The Development and Implementation of Single Molecule Protein Assays for Applications in Early Cancer Diagnostics and Single Cell Studies

A dissertation submitted by

Stephanie Mae Schubert

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Chemistry

TUFTS UNIVERSITY

May, 2015

Advisor: Professor David R. Walt

Abstract

Earlier detection of diseases, such as cancer, often leads to improved prognosis. Therefore, the ability to efficiently, accurately, sensitively, and non-invasively screen for cancer at the earliest stages possible is of paramount importance. Furthermore, a better understanding of the basic biology of molecular mechanisms, cellular pathways, and cellular heterogeneity may lead to more personalized and efficient therapies. In order to achieve this task and to also detect rare cells that may also enable earlier disease detection, specific biomarkers must be studied at single molecule resolution within individual cells. This thesis focuses on the detection of protein biomarkers at ultralow levels in serum for early cancer detection and for fundamental single cell studies using single molecule protein counting technology.

This thesis describes the technology, single molecule arrays (SiMoA), that is utilized throughout the included works and discusses the fundamental kinetics behind the method. Chapter 3 describes how a biomarker can be detected at ultralow concentrations in serum prior to palpable tumor formation using a mouse model, indicating the utility of SiMoA as an early cancer detection tool. Chapter 4 describes the use of SiMoA in a panel of breast cancer protein biomarkers and examines the utility of these biomarkers in detecting early stage breast cancer from serum

samples by using supervised multivariate regression models. Chapter 5 demonstrates the use of SiMoA technology as a straightforward approach for counting single protein molecules within single cells. Chapter 6 includes work towards creating a breast cancer mouse model for studying early cancer progression. One appendix is included that includes detailed patient information relevant to Chapter 4.

To my parents, Elaine and Robert Schubert, who have always supported me throughout all of life's twists and turns and to Manuel Palacios, whose unconditional friendship, love, and support gave me the confidence to succeed.

Acknowledgments

First and foremost, I would like to thank my advisor, Prof. David R. Walt for his support, encouragement, and confidence in me throughout my time at Tufts. I thank Prof. Samuel Thomas and Prof. Joshua Kritzer for their helpful input and support throughout all of my committee meetings. I kindly thank Prof. Marsha Moses from Harvard Medical School and Boston Children's Hospital for serving on my thesis committee.

There are several individuals who directly contributed to the works described in this thesis. I would like to acknowledge and thank Dr. Lisa Arendt, Wenhui Zhou, and Prof. Charlotte Kuperwasser from Tufts University School of Medicine, Prof. Rachel Buschsbaum from Tufts Medical Center, as well as Shazia Baig and Dr. Stephanie Walter for their contributions to the work discussed in Chapter 3. I thank Prof. Charlotte Kuperwasser, Prof. Rachel Buschsbaum, Prof. Marsha Moses, Prof. Akhilesh Pandey from the Johns Hopkins University School of Medicine, and Prof. Gail Sonenshein from Tufts University School of Medicine for their helpful conversations in discussing breast cancer biomarkers. I thank Dr. Manuel Palacios for his expert assistance in applying the statistical learning algorithms discussed in Chapter 4. I acknowledge and thank Dr. Stephanie R. Walter and Dr. Maël Manesse for their contributions to the work described in Chapter 5 and also thank Prof. Daniel Chiu and Eleanor Johnson from the University of Washington for helpful conversations regarding single cell

isolation strategies. I thank the U.S. Department of Defense for supporting all of the research described in this thesis.

I would like to thank everyone at Quanterix Corp. who helped me along the way, especially: Ray Meyer, Tomasz Piech, Purvish Patel, Derek DuPont, Jennifer Geldart-Flashman, Melissa Gardel, Dr. Cheuk Kan, Dr. David Rissin, Devon Campbell, and Steve Klose.

I thank my high school chemistry teacher, Miss Struzziero, for instilling in me a genuine curiosity in chemistry, as well as my college professor, Dr. Jaycoda Major, who made chemistry fun and interesting and inspired me to pursue a graduate degree.

Thank you to all of the Walt lab members, past and present, who have truly been a pleasure to work with as both colleagues and friends: Dr. Manuel Palacios, Dr. Maël Manesse, Cary Lipovsky, Shazia Baig, Dr. Stephanie Walter, Dr. Elena Benito Peña, Soyoon Hwang, Dr. Ryan Hayman, Dr. Marcin Rojek, Dr. Shui Nie, Pratyusha Mogalisetti, Dr. Shonda Gaylord, Dr. Mark Hartman, Dr. Kathryn Mayer, Jennifer Geldart-Flashman, Dr. Candice Etson, Dr. Barrett Duan, Prarthana Khanna, Dr. Danlu Wu, Dr. Shudan Bian, Dr. Sarah Brunker, Trinh Dinh, Dr. Bishnu Regmi, Dr. Liangxia Xie, Thy Vu, Aaron Philips, Dr. Yael Zilberman, Dr. Eli Zhang, Nick Iverson, Limor Cohen, and Payel Ghatak. I also had the pleasure to mentor two outstanding undergraduate students during my time at Tufts: Michael Lacy and Erick Velasquez.

Graduate school would not have been bearable without the support and friendship of so many caring and wonderful people. I am lucky to have worked with or simply been friends with the following people, who made the past several years fly by sharing many laughs, lunches, emergency coffee breaks, non-emergency coffee breaks, lab shenanigans, birthday cakes, dance parties, and too many good times to count: Shazia Baig, Dr. Stephanie Walter, Dr. Maël Manesse, Cary Lipovsky, Dr. Elena Benito Peña, Dr. Patricia Gumbley, Dr. Robert Pawle, Dr. Adam Visentin, Joseph Chiarelli, Dr. Ryan Hayman, Soyoon Hwang, Dr. Cristina Zamora, Jennifer Geldart-Flashman, Dr. Amanda Kowalsick, and I know I am probably forgetting some people.

I am forever grateful to have such incredibly strong, loving, and encouraging parents: Robert and Elaine Schubert, thank you for everything you have done for me and for your unconditional love and support. I also thank my siblings, Jen, Chris, Jackie and Cassie for being so supportive, understanding, and loving. Last, but not least, I thank Manuel Palacios for not only being an excellent mentor and lab mate, but for being a constant, caring, and genuine friend and companion. Your love, support, encouragement, and belief in me mean everything. I cannot thank you enough.

Table of Contents

List of Tables

List of Figures

List of Common Abbreviations

AEB ……..… average number of enzymes per bead
ASCO ……...…… American Society of Clinical Oncology
CA 15-3 ……...…. cancer antigen 15-3
CA 27.29 ……...….. cancer antigen 27.29
CDKN2D ……..…… cyclin-dependent kinase inhibitor 2D
CEA ……...……. carcinoembryonic antigen
DCIS ……..…… ductal carcinoma in situ
EDTA …… ethylenediaminetetraacetic acid
ELISA ……...…… enzyme linked immunosorbant assay
ER ……...….. estrogen receptor
HER2 ……...……. human epidermal growth factor receptor 2
PBS ……...….. phosphate buffered saline
PR ……...…….... progesterone receptor
PSA ……...…… prostate specific antigen
RGP ……..……...… resorufin-β-D-galactopyranoside
TNM ……..…… tumor node metastasis
SiMoA ……...…… single molecule array
SβG ……..…… streptavidin-β-galactosidase

Chapter 1

Introduction

1.1 Introduction

Currently available technologies used to screen and diagnose cancer, specifically breast cancer, are insufficient in terms of sensitivity, specificity, and patient experience. There is an overwhelming need to develop more accurate, sensitive, and biologically relevant non-invasive testing methodologies for earlier breast cancer detection. In the United States, breast cancer represents the second highest cancer-related cause of death, after lung cancer. Breast cancer was also the cause of approximately 40,000 deaths in 2014, which represents 6.8% of all cancer-related deaths.[1] The probability of a woman developing breast cancer in the United States is approximately 12.3%.[2] It is estimated that 232,667 new cases of breast cancer were diagnosed in 2014, or 14% of all new cancer cases.[1] On a global scale, breast cancer represents the most common cancer in women worldwide and the second most common cancer overall, accounting for approximately 1.7 million new cases in 2012.[3] Breast cancer is thus both a national and global health issue. The implementation of improved technologies that are both sensitive and specific enough to detect early stage cancers are required to reduce breast cancer mortality.[4] This chapter describes fundamental information about breast cancer, stresses the importance of early detection, discusses current screening practices and approved breast cancer biomarkers, and introduces the topic of ultra-sensitive single molecule technology as a new methodology to improve upon the currently accepted practices.

1.2 Importance of Early Detection

The earlier a disease is detected and diagnosed, the sooner it can be treated, cured, or managed to prevent further complications. In the case of breast cancer, earlier detection indicates better overall survival. **Table 1.1** depicts the relative 5-year survival rate compared to the stage at which breast cancer is diagnosed.[5] Logically, the more advanced the cancer is at diagnosis, the poorer the prognosis and overall survival are. It is thus vital that breast cancer is detected early to enable the best chance for survival as well as to decrease the physical and emotional trauma and financial burdens associated with undergoing cancer therapy.

Table 1.1 Relative 5-year breast cancer survival rates compared to stage at diagnosis (Adapted from Reference 5).

Stage	5-year Relative Survival Rate
0	100%
I	100%
II	93%
III	72%
IV	22%

1.3 Breast Cancer Risk Factors

Several different factors can put women at more risk for developing breast cancer in their lifetime. Age is one of the most obvious risk associated traits, with

95% of U.S. women being diagnosed after the age of 40.[6] Other modifiable factors that have proven to increase the risk of developing breast cancer include drinking alcohol,[7] having a first child after age 30,[8] being overweight,[9] and not breastfeeding.[6] Studies have demonstrated that 5-10% of all breast cancers are the result of inherited mutations, most commonly the BRCA1 and BRCA2 mutations, which are present in less than 1% of the population.[10, 11] Other genetic conditions also increase the risk of developing breast cancer, such as LiFraumeni Syndrome, which involves germline mutations in the TP53 tumor suppressor gene[12] and Cowden syndrome, a disorder that affects multiple bodily systems due to a germline mutation of the PTEN gene.[13] Family history, undoubtedly an undefined underlying genetic propensity, can also play a large role in whether or not one will develop breast cancer, where women with close relatives with breast cancer have a higher risk of developing the disease than those who do not.[6, 14]

1.4 Breast Cancer Staging

Cancer is broadly defined by the National Cancer Institute as the uncontrolled division of abnormal cells that are capable of invading other bodily tissues.[15] Breast cancer is not a single disease, but rather a mix of complex diseases with different clinical, pathological, morphological, and molecular characteristics.[16] The majority, approximately 95%, of breast cancers are characterized as adenocarcinomas. Adenocarcinomas can be further sub-categorized into what are known as carcinoma in situ (CIS), which are non-invasive, and invasive carcinomas.[16]

Ductal carcinoma in situ (DCIS) is a non-invasive malignant epithelial cell proliferation. DCIS is the most common type of non-invasive breast cancer and accounts for 25% of all breast cancers detected via screening.[17] DCIS does not always progress to invasive carcinomas, but increases the likelihood that the patient will either recur with DCIS or develop invasive carcinoma in the future.[17, 18]

Breast cancer is diagnosed using morphological features, such as tumor size, nuclear and cellular characteristics, necrosis, hormonal receptors, histological type, and axillary tumor lymph node status.[19] Breast cancer is staged using a modified version of the tumor node metastasis (TNM) classification system.[20] This system classifies breast cancer into four stages with several sub-stages depending on size, localization, and extension of the primary tumor into the surrounding structures. Also taken into account are the involvement of regional lymph nodes and the presence of metastases.[20] **Table 1.2** shows the most updated staging requirements using this system. 'T' refers to the primary tumor, 'N' refers to the nodes involved, and 'M' refers to metastasis. A legend explaining the TNM score is shown in **Table 1.3**.

Table 1.2 TNM staging for breast cancer. T= tumor; N = lymph nodes; M = metastasis (Adapted from Reference 20).

	T	N	M
Stage 0	Tis	N0	M0
Stage 1A	T1	N0	M0
Stage 1B	T0-T1	N1mi	M0
Stage 2A	T0 T1 T2	N1 N1 N0	M0
Stage 2B	T2 T3	N1 N0	M0
Stage 3A	T0 T1* T2 T3 T3	N2 N2 N2 N1 N2	M0
Stage 3B	T4 T4 T4	N0 N1 N2	M0
Stage 3C	Any T	N3	M0
Stage 4	Any T	Any T	M1

Table1.3 Abbreviated key for TNM breast cancer staging. T= tumor; N = lymph nodes; M = metastasis. Dotted lines indicate stages that are not applicable for the respective T, N, or M characteristics (Adapted from Reference 20).

	T	N	M
0	No primary tumor	No regional lymph node metastases	No evidence of distant metastases
is	Carcinoma in situ	----------	----------
mi	< 0.1 cm	----------	----------
1	≥2 cm	Metastases to movable ipsilateral level I,II axillary lymph nodes	Distant detectable and/or > 0.2 mm
1a	0.1 cm > 1 cm	----------	----------
1b	0.5 >1 cm	----------	----------
1c	1 cm > 2 cm	----------	----------
2	2 cm > 5 cm	Metastases to ipsilateral level I,II axillary lymph node	----------
2a	----------	Metastases in ipsilateral level I,II axillary lymph nodes fixed to one another or to other structures	----------
3	> 5 cm	Metastases to ipsilateral infraclavicular (level III) axillary lymph node(s) with or without level I,II axillary lymph node(s) involvement, ipsilateral internal mammary lymph node(s)	----------
4	Any size with direct extension to chest wall or skin	----------	----------

1.5 Molecular Subtypes of Breast Cancer

As previously mentioned, breast cancer is not a single simple disease but a rather complicated and heterogeneous group of diseases. Through gene expression studies, it has been determined that breast cancer can be categorized into three major molecular subtypes: basal, HER2, and luminal[21-23] (**Table 1.4**). Luminal breast cancer can be further subcategorized into Luminal A and Luminal B. The categories are primarily based on the expression levels of estrogen receptors (ER), progesterone receptors (PR), and human epidermal growth factor receptor 2 (HER2). The functions of ER and PR will be discussed in more detail in Chapter 4. Briefly, ER and PR are hormone receptors located within the cell that are overexpressed in approximately 70% of breast cancers.[24] HER2 is a transmembrane tyrosine kinase receptor that plays a role in nuclear gene activation.[25] When overexpressed, HER2 has been associated with the proliferation of aggressive breast cancer.[43] Also included in distinguishing Luminal A and Luminal B is Ki67, which is a nuclear protein that has been associated with proliferation.[26]

Both the Basal and HER2 subtypes are characterized by having low or absent gene expression of ER and PR. These two subtypes differ in their expression level of HER2, where it is overexpressed in the HER2 subtype and not expressed in the basal subtype. Due to the lack of expression of these three main markers, the basal subtype is often known as triple negative breast cancer (TNBC). Basal breast cancer also commonly has BRCA1 mutations. Luminal breast cancer is typically ER+ and/or PR+, where + indicates overexpression.

