *miRNA and Other Non-Coding RNAs as Promising Diagnostic Markers*. EJIFCC, 2018. **29**(3), 221-226.
63. Raisch, J., Darfeuille-Michaud A., Nguyen H. T., *Role of microRNAs in the immune system, inflammation and cancer*. World J Gastroenterol, 2013. **19**(20), 2985-2996.
64. Mishra, P. J., *MicroRNAs as promising biomarkers in cancer diagnostics*. Biomark Res, 2014. **2**, 19.
65. Cheng, T., Zhan X., *Pattern recognition for predictive, preventive, and personalized medicine in cancer*. EPMA J, 2017. **8**(1), 51-60.
66. Chen, X., Ba Y., Ma L., et al., *Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases*. Cell Res, 2008. **18**(10), 997-1006.
67. Montani, F., Marzi M. J., Dezi F., et al., *miR-Test: a blood test for lung cancer early detection*. J Natl Cancer Inst, 2015. **107**(6), djv063.
68. Shah, J. S., Soon P. S., Marsh D. J., *Comparison of Methodologies to Detect Low Levels of Hemolysis in Serum for Accurate Assessment of Serum microRNAs*. PLOS ONE, 2016.
69. Coulibaly F.S., Y. B. B., *Concanavalin A - Polysaccharides Binding Affinity Analysis Using A Quartz Crystal Microbalance*. Biosens Bioelectron, 2014.
70. Jiang, P., Liu Z.-F., Cai S.-M., *Growing Monodispersed PbS Nanoparticles on Self-Assembled Monolayers of 11-Mercaptoundecanoic Acid on Au(111) Substrate*. Langmuir, 2002. **18**(11), 4495-4499.
71. Willey, T. M., Vance A. L., Buuren T. V., et al., *Chemically Transformable Configurations of Mercaptohexadecanoic Acid Self-Assembled Monolayers Adsorbed on Au(111)*. Langmuir, 2004. **20**(7), 2746-2752.
72. Yates, E. A., Terry C. J., Rees C., et al., *Protein-GAG interactions: new surface-based techniques, spectroscopies and nanotechnology probes*. Biochemical Society Transactions, 2006. **34**(3), 427-430.

73. Šimšíková, M., Mariánantalík, Kaňuchová M., Škvarla J., *Anionic 11-mercaptoundecanoic acid capped ZnO nanoparticles*. Applied Surface Science, 2013. **282**, 342-347.
74. Jordon, C. E., Frey B. L., Kornguth S., Corn R. M., *Characterization of Poly-L-lysine Adsorption onto Alkanethiol-Modified Gold Surfaces with Polarization-Modulation Fourier Transform Infrared Spectroscopy and Surface Plasmon Resonance Measurements*. Langmuir, 1994. **10**(10), 3642-3648.
75. Almodóvar, J., Place L. W., Gogolski J., et al., *Layer-by-Layer Assembly of Polysaccharide-Based Polyelectrolyte Multilayers: A Spectroscopic Study of Hydrophilicity, Composition, and Ion Pairing*. Biomacromolecules, 2011. **12**(7), 2755-2765.
76. Leidinger, A. K. P., Vogel B., Backes C., et al., *miRNAs can be generally associated with human pathologies as exemplified for miR-144**. BMC Medicine, 2014. **12**(224).
77. Leidinger, P., Keller A., Borries A., et al., *Specific peripheral miRNA profiles for distinguishing lung cancer from COPD*. Lung Cancer, 2011. **74**(1), 41-47.
78. Macdonald, G. F. H. I. K., Reynolds C., Allen J., Murray A., *Tumor-Associated Autoantibodies: Re-Optimization of EarlyCDT-Lung Diagnostic Performance and Its Application to Indeterminate Pulmonary Nodules*. Scientific Research, 2017. **8**(5), 506-517
79. Healey, G. F., Lam S., Boyle P., et al., *Signal stratification of autoantibody levels in serum samples and its application to the early detection of lung cancer*. J Thorac Dis, 2013. **5**(5), 618-625.