Luminal A is typically HER2-, where – indicates low expression, with low Ki67 expression and Luminal B is typically HER2+ or can be HER2- with high K167 expression. The prognosis for Luminal breast cancer tends to be overall better than for both HER2 and basal breast cancer and it is also the most common form of breast cancer.[23, 27]

Table 1.4 The four major molecular subtypes of breast cancer. Subtypes are listed in columns and associated markers and other aspects are listed in the rows (Adapted from References 23 and 27).

	Basal/TNBC	HER2	Luminal A	Luminal B
ER/PR	ER-, PR-	ER-, PR-	ER+ and/or PR+	ER+ and/or PR+
HER2	HER2-	HER2+	HER2- (low Ki67)	HER2+ (or HER2-, high Ki67)
Clinical Features	-Likely high grade and poorly differentiated -BRCA1 mutation common	Likely to be high grade and node positive - TP53 mutation common	Luminal B tends to be higher histological grade than Luminal A	
Prognosis	Generally poor	Generally poor	Generally good; prognosis better for luminal A than Luminal B	
Prevalence	~15%	~15%	~70%	

1.6 Current Breast Cancer Screening Technologies

There are several methods currently used for breast cancer screening, including mammography, magnetic resonance imaging (MRI), and ultrasound. The most common technique, by far, is mammography. This section describes the impact of screening mammography on breast cancer detection as well as the utility of both MRI and ultrasound.

1.6.1 Mammography

Mammograms are x-ray images taken of the breast as it is compressed between two plates, which is an uncomfortable process. Breast cancer presents on mammograms as abnormal masses, architectural distortions, densities, and/or grouped calcifications.[28] Women over the age of 40 have been recommended to receive annual mammograms, since breast cancer risk increases with age.[6, 29]

Mammography was introduced in 1983, but was not widely used for breast cancer screening until 1986.[30] In order to demonstrate the impact that mammography screening has had on breast cancer detection and survival, **Figure 1.1** shows the trends of reported *in situ* and invasive breast cancer cases between 1975 and 2010 with the implementation of mammography screening highlighted.[29, 31] As shown in **Figure 1.1a**, cases of reported in situ, or non-invasive breast cancer, rose rapidly when mammography started to be heavily used for screening, although rates increased faster in women over 50. The rates of DCIS increased so rapidly that the number of DCIS cases in 1992 were approximately 200% higher than anticipated, based on previous trends between 1973 and 1983.[30] The rapid increase is believed to be attributed to the fact that mammography is capable of detecting indolent, or slow growing cancers, that may not have otherwise been detected for 1-3 years.[29, 31] After the rapid spike, incidence rates of in situ breast cancer stabilized from 1999-2010 for women over 50; however, incidence rates for women under 50 continued to slowly rise at approximately 1.9% per year during the same time period.[32]

In terms of cases of invasive breast cancer, the implementation of mammography screening also had a rapid spike in the rate of incidence, primarily in women over 50 who are at greater risk for developing the disease (**Figure 1.1b**). Other risk factors were also associated with increased rates of breast cancer incidence during this time period, such as increased obesity rates, delays in child birth, and having fewer children.[29] A sharp drop in the incidence rate of invasive breast cancer occurred between 2002 and 2003. This drop can be mainly attributed to findings that the use of combined hormonal therapy used for treating postmenopausal women correlated to both higher rates of larger and more advanced breast cancers. After the studies were released, use of this specific type of combined therapy was reduced.[33] The overall incidence rates of breast cancer have remained relatively stable for cases of both in situ and invasive breast cancer since 2004. It has been suggested that a combination of early breast cancer detection and improved treatment have decreased death rates from 1994 to 2010 by as much as 34%.[2, 34]

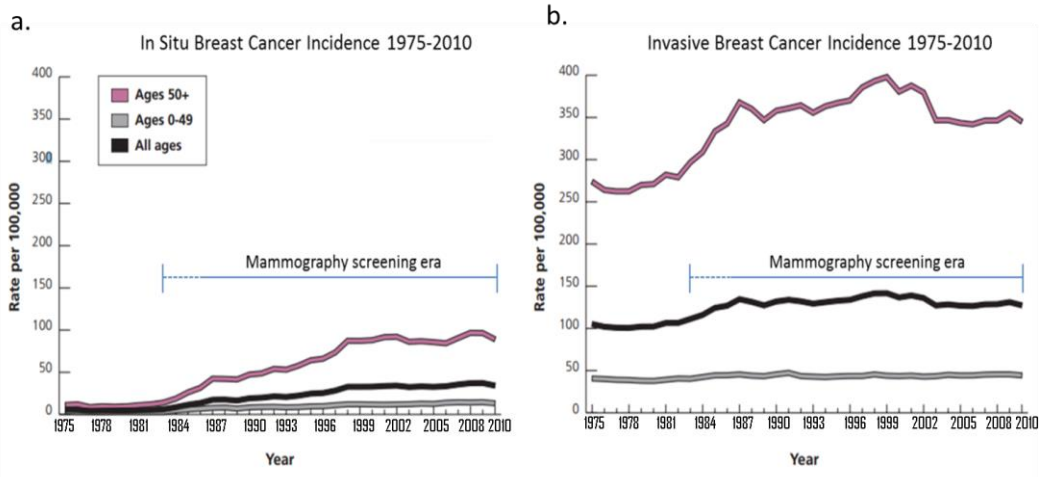


Figure 1.1 Impact of mammography screening from 1975-2010 (Adapted from Reference 29).

More recent studies with evidence correlating mortality rates and breast cancer screening urged the U.S. Preventative Services Task Force (USPSTF) to alter the recommendations for mammography screenings.[35, 36] The new recommendation for women between the age of 50 and 69 is to have biennial mammography screenings. Women in this age group were demonstrated to have a significant decrease in mortality as a result of being screened and therefore benefitted the most.[35] Results were inconclusive for screening women above 69 in terms of the benefits. It was concluded that only women who have family history of breast cancer or who have other risk factors should consider screening before age 49 since the net benefit of screening at this age is small.[36] The goal of these new recommendations is to decrease over-diagnosis and overtreatment.

Despite the improvements in early detection and decreased death rates that mammography has had on breast cancer, it has multiple disadvantages as a screening technique. The main drawback of mammography screening is the often required follow-up biopsies that patients must undergo. There are several different types of biopsies that can be performed. Biopsies involve the removal of tissue from the mass or tumor in question, which is an extremely invasive process. Fine needle aspiration (FNA) offers less invasiveness by utilizing a needle to aspirate cells from the site, however, inadequate sample is often acquired using this technique and the results are dependent on the level of experience of the cytopathologists and radiologists performing the biopsy.[37]

Mammography is also associated with a relatively high false positive rate of 8-10%.[38] False positives may be the result of benign masses, such as cysts or

fibroadenomas, or due to imaging abnormalities caused by thick or dense breast tissue. Approximately 10% of women who undergo mammography screening require a follow-up biopsy. Of the women requiring biopsy, only approximately 20-30% have cancer, meaning that 70–80% of women are biopsied unnecessarily.[39] False positives can cause anxiety and psychological stress as well as financial burden and physical discomfort from additional tests.[40]

Since mammography is strictly an imaging technique, it is not possible to distinguish cancers that may progress from non-progressive cancers. Thus, although mammography may be detecting cancers early through screening, some of the cancers it detects may not progress, leading to overtreatment. In fact, 1 in 3 invasive cancers may be over-diagnosed and some of the cancers detected through screening may regress on their own.[41, 42] Since it is difficult to decipher between a cancer that will or will not progress, they are treated the same, subjecting patients to the adverse effects of harsh cancer treatments unnecessarily.

Since mammography is limited by the ability to see tumors of approximately 5 mm,[43] it is not able to detect tumors that are fast growing or aggressive. For example, **Figure 1.2** shows the progression of four tumors, A, B, C, and D. The arrows depict when the patient had her prescribed screening mammograms. Tumor A progressed too quickly to be detected by mammography screening methods. Tumor B was detected while in a treatable state but is destined to be fatal. Tumor C was eventually detected and was still localized. This tumor did not cause symptoms, progress, or lead to death. Tumor D was microscopic and was below the detection limit of mammography. This tumor was not detected

nor did it progress, cause symptoms, or cause death. From this example, the only case that benefits from mammography screening are patients similar to Tumor B, where the cancer was destined to be fatal but was caught in time.[28]

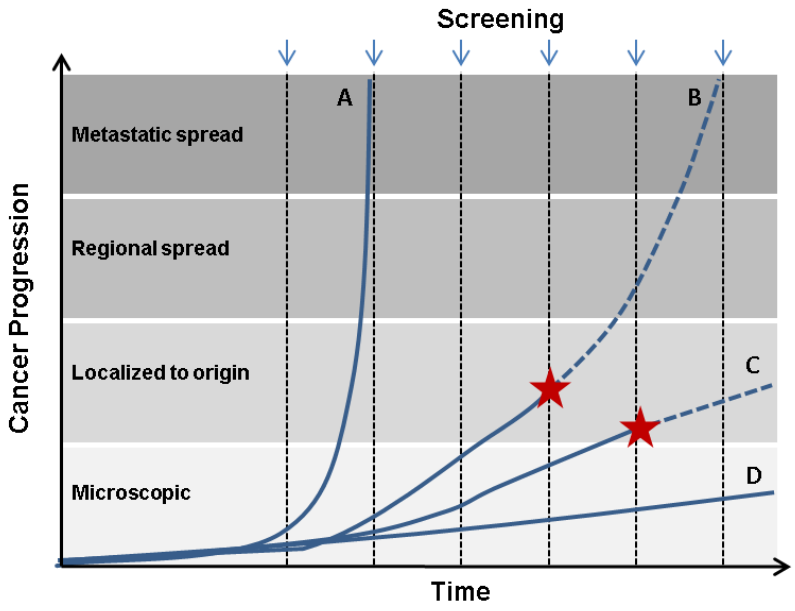


Figure 1.2 Effectiveness of mammography screening on different theoretical tumor growth rates. The growth rates of tumors A, B C, and D are displayed along with mammography screening time points. Tumor A grows too fast to be detected. Tumor B is detected while curable but is destined to be metastatic. Tumor C grows until it can be detected, but does not cause symptoms, progress, or lead to death. Tumor D remains undetectable and without morbidity during the patient's lifetime. Consequently, the only patient that benefits from mammography screening is the patient with Tumor B. Tumor detection is indicated with a red star (Adapted from Reference 28).

Mammography is also associated with false negatives, missing approximately 20% of cancers.[29] This may be due to human error or to the inability to see a tumor mass through dense breast tissue, which is more common among premenopausal women.[29] Since mammograms are x-rays, they expose the patient to potentially harmful radiation. Although the levels of radiation are low - approximately 0.4 mSv or approximately the amount of radiation an average

American would be exposed to in 7 weeks - repeated exposure can increase the causes of cancer.[44]

Mammography is a powerful imaging technique for tumor detection; however, it lacks the ability to decipher benign from cancerous tumors, is unable to detect tumors smaller than ~5 mm,[43] misses approximately 20% of breast cancers potentially present at the time of screening, and has a high false positive rate.[38] These drawbacks lead to inaccurate patient diagnosis, which can allow potentially fatal disease progression, or in the cases of over-treatment, unnecessary physical and emotional trauma.[45] In addition, biopsies are often taken if a tumor is suspected from a mammogram, which is an invasive process. In addition to mammography, several other techniques are available for diagnosing breast cancer, including MRI and ultrasound.

1.6.2 MRI

MRI is another imaging technique that is often used for breast cancer evaluation. Unlike mammography, MRI does not require ionizing radiation and instead uses magnetic interrogation that is not harmful. In comparison to both mammography and ultrasound, MRI is superior at establishing tumor size as well as showing enlarged lymph nodes to suggest metastatic spread.[28] However, since MRI cannot distinguish between lesions that are benign or cancerous and is more expensive than mammography, it is not recommended as a screening tool by itself and is only recommended as a supplement to mammography for women who have

an increased risk for breast cancer.[46] MRI is commonly used in new patient diagnoses to evaluate the extent of disease.[28]

1.6.3 Ultrasound

An additional imaging method sometimes used to follow-up on abnormal findings from screening/diagnostic mammograms or physical exams is ultrasound. Ultrasounds utilize high-frequency sound waves to produce images of structures within the body. The use of ultrasound is helpful in the case of patients with dense breast tissue, which are technically challenging to image using mammography.[28, 29] Ultrasounds are known to have negative predictive values of 99-100%, meaning that if no abnormalities are detected on the ultrasound then it is unlikely that the patient has cancer.[47] It has been reported that ultrasound actually detects more cancer in dense breast tissue than when only mammography is used; however, it has been suggested that ultrasound on its own may lead to higher false positive rates.[48] For this reason, ultrasound is not currently recommended as a screening tool for breast cancer, but rather as a diagnostic supplement.