80. Murray, A., Chapman C. J., Healey G., et al., *Technical validation of an autoantibody test for lung cancer*. Ann Oncol, 2010. **21**(8), 1687-1693.
81. Schmidt, R., Jacak J., Schirwitz C., et al., *Single-molecule detection on a protein-array assay platform for the exposure of a tuberculosis antigen*. J Proteome Res, 2011. **10**(3), 1316-1322.
82. Shirane, M., Nakayama K. I., *Inherent calcineurin inhibitor FKBP38 targets Bcl-2 to mitochondria and inhibits apoptosis*. Nat Cell Biol, 2003. **5**(1), 28-37.
83. Soussi, T., *p53 Antibodies in the sera of patients with various types of cancer: a review*. Cancer Res, 2000. **60**(7), 1777-1788.
84. Werner, S., Chen H., Tao S., Brenner H., *Systematic review: serum autoantibodies in the early detection of gastric cancer*. Int J Cancer, 2015. **136**(10), 2243-2252.
85. Kawashima, H., Takatori H., Suzuki K., et al., *Tumor suppressor p53 inhibits systemic autoimmune diseases by inducing regulatory T cells*. J Immunol, 2013. **191**(7), 3614-3623.
86. Gudkov, A. V., Gurova K. V., Komarova E. A., *Inflammation and p53: A Tale of Two Stresses*. Genes Cancer, 2011. **2**(4), 503-516.
87. Gure, A. O., Stockert E., Scanlan M. J., et al., *Serological identification of embryonic neural proteins as highly immunogenic tumor antigens in small cell lung cancer*. Proc Natl Acad Sci U S A, 2000. **97**(8), 4198-4203.
88. Kobold, S., Tams S., Luetkens T., et al., *Patients with multiple myeloma develop SOX2-specific autoantibodies after allogeneic stem cell transplantation*. Clin Dev Immunol, 2011. **2011**, 302145.
89. Shih, J., Rahman M., Luong Q. T., et al., *Dominant B-cell epitopes from cancer/stem cell antigen SOX2 recognized by serum samples from cancer patients*. Am J Clin Exp Immunol, 2014. **3**(2), 84-90.
90. Sun, Y., Zhang R., Wang M., et al., *SOX2 autoantibodies as noninvasive serum biomarker for breast carcinoma*. Cancer Epidemiol Biomarkers Prev, 2012. **21**(11), 2043-2047.
91. Donigan, K. A., Hile S. E., Eckert K. A., Sweasy J. B., *The human gastric cancer-associated DNA polymerase beta variant D160N is a mutator that induces cellular transformation*. DNA Repair (Amst), 2012. **11**(4), 381-390.
92. Dong, Z. M., Zheng N. G., Wu J. L., et al., *Difference in expression level and localization of DNA polymerase beta among human esophageal cancer focus, adjacent and corresponding normal tissues*. Dis Esophagus, 2006. **19**(3), 172-176.
93. Michiels, S., Laplanche A., Boulet T., et al., *Genetic polymorphisms in 85 DNA repair genes and bladder cancer risk*. Carcinogenesis, 2009. **30**(5), 763-768.
94. Uludag, Y., Tothill I. E., *Development of a sensitive detection method of cancer biomarkers in human serum (75%) using a quartz crystal microbalance sensor and nanoparticles amplification system*. Talanta, 2010. **82**(1), 277-282.

95. Catalona, W. J., Smith D. S., Ratliff T. L., et al., *Measurement of prostate-specific antigen in serum as a screening test for prostate cancer*. N Engl J Med, 1991. **324**(17), 1156-1161.
96. Stamey, T. A., Yang N., Hay A. R., et al., *Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate*. N Engl J Med, 1987. **317**(15), 909-916.
97. Nadler, R. B., Humphrey P. A., Smith D. S., et al., *Effect of inflammation and benign prostatic hyperplasia on elevated serum prostate specific antigen levels*. J Urol, 1995. **154**(2 Pt 1), 407-413.
98. Carter, H. B., Pearson J. D., Metter E. J., et al., *Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease*. JAMA, 1992. **267**(16), 2215-2220.
99. Park, S. H., Kim S. Y., Lee W., et al., *New decision criteria for selecting delta check methods based on the ratio of the delta difference to the width of the reference range can be generally applicable for each clinical chemistry test item*. Ann Lab Med, 2012. **32**(5), 345-354.
100. Landgraf, P., Rusu M., Sheridan R., et al., *A mammalian microRNA expression atlas based on small RNA library sequencing*. Cell, 2007. **129**(7), 1401-1414.
101. Zhang, W. C., Chin T. M., Yang H., et al., *Tumour-initiating cell-specific miR-1246 and miR-1290 expression converge to promote non-small cell lung cancer progression*. Nat Commun, 2016. **7**, 11702.
102. Li, J., Dong J., Li S., et al., *An alternative microRNA-mediated post-transcriptional regulation of GADD45A by p53 in human non-small-cell lung cancer cells*. Sci Rep, 2017. **7**(1), 7153.
103. Chen, J., Gao S., Wang C., et al., *Pathologically decreased expression of miR-193a contributes to metastasis by targeting WT1-E-cadherin axis in non-small cell lung cancers*. J Exp Clin Cancer Res, 2016. **35**(1), 173.
104. Zhou, Y., Liang H., Liao Z., et al., *miR-203 enhances let-7 biogenesis by targeting LIN28B to suppress tumor growth in lung cancer*. Sci Rep, 2017. **7**, 42680.
105. You, J., Li Y., Fang N., et al., *MiR-132 suppresses the migration and invasion of lung cancer cells via targeting the EMT regulator ZEB2*. PLOS ONE, 2014. **9**(3), e91827.
106. Inamura, K., *Major Tumor Suppressor and Oncogenic Non-Coding RNAs: Clinical Relevance in Lung Cancer*. Cells, 2017. **6**(2).
107. Fan, T., Wang W., Zhang B., et al., *Regulatory mechanisms of microRNAs in lung cancer stem cells*. Springerplus, 2016. **5**(1), 1762.
108. Ricciuti, B., Mecca C., Crino L., et al., *Non-coding RNAs in lung cancer*. Oncoscience, 2014. **1**(11), 674-705.
109. Berezikov, E., Chung W. J., Willis J., et al., *Mammalian mirtron genes*. Mol Cell, 2007. **28**(2), 328-336.
110. Butkyte, S., Ciupas L., Jakubauskiene E., et al., *Splicing-dependent expression of microRNAs of mirtron origin in human digestive and excretory system cancer cells*. Clin Epigenetics, 2016. **8**, 33.
111. Quezada-Maldonado, E. M., Sanchez-Perez Y., Chirino Y. I., et al., *miRNAs deregulation in lung cells exposed to airborne particulate matter (PM10) is associated with pathways deregulated in lung tumors*. Environ Pollut, 2018. **241**, 351-358.
112. Hu, Y., Liu J., Jiang B., et al., *MiR-199a-5p loss up-regulated DDR1 aggravated colorectal cancer by activating epithelial-to-mesenchymal transition related signaling*. Dig Dis Sci, 2014. **59**(9), 2163-2172.
113. He, J., Wang M., Jiang Y., et al., *Chronic arsenic exposure and angiogenesis in human bronchial epithelial cells via the ROS/miR-199a-5p/HIF-1alpha/COX-2 pathway*. Environ Health Perspect, 2014. **122**(3), 255-261.
114. Ahmadi, A., Khansarinejad B., Hosseinkhani S., et al., *miR-199a-5p and miR-495 target GRP78 within UPR pathway of lung cancer*. Gene, 2017. **620**, 15-22.
115. Sanfiorenzo, C., Ilie M. I., Belaid A., et al., *Two panels of plasma microRNAs as non-invasive biomarkers for prediction of recurrence in resectable NSCLC*. PLOS ONE, 2013. **8**(1), e54596.