1.7 Biomarkers Used In Breast Cancer Detection

Screening using only mammography or other imaging technologies increases the detection rate of indolent or potentially non-life threatening cancers, leading to over diagnosis. These methodologies may require painful biopsies when imaging results are suspicious and a significant number of women are biopsied unnecessarily. Current screening technologies are also not sensitive

enough to detect aggressive cancers that will lead to poor prognostic outcomes. A new direction in breast cancer research has focused on the use of biologically relevant biomarkers for breast cancer prevention and detection.[28]

1.7.1 Biomarkers

Biomarkers can be broadly defined as characteristics or objective indications capable of being quantitatively measured to imply either a healthy or diseased state.[49] Ideal biomarkers should elucidate understanding of disease prediction, cause, diagnosis, regression, and/or outcome.[50] Ideally, a biomarker test should be able to accurately, reproducibly, and reliably divide populations into at least two groups associated with significantly different clinical outcomes. The use of biomarker tests should also improve patient outcome to ensure clinical utility.[39, 51] Tumor biomarkers specifically are biomarkers produced by either cancer cells or by the body in response to the presence of cancer. Common types of tumor biomarkers include proteins, RNA, DNA abnormalities and metabolites.[39] A variety of different media can be used for biomarkers testing, but most commonly biomarkers are monitored in tissue as well as in blood, urine, stool, saliva, and breast ductal fluid.[39]

Due to the previously described high rate of biopsies that are performed as a result of the shortcomings associated with mammography, the implementation of biomarker assays may benefit breast cancer screening and diagnostics in several ways. First, more accurate screening processes for breast cancer, either stand-alone methods or in conjunction with mammography, could eliminate or

significantly reduce the number of inconclusive or abnormal mammographies that lead to unnecessary biopsies.[39] Second, biomarkers can be used to monitor therapeutic efficacy to ensure that patients respond appropriately to the treatment provided and to determine any necessary changes in care.[39] Third, the monitoring of biomarkers can indicate disease progression, recurrence, or metastasis. Next, biomarkers can offer prognostic information. Since biomarkers are molecular markers, they also offer insight into the subtype of cancer, which will aid in diagnostics. Finally, sensitive methods of biomarker detection can be used for early disease detection.[19] Currently, no single or multivariate biomarker assay has come close to attaining the above-described analytical or clinical utility requirements for breast cancer screening or diagnosis.[19, 39]

The American Society of Clinical Oncology (ASCO) has only officially recommended a short list of biomarkers as well as one 21-gene recurrence score assay (OncotypeDxR) for clinical applications in treating breast cancer. Four of the approved biomarkers are tissue biomarkers and only two are circulating biomarkers. ER, PR, HER2, OncotypeDxR, and urinary plasminogen activator/plasminogen activator inhibitor 1 (UPA/PAI-1) have all been approved as tissue biomarkers. These biomarkers are used for isolated tissue from biopsies. The only recommended blood serum biomarkers are carcinoembryonic antigen (CEA) and two soluble forms of mucin-1 (MUC-1), cancer antigens 15-3 and 27.29 (CA 15-3 and CA 27.29). [25, 39 52]

1.7.2 Currently Recommended Breast Cancer Biomarkers

The main advantage in being able to monitor biomarkers in blood components, including serum, is that obtaining blood samples is minimally invasive. In order to test tissue samples, a biopsy is typically performed. The following section describes the current status of recommended biomarkers for breast cancer in both blood and tissue.

ER and PR are nuclear steroid hormone receptors that play a critical role in mammary gland biology and are associated with breast cancer progression.[24] In the cases of ER and PR, which are both recommended as tissue-based markers, ASCO suggests that both markers be measured in both pre and post-menopausal women to establish hormone status and determine whether or not patients will benefit from endocrine therapy.[25] Only tumors in which ≥ 1% of the cell nuclei react immunohistochemically to either ER or PR are considered ER+ or PR+, respectively.[53]

Although HER2 status is a hallmark in breast cancer treatment and has been associated with poor prognosis, overexpression of either the HER2 gene or receptor has been observed in various cancers.[25, 54] Despite correlations with poor prognosis in breast cancer, HER2 has not been validated as a prognostic marker by the ASCO. The main use of HER2 is for establishing the benefit of utilizing anti-HER2 therapies, such as trastuzumab. For such purposes, the ASCO recommends that HER2 expression be assessed at the time of diagnosis or recurrence so the appropriate therapy can be prescribed.[25]

UPA is an extracellular matrix degrading protease that has been associated with both cancer invasion and metastasis.[55] PAI-1is a serine protease inhibitor that is instrumental in extracellular matrix degradation as well as in the release of angiogenic factors.[56] The use of an enzyme linked immunosorbant assay (ELISA) to study uPA/PAI-1 was added to the ASCO recommendation for breast cancer diagnostics in 2007. The use of this assay is only recommended for studying breast tissue for prognosis determination of newly diagnosed node negative breast cancer patients. Low levels of uPA/PAI-1 indicate low risk of recurrence and marginal benefits from additional chemotherapy, and vice versa.[25]

In the OncotypeDxR assay, total RNA is extracted from paraffin-embedded tumor tissue. The RNA is treated with DNase I and reverse transcribed to obtain cDNA. Quantitative PCR (qPCR) is then performed on the cDNA products of a total of 21 genes, including 16 breast cancer related genes and 5 reference genes. A recurrence-score algorithm is then normalized using the response from the reference genes to indicate the likelihood of recurrence.[57] ASCO has accepted the use of the OncotypeDxR assay for predicting risk associated with recurrence in cases of newly diagnosed node-negative and ER+ breast cancer. This assay may also be used to predict if patients will benefit from either endocrine therapy or chemotherapy.[25] Drawbacks of this method include the requirement of invasive tissue samples, special training to prevent both RNase and DNase contamination during sample preparation, and possible amplification of unspecific products during qPCR.[58]

In addition to the OncotypeDX assay, other promising gene expression profiling methods not approved by the ASCO have been studied. One example is the MammaPrint microarray or RT-qPCR based gene-expression profiling assay that examines 70 different genes. MamaPrint is the only FDA approved gene-expression assay for prognostic testing of women with node-negative breast cancers, but is not currently included in the ASCO guidelines.[59] Two additional examples include the Rotterdam 76 gene signature and the Mapquant DX. The Rotterdam 76 gene signature claims to be able to predict distant metastasis within 5 years in node negative patients who did not receive chemotherapy regardless of age or tumor size.[59, 60] Mapquant DX is another microarray technique that creates a predictive and prognostic signature in attempts to stratify tumors according to histological grade.[61] Although vital information is obtained from gene expression profiling, including what biomarkers are involved in breast cancer progression and metastasis, the clinical utility of the described tests need to be further explored. As mentioned for the OncotypeDX assay, these methods rely on obtaining tissue samples and examining RNA content, which may require extra skills to prevent contamination. Therefore, the use of gene expression profiling, although extremely important for determining information about the heterogeneity of breast cancer, is not an ideal method for screening or patient monitoring.

MUC1 is a large, glycosolated, transmembrane protein that has been associated with cell adhesion, immunity, and metastasis. MUC1 is often measured in blood via detection of either CA 15-3 or CA.29, which are antigens comprised of different portions of the MUC1 protein.[62-64] CEA is an oncofetal glycoprotein

involved in cell adhesion whose overexpression has been correlated with cancer.[63, 65] Although studies have shown correlations of CEA and CA 15-3 concentrations in blood with both disease progression and metastasis, ASCO only recommends the monitoring of CEA and CA 15-3 or CA 27.29 levels in patients with metastatic disease who are undergoing active therapy.[25 66-68] The use of these biomarkers should only be considered in addition to information acquired from adjunct diagnostic imaging, physical exams, and patient history. However, measurement of CEA and CA15-3 or CA 27.29 may be considered to indicate treatment failure in situations where other methodologies cannot readily measure disease.[25] Since expression of all three markers have been associated with both other cancers and benign conditions,[63] the use of CEA, CA15-3, and CA 27.29 are not recommended for screening, diagnosis, staging, or monitoring purposes.[25, 66]

The currently available biomarker technology for breast cancer is only sufficient following breast cancer diagnosis and to predict therapy. These biomarkers are insufficient for implementation in early breast cancer screening or for other beneficial purposes, such as monitoring therapeutic efficacy, recurrence, or offering prognostic information. The majority of biomarkers require tissue samples, which means that a biopsy must be performed—an incredibly invasive process. For the two markers that have been approved for testing in blood, the markers are not specific enough for breast cancer nor are they sensitive enough for early detection testing. In addition, the ability to discriminate between malignant and benign tumors as well as tumors that will not progress could be incredibly powerful in reducing or eliminating the overtreatment of patients.

Overtreatment is a problem currently plaguing early detection methodologies and none of the described biomarkers or other available technologies are capable of performing the described tasks.[69] It is thus of vital importance to develop biomarker assays that are sensitive and specific for breast cancer. Since a single marker is unlikely to be sufficient for such a task, a panel of biomarkers can be used to improve both analytical and clinical utility.

1.8 Ultra-sensitive biomarker detection in serum

Blood components, such as serum or plasma, are ideal testing media compared to tissue or other bodily fluids. Serum and plasma are thought to contain thousands of protein molecules that arise from either active secretion or leakage from various cells and tissues that comprehensively sample the human phenotype.[70] Although other media can be sampled non-invasively, such as saliva, tears, or urine, these media typically represent only small subsets of what is available in serum and plasma and provide a restricted local picture of cellular activity.[71] In addition, serum and plasma are already the most clinically analyzed proteome, making them easy and safe to obtain.[71] It is believed that the majority of protein biomarkers that exist in serum and plasma have yet to be discovered due to sensitivity limitations.[71] In order to measure relevant biomarkers for the early detection of breast cancer in serum, a new methodology must be implemented that is not in current practice.

Single molecule array (SiMoA) assays, also known as Digital ELISA, are ultra- sensitive ELISAs that are capable of detecting protein molecules in serum at concentrations approximately 1,000× more dilute than traditional ELISAs.[72] SiMoA technology has proven to be effective for the detection of several different types of protein biomarkers in serum, including prostate specific antigen (PSA) for monitoring the recurrence of prostate cancer after radical prostatectomy,[73] as well as tumor necrosis factor-alpha (TNF-α) and interleukin 6 (IL-6) for monitoring therapeutic efficacy in Crohn's disease.[74] Due to the vast improvement in sensitivity, the development of SiMoA assays for breast cancer biomarkers in serum could lead to improved early detection as well as recurrence monitoring and therapeutic efficacy for breast cancer patients.

1.9 Scope of thesis

The present studies aim to demonstrate the applicability of SiMoA as a platform for non-invasive early cancer detection by measuring protein biomarkers in serum. Additionally, a method for isolating single cancer cells and sensitively counting the protein content using SiMoA is described. Chapter 2 describes the SiMoA methodology that is used in subsequent chapters and discusses how single molecule technology gains added sensitivity over traditional ELISAs protein assays. Chapter 3 describes a mouse model using PSA as a gold standard protein biomarker to demonstrate that SiMoA can be utilized to monitor increasing concentrations of biomarkers in blood over time beyond the scope of ELISA and

prior to tumor formation. Chapter 4 describes the design of SiMoA assays for breast cancer biomarkers as well as results from both healthy and breast cancer patients. Chapter 5 discusses the development of a single cell platform where single cancer cells are isolated and SiMoA is used to count the number of protein molecules in each cell. Chapter 6 describes preliminary work to develop a breast cancer mouse model. One appendix is included that contains patient information for samples used in Chapter 4 as well as other relevant information regarding the data analysis in that chapter.

1.10 References

1. SEER.*SEER Stat Fact Sheets: Breast Cancer*.25 February 2015.
2. DeSantis, C., Ma, J., Bryan, L. & Jemal, A. Breast cancer statistics, 2013. *CA: A Cancer Journal for Clinicians* **64**, 52-62 (2014).
3. Ferlay, J. et al.*GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012*.World Health Organization.25 February 2015.
4. Sweep, F.C.G.J., Thomas, C.M.G. & Schmitt, a.M. in Cancer Drug Discovery and Development: Biomarkers in Breast Cancer: Molecular Diagnostics for Predicting and Monitoring Therapeutic Effect. (eds. G. Gasparini & D.F. Hayes) (Humana Press, Totowa; 2006).
5. American Cancer Society.*Breast cancer survival rates by stage*.26 February 2015.Web.6 March 2015.
6. Goldstein, N.S. & Ziegfeld, C.R. Risk Factors and Risk Assessment (Saunders, Philidelphia; 2011).
7. Horn-Ross, P.L. et al. Patterns of Alcohol Consumption and Breast Cancer Risk in the California Teachers Study Cohort. *Cancer Epidemiology Biomarkers & Prevention* **13**, 405-411 (2004).
8. Russo, J., Moral, R., Balogh, G.A., Mailo, D. & Russo, I.H. The protective role of pregnancy in breast cancer. *Breast Cancer Res* **7**, 131-142 (2005).
9. van den Brandt, P.A. et al. Pooled Analysis of Prospective Cohort Studies on Height, Weight, and Breast Cancer Risk. *American Journal of Epidemiology* **152**, 514-527 (2000).
10. Easton, D.F. Familial risks of breast cancer. *Breast Cancer Res* **4**, 179-181 (2002).
11. Schwartz, G.F. et al. Proceedings of the International Consensus Conference on Breast Cancer Risk, Genetics, & Risk Management, April, 2007. *The Breast Journal* **15**, 4-16 (2009).
12. Varley, J.M. Germline TP53 mutations and Li-Fraumeni syndrome. *Human Mutation* **21**, 313-320 (2003).
13. Liaw, D. et al. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat. Genet.* **16**, 64-67 (1997).
14. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *The Lancet* **358**, 1389-1399 (2001).
15. National Cancer Institute.*What Is Cancer?*9 February 2015.25 February 2015.
16. Cell and Molecular Biology of Breast Cancer. (Humana Press, New York; 2013).
17. Pinder, S.E. & O'Malley, F.P. in Breast Pathology. (eds. F.P.O.M.E.P. Editor & J.R. Goldblum) 191-200 (Churchill Livingstone, 2006).
18. Singh, M. & Rittenbach, J.V. in Early Diagnosis and Treatment of Cancer: Breast Cancer. (eds. L. Jacobs & C.A. Finlayson) (Saunders, Philidelphia; 2011).
19. Laguens, G., Coronato, S. & Girolamo, W.D. Biomarkers in Breast Cancer. *Central European Journal of Medicine* **1** (2006).
20. Haydaroglu, A. & Ozyigit, G. (eds.) Principles and Practice of Modern Radiotherapy Techniques in Breast Cancer. (Springer, New York; 2013).
21. Sørlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *PNAS* **98**, 10869-10874 (2001).
22. Fan, C. et al. Concordance among Gene-Expression–Based Predictors for Breast Cancer. *N. Engl. J. Med.* **355**, 560-569 (2006).