116. Mizuno, S., Bogaard H. J., Gomez-Arroyo J., et al., *MicroRNA-199a-5p is associated with hypoxia-inducible factor-1alpha expression in lungs from patients with COPD*. Chest, 2012. **142**(3), 663-672.

7. APPENDIX

Supplementary Table 1: Top 100 Antibody Combination ROC results among based on AUC

Panel Rank	Antibody Combination	AUC	p-value	SCLC (n)	Healthy (n)	Total Sample Size (n)
1	FKBP8 - ENSA - SOX2 - POLB	0.96	1.59E-13	34	31	65
2	FKBP8 - SOX2 - POLB	0.96	2.31E-14	36	34	70
3	FKBP8 - ENSA - PPIA - SOX2 - POLB	0.95	2.69E-13	34	31	65
4	HuD - TUBB4 - FKBP8 - ENSA - SOX2	0.95	3.54E-12	33	27	60
5	FKBP8 - PPIA - SOX2 - POLB	0.95	4.40E-14	36	34	70
6	FKBP8 - DUSP8 - SOX2 - POLB	0.95	5.15E-14	36	34	70
7	FKBP8 - DUSP8 - ENSA - SOX2 - POLB	0.95	5.30E-13	34	31	65
8	FKBP8 - P53 - SOX2 - POLB	0.95	1.00E-13	35	34	69
9	HuD - FKBP8 - ENSA - SOX2	0.95	4.05E-12	34	27	61
10	GON4L - FKBP8 - SOX2 - POLB	0.95	4.09E-13	33	33	66
11	FKBP8 - DUSP8 - ENSA - PPIA - SOX2 - POLB	0.95	7.38E-13	34	31	65
12	FKBP8 - P53 - PPIA - SOX2 - POLB	0.95	1.37E-13	35	34	69
13	TUBB4 - FKBP8 - SOX2 - POLB	0.95	4.81E-13	33	33	66
14	FKBP8 - DUSP8 - PPIA - SOX2 - POLB	0.95	9.56E-14	36	34	70
15	FKBP8 - TIMM44 - SOX2 - POLB	0.95	1.77E-12	33	30	63
16	GON4L - SOX2 - POLB	0.95	5.65E-13	33	33	66
17	GON4L - PPIA - SOX2 - POLB	0.95	5.65E-13	33	33	66
18	GON4L - FKBP8 - PPIA - SOX2 - POLB	0.95	5.65E-13	33	33	66
19	HuD - GON4L - TUBB4 - FKBP8 - ENSA - SOX2	0.95	8.65E-12	33	27	60
20	HuD - TUBB4 - FKBP8 - ENSA - PPIA - SOX2	0.95	8.65E-12	33	27	60
21	FKBP8 - DUSP8 - P53 - SOX2 - POLB	0.95	2.16E-13	35	34	69
22	TUBB4 - FKBP8 - P53 - SOX2 - POLB	0.95	1.13E-12	32	33	65
23	HuD - FKBP8 - ENSA - PPIA - SOX2	0.95	8.07E-12	34	27	61
24	TUBB4 - TIMM44 - SOX2 - POLB	0.95	9.43E-12	31	29	60
25	TUBB4 - DUSP8 - TIMM44 - SOX2 - POLB	0.95	9.43E-12	31	29	60
26	FKBP8 - P53 - ENSA - SOX2 - POLB	0.95	1.95E-12	33	31	64
27	TUBB4 - SOX2 - POLB	0.94	9.05E-13	33	33	66
28	GON4L - TUBB4 - SOX2 - POLB	0.94	9.05E-13	33	33	66
29	GON4L - FKBP8 - DUSP8 - SOX2 - POLB	0.94	9.05E-13	33	33	66
30	TUBB4 - FKBP8 - PPIA - SOX2 - POLB	0.94	9.05E-13	33	33	66