23. Schnitt, S.J. Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Mod Pathol* **23**, S60-S64 (2010).
24. Saha Roy, S. & Vadlamudi, R.K. Role of Estrogen Receptor Signaling in Breast Cancer Metastasis. *International Journal of Breast Cancer* **2012**, 8 (2012).
25. Harris, L. et al. American Society of Clinical Oncology 2007 Update of Recommendations for the Use of Tumor Markers in Breast Cancer. *Journal of Clinical Oncology* **25**, 5287-5312 (2007).
26. Dowsett, M. et al. Assessment of Ki67 in Breast Cancer: Recommendations from the International Ki67 in Breast Cancer Working Group. *Journal of the National Cancer Institute* **103**, 1656-1664 (2011).
27. Brenton, J.D., Carey, L.A., Ahmed, A.A. & Caldas, C. Molecular Classification and Molecular Forecasting of Breast Cancer: Ready for Clinical Application? *JOURNAL OF CLINICAL ONCOLOGY* **23**, 7350-7360 (2005).
28. Lewin, J.M. in Early Diagnosis and Treatment of Cancer: Breast Cancer. (eds. L. Jacobs & C.A. Finlayson) 125-139 (Saunders, Philidelphia; 2011).
29. American Cancer Society. *Breast Cancer Facts & Figures 2013-2014.*Atlanta: (2013)
30. Ernster, V.L., Barclay, J., Kerlikowske, K., Grady, D. & Henderson, I. Incidence of and treatment for ductal carcinoma in situ of the breast. *JAMA* **275**, 913-918 (1996).
31. Esserman, L., Shieh, Y. & Thompson, I. Rethinking screening for breast cancer and prostate cancer. *JAMA* **302**, 1685-1692 (2009).
32. Breen, N., Gentleman, J.F. & Schiller, J.S. Update on mammography trends. *Cancer* **117**, 2209-2218 (2011).
33. Writing Group for the Women's Health Initiative Investigators Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the women's health initiative randomized controlled trial. *JAMA* **288**, 321-333 (2002).
34. Berry, D.A. et al. Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer. *N. Engl. J. Med.* **353**, 1784-1792 (2005).
35. Wang, A.T. et al. Impact of the 2009 US Preventive Services Task Force Guidelines on Screening Mammography Rates on Women in Their 40s. *PLoS ONE* **9**, e91399 (2014).
36. U.S. Preventive Services Task Force Screening for Breast Cancer: Recommendations and Rationale. *Ann. Intern. Med.* **137**, 344-346 (2002).
37. Rautiainen, S. et al. Axillary Lymph Node Biopsy in Newly Diagnosed Invasive Breast Cancer: Comparative Accuracy of Fine-Needle Aspiration Biopsy versus Core-Needle Biopsy. *Radiology* **269**, 54-60 (2013).
38. Taplin, S. et al. Mammography facility characteristics associated with interpretive accuracy of screening mammography. *J Natl Cancer Inst* **100**, 876-887 (2008).
39. Paoletti, C. & Hayes, D.F. Molecular Testing in Breast Cancer. *Annu. Rev. Med.* **65**, 95-110 (2014).
40. Goldman, L.E., Walker, R., Miglioretti, D.L., Smith-Bindman, R. & Kerlikowske, K. Facility characteristics do not explain higher false positive rates in diagnostic mammography at facilities serving vulnerable women. *Medical care* **50**, 210-216 (2012).
41. Jørgensen, K.J. & Gøtzsche, P.C. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends, Vol. 339. (2009).

42. Zahl, P., Mæhlen, J. & Welch, H. THe natural history of invasive breast cancers detected by screening mammography. *Archives of Internal Medicine* **168**, 2311-2316 (2008).
43. Michaelson, J. et al. Estimates of the Sizes at Which Breast Cancers Become Detectable on Mammographic and Clinical Grounds. *Journal of Women's Imaging* **5**, 3-10 (2003).
44. American Cancer Society.*Radiation exposure from mammography*.13 February 2015.26 February 2015.
45. Jørgensen, K.J. & Gøtzsche, P.C. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* **339** (2009).
46. Saslow, D. et al. American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography. *CA: A Cancer Journal for Clinicians* **57**, 75-89 (2007).
47. Soo, M.S., Rosen, E.L., Baker, J.A., Vo, T.T. & Boyd, B.A. Negative Predictive Value of Sonography with Mammography in Patients with Palpable Breast Lesions. *American Journal of Roentgenology* **177**, 1167-1170 (2001).
48. Berg, W.A., Zhang, Z., Lehrer, D. & et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening mri to mammography in women with elevated breast cancer risk. *JAMA* **307**, 1394-1404 (2012).
49. Strimbu, K. & Tavel, J.A. What are Biomarkers? *Current opinion in HIV and AIDS* **5**, 463-466 (2010).
50. Mayeux, R. Biomarkers: Potential Uses and Limitations. *NeuroRx* **1**, 182-188 (2004).
51. Teutsch, S.M. et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. *Genetics in Medicine* **11**, 3-14 (2009).
52. Carlson, R.W. et al. Metastatic Breast Cancer, Version 1.2012: Featured Updates to the NCCN Guidelines. *Journal of the National Comprehensive Cancer Network* **10**, 821-829 (2012).
53. Hammond, M.E.H. et al. American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. *Journal of Clinical Oncology* **28**, 2784-2795 (2010).
54. Cook, T. What is HER2? *European Journal of Oncology Nursing* **4**, 2-9 (2000).
55. Duffy, M.J., McGowan, P.M., Harbeck, N., Thomssen, C. & Schmitt, M. uPA and PAI-1 as biomarkers in breast cancer: validated for clinical use in level-of-evidence-1 studies. *Breast Cancer Res* **16**, 428- (2014).
56. Iwadate, Y. et al. High Serum Level of Plasminogen Activator Inhibitor-1 Predicts Histological Grade of Intracerebral Gliomas. *Anticancer Res.* **28**, 415-418 (2008).
57. Paik, S. et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *N. Engl. J. Med.* **351**, 2817-2826 (2004).
58. Pfaffl, M.W. in A-Z of quantitative PCR (ed. S. Bustin) 87-112 (International University Line, La Jolla; 2004).
59. Arango, B.A., Rivera, C.L. & Glück, S. Gene Expression Profiling in Breast Cancer. *American Journal of Translational Research* **5**, 132-138 (2013).

60. Foekens, J.A. et al. Multicenter Validation of a Gene Expression–Based Prognostic Signature in Lymph Node–Negative Primary Breast Cancer. *JOURNAL OF CLINICAL ONCOLOGY* **24**, 1665-1671 (2006).
61. Metzger-Filho, O. et al. Genomic Grade Index (GGI): Feasibility in Routine Practice and Impact on Treatment Decisions in Early Breast Cancer. *PLoS ONE* **8**, e66848 (2013).
62. Duffy, M.J. et al. High Preoperative CA 15-3 Concentrations Predict Adverse Outcome in Node-Negative and Node-Positive Breast Cancer: Study of 600 Patients with Histologically Confirmed Breast Cancer. *Clin. Chem.* **50**, 559-563 (2004).
63. Perkins, G.l., Slater, E.D., Sanders, G.K. & Prichard, J.G. Serum Tumor Markers. *American Family Physician* **68**, 1075-1082 (2003).
64. Duffy, M.J., Evoy, D. & McDermott, E.W. CA 15-3: Uses and limitation as a biomarker for breast cancer. *Clin. Chim. Acta* **411**, 1869-1874 (2010).
65. Benchimol, S. et al. Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. *Cell* **57**, 327-334 (1989).
66. Laessig, D. et al. Importance of CEA and CA 15-3 during Disease Progression in Metastatic Breast Cancer Patients. *Anticancer Res.* **27**, 1963-1968 (2007).
67. Ebeling, F.G. et al. Serum CEA and CA 15-3 as prognostic factors in primary breast cancer. *Br J Cancer* **86**, 1217-1222 (2002).
68. Lee, J. et al. Elevated levels of serum tumor markers CA 15-3 and CEA are prognostic factors for diagnosis of metastatic breast cancers. *Breast Cancer Research and Treatment* **141**, 477-484 (2013).
69. Esserman, L.J. et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *Lancet Oncol* **15**, e234-e242 (2014).
70. Pieper, R. et al. The human serum proteome: Display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. *Proteomics* **3**, 1345-1364 (2003).
71. Anderson, N.L. & Anderson, N.G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Molecular & Cellular Proteomics* **1**, 845-867 (2002).
72. Rissin, D.M. et al. Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nat. Biotechnol.* **28**, 595-599 (2010).
73. Wilson, D.H. et al. Fifth-Generation Digital Immunoassay for Prostate-Specific Antigen by Single Molecule Array Technology. *Clin. Chem.* **57**, 1712-1721 (2011).
74. Song, L. et al. Single molecule measurements of tumor necrosis factor α and interleukin-6 in the plasma of patients with Crohn's disease. *J. Immunol. Methods* **372**, 177-186 (2011).

Chapter 2

Methodology

2.1 Introduction

One of the most significant challenges that restrict biomarkers studies for early cancer detection is the limited sensitivity of existing methods to detect useful biomarkers that may be indicative of early stage disease. As previously discussed in Chapter 1, a significant portion of currently studied biomarkers are proteins. In order to sensitively measure the concentration of biomarkers at the earliest stages of cancer, a methodology capable of detecting low abundant proteins must be used. This chapter discusses the methodology and fundamental principles of SiMoA, the technology used throughout this work. First, a general overview of ELISA is provided. Next, the use of SiMoA in optical fibers is introduced and discussed. A fully automated SiMoA platform is then described followed by the principles and fundamentals of SiMoA. Finally, a general materials and methods section provides details for how SiMoA experiments are performed.

2.2 ELISA

ELISA is both a rapid and convenient method for protein detection that can be both quantitative and qualitative and is often used for the detection of protein analytes in blood or blood components, such as serum or plasma.[1-3] ELISA is a widely used immunoassay for protein detection and is also the current gold standard for analyte quantitation in clinical samples.[3, 4] In ELISA assays, antibodies are used to specifically recognize a protein or antigen of interest by

forming an immunocomplex.[2] There are several forms of ELISA assays, such as direct, indirect, sandwich, and competitive ELISA.[5] Of these formats, sandwich ELISA offers the ability to quantify low concentrations of antigen; in the range of ng/mL, or in some cases, pg/mL, of protein.[2, 6]

In sandwich ELISA formats, a capture antibody is adsorbed to the bottom of a well, typically in a 96-well plate. A sample containing the antigen is then added and allowed to bind with the surface bound capture antibodies. A second antibody, known as a detection antibody, is added, followed by subsequent labeling with an enzyme, typically via biotin-streptavidin binding. Some commonly used enzymes in ELISA are alkaline phosphatase, horseradish peroxidase and streptavidin-β-galactosidase (SβG). When substrate is introduced, the enzyme molecules turn over the substrate molecules to produce product molecules. The substrate used typically undergoes a color change or becomes fluorescent upon turn over to facilitate detection. Several substrates are available for each enzyme, depending on whether the desired readout is colorimetric or fluorescent. Since each enzyme is capable of turning over many substrate molecules to produce product molecules, the signal from each molecule is amplified, a process known as signal amplification (**Figure 2.1**).

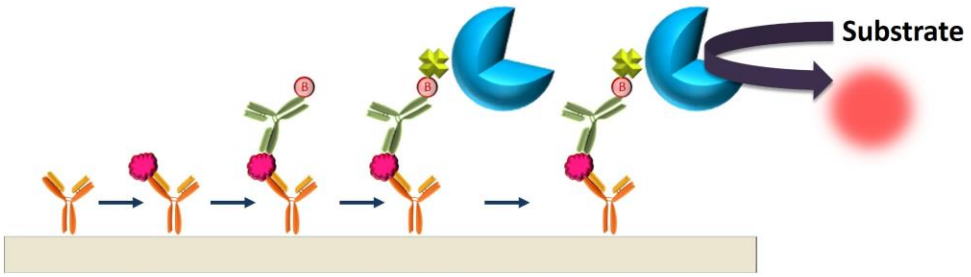


Figure 2.1 Scheme of traditional sandwich ELISA. Capture antibodies (orange) are adsorbed onto the surface of a 96-well plate. When target molecules (pink) are close enough to the surface, they bind to capture antibodies. Biotinylated detection antibodies (green with red circle for biotin tag) are added to tag all captured proteins. A streptavidin labeled enzyme (blue and yellow) binds to the biotinylated detection antibody. When substrate is added, a colorimetric or fluorescence (large red circle) readout is obtained to determine protein concentration.

Sandwich ELISAs are specific in that different antibodies can be used to target a specific protein of interest.[3] The resulting enzymatic amplification leading to product release within the wells is proportional to the amount of target present and can be quantified.[2, 3] Since ELISA is so widely used, kits are available from many vendors and results are generally reproducible.[3] The scope of this thesis focuses on the use of sandwich ELISA, and thus from here on the term "ELISA" will strictly refer to sandwich ELISA.