31	TUBB4 - FKBP8 - ENSA - SOX2 - POLB	0.94	4.55E-12	31	31	62
32	FKBP8 - TIMM44 - ENSA - SOX2 - POLB	0.94	1.23E-11	32	28	60
33	GON4L - FKBP8 - SOX2	0.94	1.33E-14	39	38	77
34	GON4L - TUBB4 - FKBP8 - SOX2 - POLB	0.94	1.06E-12	33	33	66
35	GON4L - FKBP8 - DUSP8 - PPIA - SOX2 - POLB	0.94	1.06E-12	33	33	66
36	HuD - GON4L - FKBP8 - ENSA - PPIA - SOX2	0.94	1.45E-11	33	27	60
37	GON4L - FKBP8 - ENSA - PPIA - SOX2 - POLB	0.94	5.36E-12	31	31	62
38	FKBP8 - DUSP8 - TIMM44 - SOX2 - POLB	0.94	4.01E-12	33	30	63
39	TUBB4 - FKBP8 - TIMM44 - SOX2 - POLB	0.94	1.32E-11	31	29	60
40	TUBB4 - FKBP8 - DUSP8 - TIMM44 - SOX2 - POLB	0.94	1.32E-11	31	29	60
41	FKBP8 - DUSP8 - P53 - PPIA - SOX2 - POLB	0.94	3.91E-13	35	34	69
42	GON4L - FKBP8 - ENSA - SOX2 - POLB	0.94	6.31E-12	31	31	62
43	TUBB4 - FKBP8 - ENSA - PPIA - SOX2 - POLB	0.94	6.31E-12	31	31	62
44	HuD - GON4L - FKBP8 - ENSA - SOX2	0.94	1.71E-11	33	27	60
45	HuD - TUBB4 - FKBP8 - DUSP8 - ENSA - SOX2	0.94	1.71E-11	33	27	60
46	GON4L - FKBP8 - P53 - SOX2 - POLB	0.94	2.11E-12	32	33	65
47	TUBB4 - P53 - PPIA - SOX2 - POLB	0.94	2.11E-12	32	33	65
48	TUBB4 - FKBP8 - P53 - PPIA - SOX2 - POLB	0.94	2.11E-12	32	33	65
49	TUBB4 - FKBP8 - ENSA - SOX2	0.94	9.08E-14	37	36	73
50	GON4L - TIMM44 - SOX2 - POLB	0.94	1.56E-11	31	29	60
51	HuD - GON4L - TUBB4 - FKBP8 - SOX2	0.94	1.04E-11	35	27	62
52	FKBP8 - DUSP8 - SOX2	0.94	4.77E-15	42	39	81
53	FKBP8 - P53 - ENSA - PPIA - SOX2 - POLB	0.94	3.66E-12	33	31	64
54	TUBB4 - P53 - SOX2 - POLB	0.94	2.46E-12	32	33	65
55	TUBB4 - FKBP8 - DUSP8 - P53 - SOX2 - POLB	0.94	2.46E-12	32	33	65
56	TUBB4 - FKBP8 - DUSP8 - SOX2 - POLB	0.94	1.67E-12	33	33	66
57	GON4L - TUBB4 - FKBP8 - PPIA - SOX2 - POLB	0.94	1.67E-12	33	33	66
58	TUBB4 - FKBP8 - DUSP8 - PPIA - SOX2 - POLB	0.94	1.67E-12	33	33	66
59	FKBP8 - P53 - PPIA - SOX2	0.94	7.15E-15	41	39	80

60	ENSA - SOX2 - POLB	0.94	2.61E-12	34	31	65
61	TUBB4 - ENSA - SOX2 - POLB	0.94	8.69E-12	31	31	62
62	GON4L - P53 - PPIA - SOX2 - POLB	0.94	2.86E-12	32	33	65
63	GON4L - FKBP8 - P53 - PPIA - SOX2 - POLB	0.94	2.86E-12	32	33	65
64	ENSA - PPIA - SOX2 - POLB	0.94	3.03E-12	34	31	65
65	HuD - TUBB4 - FKBP8 - P53 - SOX2	0.94	1.84E-11	34	27	61