Despite the many benefits of using traditional ELISA, there are two major disadvantages. Since traditional ELISA technology is performed in wells containing volumes of approximately 100 µL or more, in order to detect a distinguishable change in color/fluorescence above background, a large number of product molecules must be produced. Despite signal amplification from the enzyme, if the concentration of target is too low, enough target molecules within

the solution may not bind to the capture antibodies on the plate surface to result in enough signal to read above the background. This requirement raises the detection limit of ELISA, making it less sensitive. Second, since one type of capture antibody is coated to the well to produce product, ELISA is limited by its inability to multiplex, or simultaneously investigate multiple biomarkers within the same sample. In order to probe a sample for more than one biomarker using ELISA, multiple wells must be used, which requires more sample volume and decreases throughput.[3] The use of smaller wells and thus lower volumes would lead to the generation of more concentrated products from smaller initial target concentrations, facilitating lower detection limits. The following section describes an ultra-sensitive, single molecule ELISA platform to count individual protein molecules in femtoliter sized wells, known as SiMoA.

2.3 SiMoA in Optical Fibers

The theoretical limit of detection of an ultra-sensitive assay is a single molecule.[1] SiMoA enables the sensitive detection of protein molecules by capturing individual molecules, isolating them, and digitally counting them. SiMoA is very similar to traditional ELISA assays in that it uses a sandwich format. However, instead of using the surface of wells in a 96-well plate to adsorb the capture antibodies, capture antibodies are instead covalently linked to magnetic microspheres (**Figure 2.2**). Samples containing low abundance proteins, several thousand to tens of thousands of molecules per 100 μL, are added to approximately 500,000 microspheres, each containing hundreds of thousands of

capture antibodies on the surface. Statistically, due to the significant excess of beads in comparison to target protein molecules, some beads will capture only one protein molecule and the majority of beads will remain unbound. This process follows a Poisson distribution, which is explained further in a following section. The beads are allowed to incubate with the sample at room temperature while shaking for 2 hours. After the protein molecules are captured by the beads, the beads are washed to remove the remaining complex media from the sample. Since the beads are magnetic, this step is performed using magnetic separation. Next, an excess amount of biotinylated detection antibody is added to the beads and incubated for 1 hour with shaking at room temperature. The detection antibody is added in excess to ensure that all of the captured protein molecules will be labeled. This step is also followed by several washing steps to remove any remaining unbound or nonspecifically-bound detection antibody. Finally, SβG, the enzyme used in SiMoA, is added to complete the immunocomplex. The substrate used is resorufin-β-D-galactopyranoside (RGP), which is not fluorescent. However, upon hydrolysis catalyzed by SβG, RGP yields the fluorescent product resorufin, which is measured at excitation/emission wavelengths of 577/620 nm.

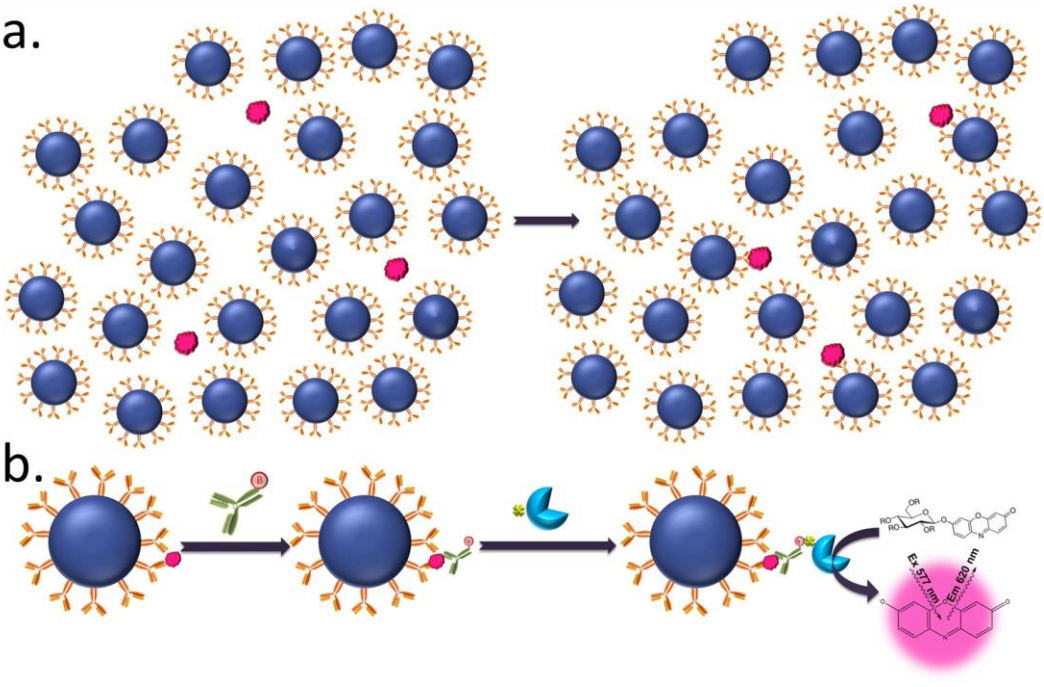


Figure 2.2 Scheme of SiMoA using beads. **a)** Excess beads are incubated with target molecules such that some beads will bind to only one molecule and most beads will remain unbound. **b)** Scheme showing one bead that captured a protein target. Beads are incubated with biotinylated detection antibodies followed by SβG. RGP is then turned over to produce the fluorescent product resorufin.

As previously mentioned, one of the major disadvantages of ELISA is that in the wells of standard 96-well plates and at low sample concentrations, the amount of substrate produced cannot be measured because it is being diluted in a relatively large volume. In order to measure the fluorescent signal being produced from single molecules, SiMoA first utilized micron-sized wells created in optical fiber bundles.[7] To create these wells, optical fiber bundles containing approximately 50,000 individual fibers are polished and then etched in a weak hydrochloric acid solution. The core and the cladding materials of the fibers etch at different rates, leading to the formation of micro-wells (**Figure 2.3**). The resulting array consists of wells that are approximately 4.5 μm wide and 3.25 μm

deep and hold a volume of about 50 fL – or about 2 billion times less than the 100 μL volume usually used in ELISA.

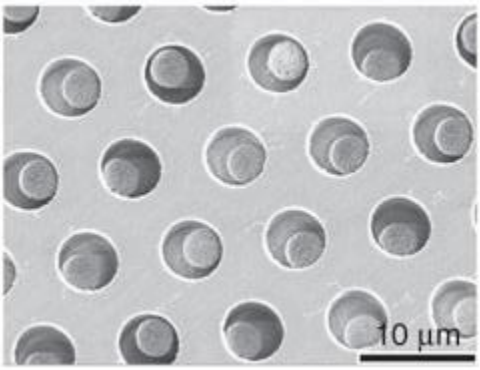


Figure 2.3 Scanning electron micrograph of wells formed in a fiber optic array after bead loading (Reprinted with permission from Reference 7).

Beads are loaded into the wells of the array via centrifuge. The beads used are ~2.7 μm in diameter and, thus by size exclusion, only one bead can physically fit in each well. The loading efficiency using this technique is approximately 50%, and only about 20,000-25,000 beads are loaded into the array (**Figure 2.4**). The wells are enclosed via mechanical sealing. The seal consists of a glass slide covered with a thin sheet of polydimethylsiloxane (PDMS). A droplet of RGP is placed on the PDMS seal such that the wells are filled with substrate upon sealing. Good sealing is vital to prevent leaking of fluorescent product into neighboring wells. Only wells that contain a bead and an immunocomplex will create a fluorescent signal. Since the volume of the wells is so small, the local concentration of product is high enough to be easily detected using standard fluorescence microscopes and an uncooled CCD camera. The concentration of analyte is determined by digitally counting the number of wells that light up and

comparing it to the number of total beads present. In order to reduce background signal and eliminate possible contaminants, only wells that contain a bead and increase in fluorescence by 20% over five frames (due to enzymatic activity) are considered to be a positive signal. The dynamic range of SiMoA can be extended beyond the single molecule regime. Samples containing higher protein concentrations will lead to beads with multiple enzyme labels. In this case, analog measurements are made where the average fluorescence intensity of each bead is quantified. The dynamic range of SiMoA spans from 10^{-19} M to 10^{-13} M.[8] In order to increase throughput, strips containing eight fiber bundles can be assembled via custom glass holders.

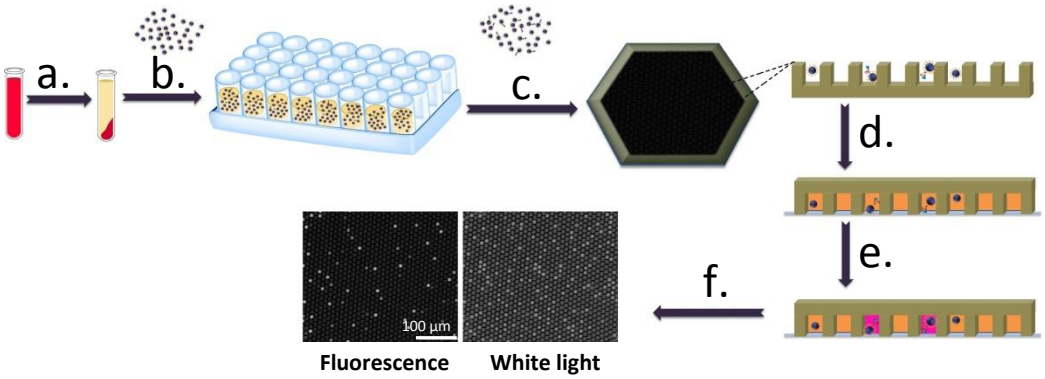


Figure 2.4 Schematic of SiMoA assay in a fiber. **a)** Spin and coagulate whole blood sample to obtain serum. **b)** Incubate beads with 100uL of serum. Follow with incubations with detection antibody and SβG to create complete immunocomplexes. **c)** Load beads with complete immunocomplex into fiber array. **d)** Seal with substrate. **e)** Fluorescent product is produced in wells containing a bead with an enzyme. **f)** Fluorescence and white light images are taken of the entire fiber. Only wells that increase in fluorescence intensity by 20% over five frames are considered a true positive.

Despite the incredible increased sensitivity that SiMoA presented in optical fibers, it still suffers from a few limitations. Preparation time for experiments is lengthy. Fibers must be polished uniformly, which can take up to an hour for every fiber strip. In order to obtain optimal sealing, fibers are then silanized, which requires preparation under nitrogen and large quantities of ethanol. This process takes approximately 4 hours for up to 120 fibers. Seals must also be prepared, which involves precisely cutting pieces of PDMS to fit onto glass cover slides, cleaning the PDMS, and carefully placing the PDMS on the glass slide such that there are no air bubbles. Since optical fibers are made of glass, they sometimes break during imaging, losing valuable data. There are multiple and repetitive pipetting steps, and all reagents are pipetted manually, although washes can be performed on automatic washers. In order to load beads into the fibers, tubing must be affixed to the fibers as holders for the bead solution. Bead loading is not always uniform using this methodology and coefficients of variation (CVs) were highly variable. The number of steps involved in this process and all of the manual pipetting introduces potential errors as well as low reproducibility. In order to obtain triplicate measurements and calibration data for each experiment a large number of fibers must be used. In general, this methodology is extremely tedious, time consuming, and requires expert users to achieve optimal results.

2.4 Fully Automated SiMoA

In 2013, Quanterix Corp. released the HD-1 Analyzer, a fully automated instrument with integrated analysis software capable of performing SiMoA assays with significantly less preparation time than fiber assays.(**Figure 2.5a**)[9] The wells used in the automated platform are of the same size used for fiber assay, but the array size is significantly larger; 216,000 wells compared to the 50,000 wells in each optical fiber. In place of individual fibers, the larger arrays are contained within a microfluidic device, or 'disc' in the shape of a standard DVD, where each disc holds 24 3×4 mm arrays (**Figure 2.5b-d**).[10] The wells are contained within a microfluidic channel that is 500 μm deep and contains both an inlet and vent port. After samples have been incubated with the appropriate reagents according to the designated protocol described above, the beads are resuspended in RGP and are loaded into the disc array via the fluidic inlet port, sealed with oil and imaged. When compared to fiber arrays, the use of discs dramatically reduces the required preparation time since no polishing or silanization are necessary and up to 24 experiments can be performed on one disc.

The HD-1 Analyzer utilizes a precise automated liquid handling system to add all reagents and samples to cuvettes for processing, thus eliminating the majority of manual pipetting steps and limiting opportunities to introduce errors. The use of liquid handling generally produces reduced CVs compared to fiber assays performed with manual pipetting. CVs using the HD-1 Analyzer are consistently under 10%. In addition, cuvettes containing samples are processed in succession such that the timing of each sample is exactly the same throughout the

SiMoA process, which is not the case when performing washes and reagent additions by hand.

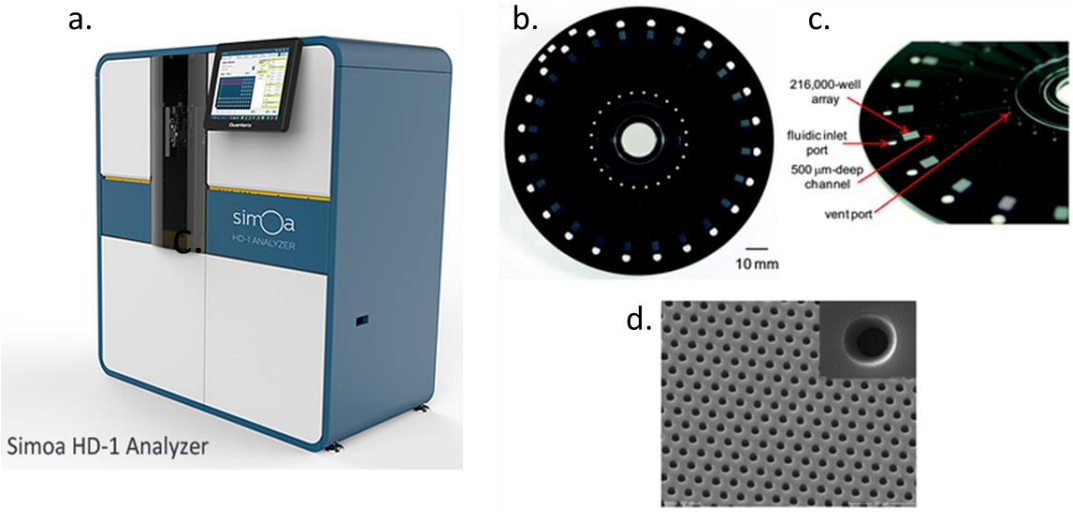


Figure 2.5 Simoa HD-1 Analyzer and Simoa discs. a) Picture of Simoa HD-1 Analyzer instrument. b) Top view of Simoa disc c.) Close up view of Simoa disc showing channel features, including array, fluidic inlet, channel, and vent port. d) SEM images of 216,000 well array in disc and inset is close-up of one individual well. Wells are 4.2 µm wide and 3.3 µm deep (Reprinted with permission from Reference 10).

In place of PDMS seals, a fluorocarbon oil is flowed through the channel to seal the wells. Vacuum is applied to the channel exit to facilitate bead loading and sealing. The oil also removes any excess beads that were not loaded into the wells. Once the excess beads and substrate are removed, the beads are effectively sealed inside the wells and can generate locally high concentrations of fluorescent product (**Figure 2.6**).[10]

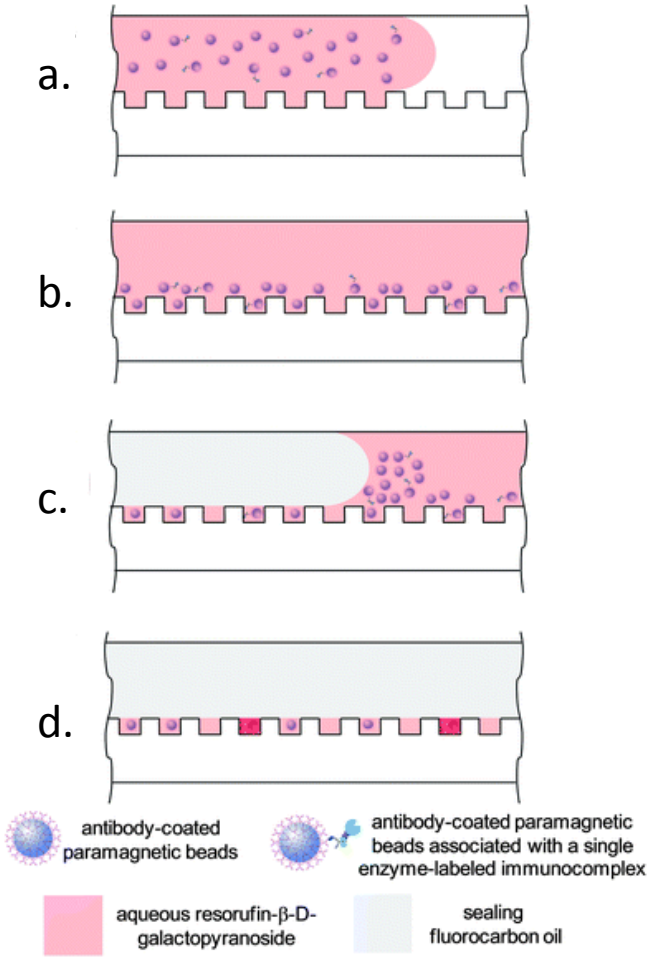


Figure 2.6 Schematic describing the oil sealing process in disc arrays for fully automated SiMoA. **a)** A mixture of beads with and without enzyme complexes suspended in RGP is flowed into the array via pipette. **b)** Beads are allowed to settle into the array via gravity **c)** A fluorocarbon oil is flowed through to remove excess RGP and beads and to seal the wells **d)** The channel is filled with oil and all wells are isolated, enabling imaging of the fluorescent signals generated by individual molecules (Reprinted with permission from Reference 10).

Reagent bottles are used to hold beads, detector antibody, SβG, sample diluent, and RGP. Samples can be loaded in either a 96-well plate or test tube rack. The HD-1 Analyzer performs all incubation and wash steps, loads the reagents, images, and processes the data. Assays can be performed in 3, 2, or 1

step, where the steps indicate the number of incubations. Standard 3-step assays consist of a 15 minute target incubation, 5 minute detector antibody incubation, and 5 minute SβG incubation. Standard 2-step assays consist of a 35 minute incubation of both target and detector antibody and a 5 minute SβG incubation. One step assays are not common as they only call for one washing step, which may result in higher background. These short incubation times paired with full automation enable extremely high throughput sampling: 96 samples can be fully processed in under 3 hours. This processing time is significantly shorter than fiber assays and is even shorter than most ELISA protocols.

Not only does the HD-1 maximize throughput, minimize CVs, and reduce assay time and preparation, but it also enables multiplexing by using dye-encoded beads, which can be incorporated with fiber assays as well, but is not possible using standard ELISA.[11] Multiplexing has been demonstrated using up to four different analytes in plasma samples with no loss in sensitivity.[11] The general protocols for the preparation of SiMoA reagents as well as SiMoA assays used throughout this work are described in the Materials and Methods section below.

2.5 Principles of SiMoA

2.5.1 Digital Counting of Molecules

As described previously, at low concentrations, a majority of beads will contain no enzyme complexes ("off" beads). In order to digitally count molecules, the fraction of beads that are associated with enzyme complexes (active or "on"

beads) is used to determine the protein concentration via Poisson Statistics.[8] The Poisson distribution (**Equation 1**) describes the probability that a discrete number of events will occur within a fixed time interval if the average number of events is known, where μ is the expected average number of events (e.g. average number of labeled molecules/number of beads), ***x*** is the number of expected events (e.g. 0, 1, 2, 3…), and *P* is the probability of observing ***x*** events.[12] By calculating values of μ from SiMoA data, one is actually calculating the average number of enzymes per bead, or the AEB.

$$P_{\mu}(x) = \frac{e^{-\mu}\mu^{x}}{x!} \qquad \text{(Eq. 1)}$$

Due to the static heterogeneity of enzymes, a bead can only be considered 'off' at $x = 0$. Therefore, if **Equation 1** is solved for $P_{\mu}(0)$ while also assuming that the fraction of off beads (f_{off}) can be defined as $1 - f_{on}$, where f_{on} is the fraction of 'on' beads (total active beads/total beads), the AEB for the digital regime can be described as (**Equation 2**):[8]

$$\mu = AEB_{digital} = -\ln[1 - f_{on}] \qquad \text{(Eq. 2)}$$

2.5.2 Analog Counting in SiMoA

When in the digital regime, the number of counted "on" beads is independent of the fluorescence intensity generated in each well. However, once the ratios of enzymes to beads get higher, digital counting is no longer accurate. In order to extend the dynamic range of SiMoA, an analog approach is used when >70% of the beads in an array are active.[8] This approach determines the average

fluorescence intensity of a bead in a given well ($\bar{I}_{bead}$**)** to determine the AEB. Digital measurements are required in order to obtain analog values so that the average intensity of a single molecule ($\bar{I}_{single}$**)** for that assay can be established. **Equation 3** describes the derivation of AEB_{analog}:[8]

$$\boldsymbol{AEB_{analog}} = \frac{\boldsymbol{f_{on} \times \bar{I}_{bead}}}{\boldsymbol{\bar{I}_{single}}} \quad \text{(Eq. 3)}$$

By combining both digital and analog quantification, SiMoA is able to span over four orders of magnitude for PSA concentration, a gold standard assay.[8]

2.5.3 SiMoA Sensitivity and Efficiency

By utilizing bead-based capture, SiMoA strives to both capture and label all, or close to all, of the target protein molecules in each sample. The efficiency of these processes is what leads to the unprecedented sensitivity of this technique.[13] As described above, there are three major binding events that occur during standard SiMoA assays: 1- target capture, 2- detector binding, and 3- SβG labeling. Chang *et al.* has described the conditions that lead to SiMoA's excellent capture efficiency by using simple kinetics.[13]

Each capture bead used in SiMoA is covered with approximately 274,000 antibodies, which correlates to a capture antibody concentration of 2.3 nM when using 500,000 beads in 100 uL.[13] The capture efficiency of SiMoA has been calculated using dissociation constants (K_D) ranging from 10^{-8} M to 10^{-11} M. The results of this study demonstrated that effectively all of the proteins in solution

can be captured when $K_D \leq 10^{-10}$ M. For K_D values that are slightly higher ($K_D=10^{-9}$ M), the efficiency drops to about 70%.[13]

It has also been demonstrated that the optimal antibodies for protein capture have high on rates, where $k_{on} = 10^5$-10^6 $M^{-1}s^{-1}$. However, SiMoA is also very efficient at labeling the proteins that are captured; where close to 100% of captured proteins are labeled with detection antibodies with K_D values of 10^{-9} M at concentrations varying from 1-100 nM. In addition, assuming a K_D of 10^{-15} M for biotin-streptavidin binding, it has been determined that the efficiency of binding for SβG is approximately 94% at 150 pM.[13] Thus, even in assays where low capture efficiencies result in a low percentage of bound protein, effectively all of the bound protein can be labeled.

The concentration of SβG used also has an effect on the assay LOD by leading to increased background signal. At lower concentrations of SβG (e.g. 15 pM), lower percentages of captured molecules are labeled intentionally to reduce non-specific binding and lower the background signal.[13] By reducing the concentration of labeling reagents, the background signal in SiMoA is equivalent to approximately 100 active beads. The overall Poisson noise is thus reduced to ≤ 10%, where Poisson noise is defined as $\sqrt{N}/N$ and N is the number of active beads.[13, 14]

In addition, it has been observed that the off-rates of proteins bound to immobilized antibodies are lower in comparison to antibodies in solution. Explanations for the differences in off-rates are likely due to possible multivalent interactions as well as multiple rebinding effects. The lower off-rates of the bound

protein and antibody complexes on the beads in SiMoA enable vigorously washing, further decreasing background signals by washing away lower affinity proteins.[13, 14] Due to efficient capture efficiency and labeling, as well as low off-rates leading to rebinding and polavalency effects which then lead to low background, SiMoA is able to detect proteins at concentrations in the femtomolar regime using antibodies that may only have nanomolar affinity.

One of the limitations of SiMoA that may detract from its ability to detect 'every' protein molecule in solution, is the low efficiency associated with bead loading. Of the 500,000 beads used for the assay, only 25,000–30,000 beads are loaded and analyzed in the 216,000 well array, which equates to an overall efficiency of < 6%. The ability to load and count more beads may enable detection at even lower concentrations, especially for assays that have low capture efficiency.

2.6 Conclusion

SiMoA, both in the fiber optics and the fully-automated commercialized versions, is an extremely powerful technique that has vastly improved upon the current gold standard ELISA. SiMoA is both a sensitive and specific technique that utilizes beads to obtain extremely high capture efficiencies of low concentrations of target molecules despite sub-optimal antibody binding affinities. The use of both digital and analog quantification strategies enables sampling over a span of four orders of magnitude in concentration. The ability to create high throughput SiMoA assays to test many samples in replicates with low CVs and

limited preparation time makes SiMoA an ideal tool for academia, industry, and clinical settings.

2.7 Materials and Methods

Simoa discs, Wash Buffers 1&2, HD-1 pipette tips, cuvettes, sealing oil, RGP, 2.7 μm carboxyl magnetic beads, Bead Diluent, SβG Diluent, SβG concentrate, Detector Diluent, Bead Conjugation Buffer, Bead Biotinylation Buffer, Bead Wash Buffer, and Bead Blocking Buffer were all purchased from Quanterix Corp. 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride **(**EDC**)** was purchased from Thermo Pierce Scientific. Reagent bottles were purchased from VWR and 96-well plates were from Axygen. Amicon centrifugal spin filters were purchased from Millipore. 1.7 mL microcentrifuge tubes were purchased from Eppendorf. Biotin reagent was purchased from Thermo Scientific. Antibodies were purchased from various sources as indicated in each chapter. All water used was Milli-Q water. All buffers are filter sterilized using various sized vacuum filtration systems with PES membranes purchased from VWR.

2.7.1 Bead Conjugation

Approximately 0.12 mg of capture antibody is buffer exchanged into Bead Conjugation Buffer (Quanterix, 101357) using Amicon centrifugal spin filters (Amicon Ultra 50KD, UFC50596) by adding the antibody to the spin column and filling the column to 500 μL with Bead Conjugation buffer. The columns are spun at 14,000 g for 5 min. The supernatant is removed and followed with two consecutive washes with 500 μL of Bead Conjugation Buffer. To recover the

antibody, the column is inverted inside a new tube and spun at 1,000 g for 2 min. The solution is kept, and the column is washed with 50 μL of Bead Conjugation Buffer to remove residual antibody from the column and spun for an additional 2 min at 1000 g. The buffer exchanged antibody concentration is measured by reading the absorbance at 280 nm using a NanoDrop 1000 (Thermo Scientific). A 200 μL solution of the buffer exchanged antibody is prepared in Bead Conjugation Buffer at 0.5 mg/mL and is kept on ice. Approximately 2×10^8 (100 uL) of stock beads (Quanterix, 101360) are washed three times with 200 μL Bead Wash Buffer (Quanterix, 101355) and 2× with 200μL Bead Conjugation Buffer. Washes are performed by vortexing the liquid and quickly centrifuging the tube to remove liquid from the cap. Beads are collected using a magnet so the supernatant can be removed. Beads are then resuspended in 190 μL of Bead Wash Buffer and kept on ice. A 10 mg/mL solution of EDC is prepared by adding 1 mL of ice cold Bead Conjugation Buffer directly to a 10 mg bottle of EDC (Thermo Pierce, 77149) and completely dissolving the contents. 10 μL of the 10 mg/mL EDC solution is added to the 190 μL of beads. The bead solution is immediately vortexed for approximately 10 s and is placed on a shaker for 30 min. at room temperature. The supernatant is then removed and the beads are washed 1× with 200 μL ice cold Bead Conjugation Buffer. The Bead Conjugation Buffer is aspirated and the 200 μL solution of 0.5 mg/mL buffer exchanged antibody is added. The beads are mixed and placed on a shaker for 2 h at room temperature. After the antibody conjugation, the supernatant is removed and saved for analysis. The beads are washed 2× with Bead Wash Buffer and the first wash is also saved.