66	GON4L - TIMM44 - PPIA - SOX2 - POLB	0.94	2.17E-11	31	29	60
67	GON4L - TUBB4 - FKBP8 - TIMM44 - SOX2 - POLB	0.94	2.17E-11	31	29	60
68	HuD - GON4L - FKBP8 - SOX2	0.94	1.43E-11	35	27	62
69	GON4L - ENSA - PPIA - SOX2 - POLB	0.94	1.02E-11	31	31	62
70	FKBP8 - ENSA - SOX2	0.94	4.88E-14	40	36	76
71	HuD - GON4L - TUBB4 - ENSA - SOX2	0.94	2.81E-11	33	27	60
72	HuD - GON4L - TUBB4 - FKBP8 - ENSA - PPIA - SOX2	0.94	2.81E-11	33	27	60
73	FKBP8 - DUSP8 - TIMM44 - SOX2	0.94	1.14E-13	39	35	74
74	GON4L - TUBB4 - TIMM44 - SOX2 - POLB	0.94	2.55E-11	31	29	60
75	TUBB4 - DUSP8 - TIMM44 - PPIA - SOX2 - POLB	0.94	2.55E-11	31	29	60
76	TUBB4 - FKBP8 - TIMM44 - SOX2	0.94	3.76E-13	37	34	71
77	FKBP8 - P53 - SOX2	0.94	1.17E-14	41	39	80
78	GON4L - FKBP8 - P53 - ENSA - SOX2 - POLB	0.94	1.73E-11	30	31	61
79	TUBB4 - FKBP8 - P53 - ENSA - SOX2 - POLB	0.94	1.73E-11	30	31	61
80	P53 - SOX2 - POLB	0.94	8.01E-13	35	34	69
81	P53 - PPIA - SOX2 - POLB	0.94	8.01E-13	35	34	69
82	GON4L - P53 - SOX2 - POLB	0.94	3.86E-12	32	33	65
83	GON4L - TUBB4 - P53 - SOX2 - POLB	0.94	3.86E-12	32	33	65
84	GON4L - TUBB4 - P53 - PPIA - SOX2 - POLB	0.94	3.86E-12	32	33	65
85	FKBP8 - TIMM44 - PPIA - SOX2 - POLB	0.94	8.78E-12	33	30	63
86	FKBP8 - DUSP8 - TIMM44 - PPIA - SOX2 - POLB	0.94	8.78E-12	33	30	63
87	TUBB4 - FKBP8 - P53 - SOX2	0.94	5.71E-14	38	38	76
88	TUBB4 - FKBP8 - DUSP8 - TIMM44 - SOX2	0.94	4.31E-13	37	34	71
89	GON4L - TUBB4 - FKBP8 - SOX2	0.94	4.25E-14	39	38	77
90	GON4L - TUBB4 - PPIA - SOX2 - POLB	0.94	3.01E-12	33	33	66
91	GON4L - TUBB4 - FKBP8 - P53 - SOX2 - POLB	0.94	4.48E-12	32	33	65
92	GON4L - FKBP8 - PPIA - SOX2	0.94	4.81E-14	39	38	77
93	FKBP8 - DUSP8 - P53 - SOX2	0.94	1.69E-14	41	39	80

94	HuD - GON4L - TUBB4 - FKBP8 - P53 - SOX2	0.94	2.96E-11	34	27	61
95	FKBP8 - ENSA - PPIA - SOX2	0.94	8.15E-14	40	36	76
96	GON4L - TUBB4 - FKBP8 - DUSP8 - TIMM44 - SOX2 - POLB	0.94	3.51E-11	31	29	60
97	TUBB4 - FKBP8 - SOX2	0.94	5.45E-14	39	38	77
98	TUBB4 - DUSP8 - P53 - SOX2 - POLB	0.94	5.18E-12	32	33	65
99	FKBP8 - DUSP8 - P53 - ENSA - SOX2 - POLB	0.94	7.81E-12	33	31	64
100	FKBP8 - DUSP8 - TIMM44 - ENSA - PPIA - SOX2 - POLB	0.94	3.86E-11	32	28	60