Next, 1 mL of Bead Blocking Buffer (Quanterix, 100457) is added to the beads and they are placed on a shaker at room temperature for 30 min. The beads are then collected on a magnet and the supernatant is removed. The beads are washed 1× with 200 µL of Bead Conjugation buffer followed by resuspension in 200 µL of Bead Conjugation. The beads are then transferred to a clean microcentrifuge tube. The beads are counted using a Z2 Coulter Counter (Beckman Coulter) to calculate the concentration. The saved supernatant and first wash are measured using a NanoDrop 1000 to approximate the antibody coating efficiency. The coating efficiency is calculated by subtracting the values obtained for the supernatant and first wash from the initial antibody concentration, while accounting for volume.

2.7.2 Detection Antibody Biotinylation

Approximately 0.13 mg of detection antibody is buffer exchanged into Biotinylation Reaction Buffer (Quanterix, 101358) using Amicon centrifugal spin filters. The buffer exchanged antibody concentration is measured by reading the absorbance at 280 nm using a NanoDrop 1000. The buffered exchanged antibody concentration is adjusted to 1 mg/mL in Biotinylation Reaction Buffer. 100 µL of Milli-Q water is added to one 2 mg vial EZ-link NHS-PEG_4-Biotin (Thermo Scientific, 21329). The solution is pipette mixed completely and transferred to a new microcentrifuge tube. This process is repeated one time followed by the addition of 800 µL of Milli-Q water to the microcentrifuge tube. The resulting

biotin solution is 3.4 mM. The biotin to antibody ratio used for the reaction is 40:1. A total of 100 µL of 1 mg/mL detection antibody, 3 µL 3.4 mM biotin and 1.5 µL of Biotinylation Reaction Buffer are added into a 1.7 µL microcentrifuge tube. The tube is briefly vortexed and centrifuged to remove liquid from the cap. The reaction is left at room temperature without shaking for 30 min. Next, antibodies are purified using the same Amicon filters used for buffer exchange and following the same procedure. The final antibody concentration is verified using absorbance measurements at 280 on a NanoDrop 1000.

2.7.3 SiMoA Reagent Preparation and Assay Set-up

A solution of beads conjugated with capture antibodies is prepared at a concentration of 5×10^6 beads/mL in Bead Diluent (Quanterix, 101362) in a 15 mL bottle (VWR, 16067-000). The total volume of bead solution required is calculated by multiplying the number of samples per assay by 110 µL and adding 3.5 mL to account for the dead volume of the bottle. A solution of biotinylated detector antibody is prepared at an optimized concentration for the assay, typically 1 µg/mL unless otherwise stated. For 2-step and 3-step assays, the total volume of detector solution required is calculated by multiplying the number of samples per assay by 30 µL and 110 µL, respectively, and adding 3.5 mL to account for the dead volume of the bottle. A solution of SβG is made by diluting SβG concentrate (Quanterix, 100439) in SβG Diluent (Quanterix, 100375). The concentration of SβG used is optimized per assay. The total volume of SβG

required is calculated by multiplying the number of samples per assay by 110 µL and adding 3.5 mL to account for the dead volume of the bottle. Sample Diluent is prepared depending on the assay, and for a 1:4 dilution 75 µL of Sample Diluent is required for each sample being diluted plus an additional 3.5 mL for dead volume. Beads, Detector, SβG, and Sample Diluent are all loaded into the reagent bays of the HD-1 analyzer and assigned in the software. Beads must be placed in a rack capable of shaking so that no bead settling occurs during imaging (spots 1-3). RGP (Quanterix, 100030) is assigned in the appropriate bay in the HD-1 Analyzer and assigned in the software. Calibrators are made by diluting a known protein standard in 25% Newborn Calf Serum (Life Technologies, 16010-142), 0.01% Tween-20 (Sigma, P7949), 5 mM Ethylenediaminetetraacetic acid (EDTA) (Invitrogen, 15575-020), 0.15% ProClin300 (Sigma Aldrich, 48914-U), and PBS (Sigma Aldrich, P5493-1L). The volume of all calibrators required is 100 µL plus an additional 50 µL (150 µL total) for each replicate per concentration with one replicate per well. For samples that are not diluted, the same volume is required. For samples that are diluted 1:4, a total of 125 µL is required for three replicates and can be placed in one well (25 µL x 3 replicates + 50 µL dead volume). Calibrators and samples are loaded into a 96-well plate (Axygen, PCR-96-FS-C), loaded into the instrument, and properly assigned in the software.

2.7.4 General HD-1 Procedure

The following steps are performed automatically by the HD-1 Analyzer. A fixed tip pipette is used to remove 110 μL of bead solution from the assigned reagent bottle and dispenses it into a cuvette. A magnet collects the beads and the pipette is used to remove the supernatant. A disposable tip pipette is used to remove 100 μL of sample from the inserted 96-well plate and add it to the beads. If the assay is a 3step assay, the cuvette is then incubated for 15 min. If the assay is a 2-step assay, 20 μL of detector are added and then the solution is incubated for 35 min. After the first incubation, the cuvette enters a wash ring where it is washed 4 times to remove excess reagent and/or sample. In the 3-step assay, this initial wash is followed by the addition of 100 μl of detector, a 5 min incubation, and 4 more washes. In both cases, the next step is now the addition of 100 μL of SβG followed by a 5 min incubation and 4 subsequent washes. The SβG is removed and the beads are resuspended in of RGP using the disposable tip pipette. The bead solution in RGP is then loaded onto the Simoa disc (Quanterix, 100001) and Simoa oil (Quanterix, 100206) is flowed in to seal the wells. White light and fluorescence images are taken and stored as IPL files.

2.7.5 Data Analysis

The HD-1 Analyzer calculates the AEB and respective concentration in pg/mL for each sample based off of the assigned calibrators. The calibration curve

is fit using a $1/y^2$ weighted 4-PL fit. The LOD is calculated by extrapolation using 3 standard deviations from the blank measurement.

2.8 References

1. Anderson, N.L. & Anderson, N.G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Molecular & Cellular Proteomics* **1**, 845-867 (2002).
2. Berg, J.M., Tymoczko, J.L. & Stryer, L. (eds.) Biochemistry, Edn. 5th. (W.H. Freeman, New York; 200).
3. Leng, S.X. et al. ELISA and Multiplex Technologies for Cytokine Measurement in Inflammation and Aging Research. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **63**, 879-884 (2008).
4. Simpson, R.J. & Greening, D.W. Serum/plasma proteomics: methods and protocols, Vol. 728.; 728. (Springer, New York; 2011).
5. Crowther, J.R. in ELISA: Theory and Practice, Vol. 42 35-61 (Humana Press, Totowa; 1995).
6. Rich, R.R. et al. (eds.) Clinical Immunology: Principles and Practice, Edn. 4. (Elsevier Health Sciences, 2013).
7. Rissin, D.M. et al. Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nat. Biotechnol.* **28**, 595-599 (2010).
8. Rissin, D.M. et al. Simultaneous Detection of Single Molecules and Singulated Ensembles of Molecules Enables Immunoassays with Broad Dynamic Range. *Anal. Chem.* **83**, 2279-2285 (2011).
9. Quanterix Corp.*Quanterix Announces Commercial Availability of its Simoa Single Molecule Array Technology*.PAN Communications.30 July 2013.Web.4 March 2015.
10. Kan, C.W. et al. Isolation and detection of single molecules on paramagnetic beads using sequential fluid flows in microfabricated polymer array assemblies. *Lab on a Chip* **12**, 977-985 (2012).
11. Rissin, D.M. et al. Multiplexed single molecule immunoassays. *Lab on a Chip* **13**, 2902-2911 (2013).
12. Massart, D.L. Chemometrics: a textbook, Vol. 2; 2. (Elsevier, Amsterdam; New York, NY, U.S.A; New York; 1988).
13. Chang, L. et al. Single molecule enzyme-linked immunosorbent assays: Theoretical considerations. *J. Immunol. Methods* **378**, 102-115 (2012).
14. Wild, D. (ed.) The Immunoassay Handbook: Theory and applications of ligand binding, ELISA and related techniques, Edn. 4th. (Elsevier, Oxford; 2013).

Chapter 3

Ultra-Sensitive Protein Detection via Single Molecule Arrays Towards Early Stage Cancer Monitoring

3.1 Introduction

The ultimate goal in cancer diagnostics is to develop tools to detect harmful disease states as early as possible. The ability to detect and treat cancer at early stages is of paramount importance in clinical practice, because for multiple types of cancer, including prostate,[1] breast,[2] ovarian,[3] and colon,[4] it has been demonstrated that early detection leads to both improved prognosis and improved survival rates. Tumor biomarkers typically consist of protein, RNA, or DNA abnormalities produced by either cancer cells or by the body in response to the presence of cancer.[5] While cancer-screening methods continue to push the boundaries in terms of detection limits, the development of ultrasensitive blood-based detection methods for cancer and cancer biomarkers could significantly benefit cancer diagnostics by enabling earlier detection than is currently possible. Furthermore, the use of biomarkers in blood-based diagnostics has the potential for differentiating biologically relevant disease from tumors that may never become symptomatic—a dimension that image-based diagnostics lack.[6] Unlike methods that require tumor tissue, such as tissue genotyping,[7] measuring protein biomarkers in blood represents a less invasive process.

ELISAs are the most common test available for measuring protein concentrations in blood; however, due to a lack of sensitivity, ELISAs may not be able to detect clinically relevant protein biomarkers in serum at very low levels.[8] In order to detect ultra-low concentrations of the cancer biomarker PSA, this work utilized SiMoA, a recently developed ultra-sensitive ELISA based on single molecule counting technology.[9] The LOD of a leading clinical diagnostic PSA

ELISA assay (Siemens)[10] is 100 pg/mL and the LOD of commercially available ultra-sensitive ELISA-based PSA tests is between 3-10 pg/mL.[11] Results from the experiments described in this chapter are compared to one or both of these benchmark assays, depending on the range of the described data. We obtained a LOD of 0.005 pg/mL for PSA using SiMoA, which is significantly more sensitive than standard and ultra-sensitive ELISA and is comparable to previous SiMoA work.[9, 12] Other literature reports have also demonstrated sensitive PSA tests, including recent work by Liu *et al.* where they utilized gold nanoparticles to create a colorimetric ultra-sensitive assay for PSA with a LOD of 0.0031 pg/mL.[13]

Recently, several groups have produced ultra-sensitive assays for the detection of various protein biomarkers. Notable works have utilized electrochemical microfluidic arrays,[14] electrochemical immunosensors with gold nanoparticles functionalized with magnetic multi-walled carbon nanotubes,[15] and novel laser-induced fluorescence systems[16] to detect cancer biomarkers. Although these reports advanced the field of ultra-sensitive biomarker detection, many of them require complicated assay set-ups, lengthy preparation, or are subject to sensor fouling.

SiMoA has been previously implemented in studies utilizing PSA for monitoring recurrence of prostate cancer after radical prostatectomy,[12] as well as tumor necrosis factor-alpha (TNF-α) and interleukin 6 (IL-6) for monitoring therapeutic efficacy in Crohn's disease.[17] Importantly, recent work by Warren *et al.* describes the use of SiMoA to noninvasively discriminate between mice with and without thrombosis by detecting microdosed disease-tailored nanoparticles at

ultra-low levels.[18] SiMoA represents an already commercialized method that is simple, straightforward, and can be performed in a fully automated instrument in a few hours with minimal preparation making it an ideal tool for implementation in clinical settings.

We chose to use PSA as a cancer biomarker for a proof-of-concept study for early cancer detection due to the extremely sensitive SiMoA LOD and its high reproducibility. This work describes the use of a prostate cancer cell line to create a novel murine xenograft model to monitor tumor formation in mice. Since the only source of human PSA within the mouse model are the human cells that are injected and then replicated in the mouse, the concentration of PSA in the mouse's bloodstream should correspond to the relative number of cancer cells that may lead to tumor formation. We demonstrate that due to its high sensitivity, SiMoA can be used to detect the presence of nascent tumors at a much earlier stage than is possible with any other protein assay. Although this work is demonstrated with PSA, it should be applicable to any biomarker associated with tumor growth that is found in the blood.

3.2 Mouse Model Development

3.2.1 Preliminary Mouse Models

The main goal of this work was to create a mouse model for prostate cancer and to track the progression of serum PSA concentration using SiMoA technology prior to the formation of palpable tumors. In order to accomplish this

goal, cell line-based mouse xenografts were utilized. This type of mouse model is commonly used in cancer research and is employed in this study using LNCaP cells, which secrete PSA. LNCaP cells are an epithelial cell line that originated from a metastatic lesion of a human prostatic adenocarcinoma and are commonly used in prostate cancer research. The doubling time for this cell line *in vitro* is approximately 60 hours.[19]

Previous studies have demonstrated that upon inoculation of 3.0×10^6 LNCaP cells, tumors will form in male immunocompromised mice within eight weeks.[19] Building on this work, experiments were performed to determine a baseline level of PSA in the serum of mice inoculated with standard concentrations of LNCaP cells. In a preliminary study, a cohort of five female NOD/SCID mice was subcutaneously injected with LNCaP cells. Four mice received two injections for a total of 3.0×10^6 LNCaP cells while the final mouse received a single injection of 1.5×10^6 cells. The experimental scheme is shown in **Figure 3.1**. Time course bleeds were taken on a weekly basis from each mouse for a total of 77 days with the final time point being a terminal bleed.

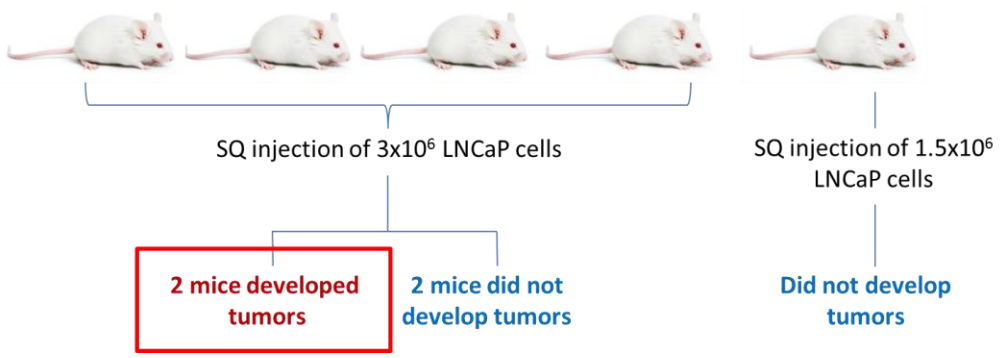
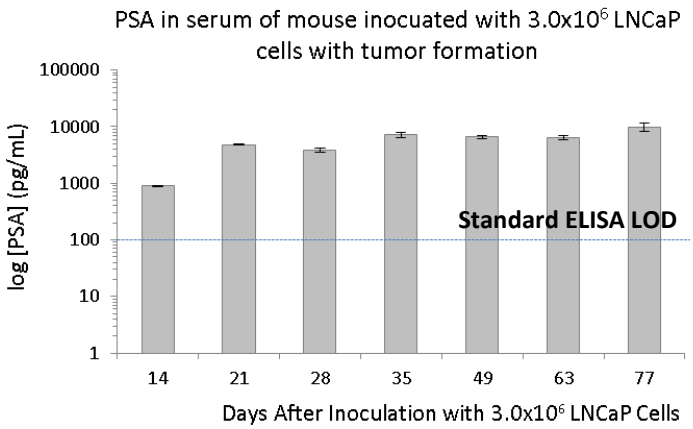


Figure 3.1 Scheme describing mouse study. Five female NOD/SCID mice were subcutaneously injected with either 3.0×10^6 or 1.5×10^6 LNCaP cells. Tumors developed in two of the mice that received a 3.0×10^6 cell injection.

The PSA content of the collected serum samples was measured using SiMoA. Of the five mice included in this study, two of the mice that were inoculated with 3.0×10^6 LNCaP cells developed large palpable tumors, while the remaining three mice did not. **Figure 3.2** shows the results of the single molecule analysis from one of the mice that developed a tumor. Since the sample volume was limited and it was unclear whether or not the concentration of PSA would be detectable within the serum, 10 fold dilutions were made. Any remaining serum was then diluted 20 fold to obtain as many replicates as possible from each sample. Dilutions were made in 5 mM EDTA/PBS containing 10 μg/mL TruBlock, which is a heterophilic blocker. As is demonstrated from the bar graph, the concentrations of PSA within these samples were significantly higher than the LOD of standard PSA ELISA assays. PSA from these samples can easily be detected using traditional ELISA, even as early as two weeks after inoculation. These results agree well with what was presented in the literature in terms of tumor growth in the expected time frame and cell inoculum; however, all of the

concentrations of PSA were too high to properly utilize SiMoA for the monitoring of tumor formation in this mouse.

a.



b.

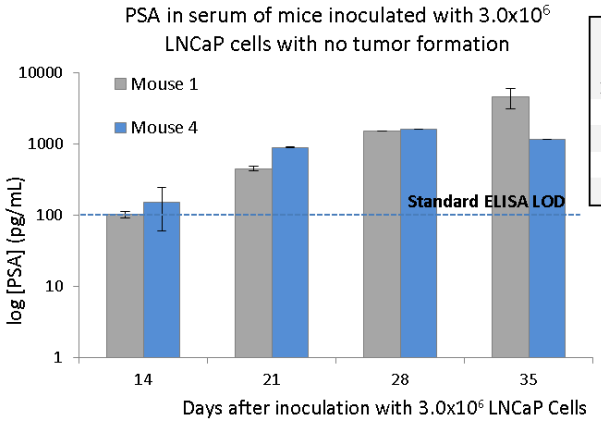
Days after inoculation	[PSA] (pg/mL)	n
14	897±20	3
21	4835±39	2
28	3918±333	5
35	7151±808	3
49	6651±453	3
63	6433±505	2
77	9747±1620	15

Figure 3.2 a) Bar graph of measured PSA concentrations in the serum of one mouse inoculated with 3.0×10^6 LNCaP cells. Measurements were taken over the course of 77 days and ultimately the mouse developed a palpable tumor. Error bars represent the standard deviation. **b)** Tabulated values for each measurement.

Interestingly, two of the mice from the same 3.0×10^6 inoculation cohort did not develop large tumors. Single molecule analysis of the PSA concentration in these samples is shown in **Figure 3.3.** Although samples were collected weekly for a total of 77 days, samples collected after day 35 had PSA concentrations that saturated the instrument and were thus not measurable and are not contained on the graph below. Unlike the mouse that developed a large tumor, the two mice, indicated as mouse 1 and mouse 4, which received the same LNCaP inoculum but did not develop tumors had significantly lower concentrations of PSA in their serum. Fourteen days after inoculation, mice without tumors had approximately a 7-fold lower PSA concentration than the corresponding mouse that developed a large tumor. Despite the decreased concentration of PSA, the overall

concentrations of PSA in the majority of samples were either at or above the LOD for standard PSA ELISA. Although these mice did not form large palpable tumors, it was noted that a few cells were stuck in the Matrigel plug used for inoculation. These few cells never formed a proper tumor, but could be the cause for the notable increase in PSA concentration over time.

a.



b.

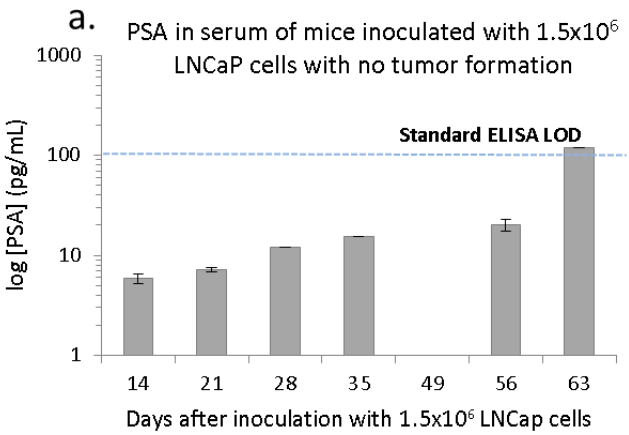
Days after inoculation	Mouse 1 [PSA] (pg/mL)	n	Mouse 4 [PSA] (pg/mL)	n
14	102±10	2	152±92	3
21	451±34	2	896±4	3
28	1520±4	4	1600±1	4
35	4590±1472	3	1150±1	2

Figure 3.3 a) Bar graph of measured PSA concentrations in the serum of two mice inoculated with 3.0×10^6 LNCaP cells. Measurements were taken over the course of 35 days and the mice did not develop tumors after 77 days. Error bars represent the standard deviation. **b)** Tabulated values for each measurement.

The final mouse in this cohort was only inoculated with $1.5\text{x}10^6$ LNCaP cells, which was half of the concentration that the literature stated would form tumors in immunocompromised mice within eight weeks. This mouse did not form tumors, but similarly to two of the mice inoculated with 3.0×10^6 cells described above, it was observed that cells were stuck in the Matrigel plug used for injection. Single molecule analysis of the PSA concentration in these samples is shown in **Figure 3.4**. As shown in the graph, the concentration of PSA steadily increased over time and only surpasses the LOD of standard ELISA after day 63.

The LOD for the PSA assay for this run was 0.06pg/mL and all values were well above the SiMoA threshold. Despite the lack tumor formation, these results demonstrate that PSA can be measured in mouse serum with only the presence of a few cells and it can also be measured using SiMoA over time for potential tumor growth monitoring.

a.



b.

Days after inoculation	[PSA] (pg/mL)	n
14	5.88±0.65	3
21	7.19±0.33	2
28	12.2	1
35	15.4	1
56	20.2±2.6	2
63	119±0.05	2

Figure 3.4 a) Bar graph of measured PSA concentrations in the serum of one mouse inoculated with $1.5.0\times10^6$ LNCaP cells. Measurements were taken over the course of 63 days and did not ultimately lead to tumor formation after 77 days. Error bars represent the standard deviation. **b)** Tabulated values for each measurement.

Due to the method used for obtaining weekly bleeds, only a small volume of whole blood is obtained from each sample and thus an even smaller volume of serum (~30-50 μL) is available for analysis. For this reason, it is not always possible to obtain replicate measurements, such as the case for days 28 and 35, despite dilution. The sample from day 49 was tested and discarded as an outlier as determined via q-test. This sample had a red color, indicating it had undergone hemolysis. Hemolysis likely occurred during sample collection. The presence of

burst red blood cells was likely the cause of the false positive signal. The sample for day 77 was also hemolyzed and was not measurable using SiMoA.

The experiments described above demonstrate that LNCaP cells can be used in a mouse xenograft model to measure increasing PSA concentrations over time with lower than typical cell inoculums. One oversight in the design of the above experiments was that the mice used were female and the cell line used is derived from male prostate tissue. Since the concentration of cells used in these studies should have all produced tumors, it is likely that the reason some of the mice did not produce tumors is linked to their sex. In order to confirm that results using male NOD/SCID mice would be similar, two mice were inoculated with $4x10^6$ LNCaP cells and terminal bleeds were collected after four weeks. At end stage, tumors from these mice measured about 10 mm in diameter. Serum PSA concentrations were extremely high, measuring approximately 30,000 and 78,000 pg/mL for each mouse (**Figure 3.5**). As expected, these values were significantly higher than the LOD of standard ELISA PSA assays and were also greater than the PSA concentrations from the initial study.

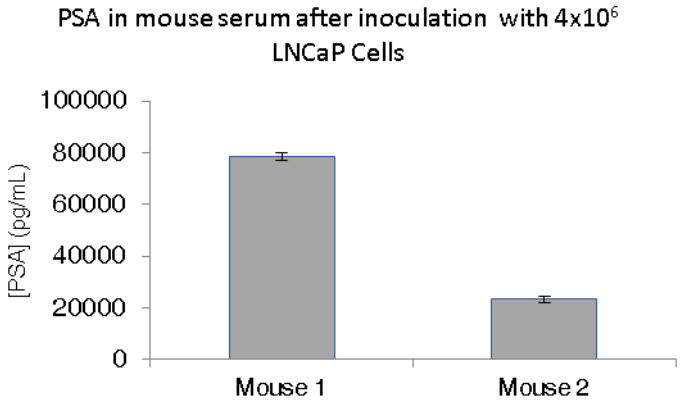


Figure 3.5 Bar graph of measured PSA concentrations in the serum of mice inoculated with 4×10^6 LNCaP cells. Measurements were taken four weeks after inoculation. Both measurements were well above the detection limits of standard ELISA and beyond the scope of SiMoA. Error bars represent the standard deviation of triplicate measurements.

All further mouse models described in this text using LNCaP cells were performed with male NOD/SCID mice. Despite the original oversight, meaningful information was gained regarding the ability to track PSA prior to potential tumor formation. Since the male mice that were tested and formed tumors also had extremely high concentrations of PSA within their serum, another study was designed to determine how soon an inoculum of just 1×0 $x10^6$ cells could be detected in mouse serum. In this study, 19 male NOD/SCID mice were inoculated with $1x10^6$ LNCaP cells subcutaneously. In order to obtain more sample volume per mouse and thus more replicates, terminal bleeds were used for this study instead of the previously described weekly bleeds. Terminal bleeds were taken daily for 19 days, beginning the day after inoculation, thus the data presented here are from three different mice at each time point. Although several measurements were below the detection limit, PSA was detected in the serum of one mouse after

only three days (**Figure 3.6**). A general trend of increasing PSA concentration over the 19 days was observed and all values were below the detection limit of standard ELISA. Ultra-sensitive ELISA was able to detect PSA in samples from days 12 and 16-19. None of the mice in this study developed tumors; however, the study only spanned three weeks. Follow-up experiments (*vide infra*) using luciferase tagged LNCaP cells with the same cell concentration for inoculation suggest that it is extremely likely these mice would have developed tumors within a few more weeks. These results further demonstrate the sensitivity of SiMoA for detecting low-level serum-based biomarkers within cancer models.

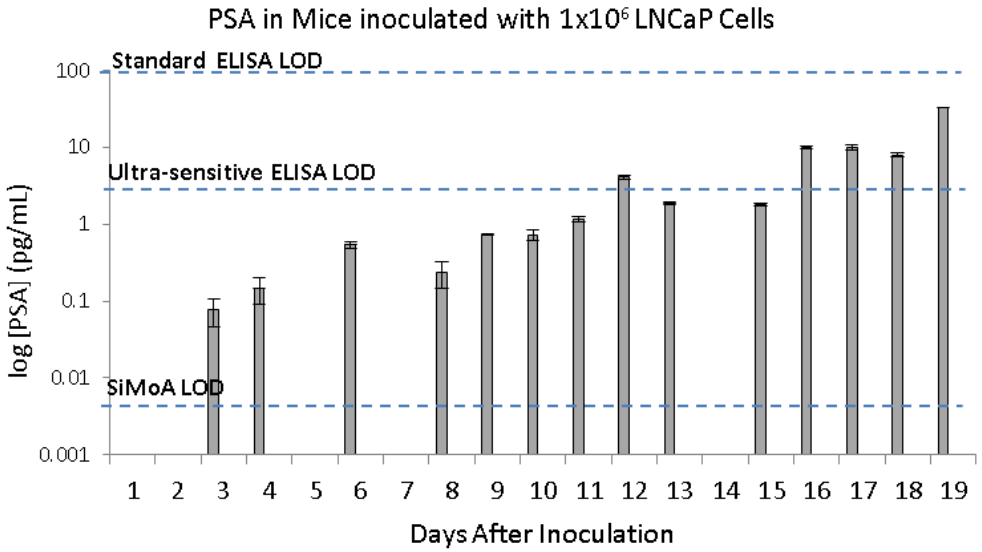


Figure 3.6 Bar graph showing the log of PSA concentrations in the serum of mice inoculated with $1x10^6$ LNCaP cells over 19 days. All measurements were taken using serum from terminal bleeds of individual mice where one mouse was sacrificed for each time point. A general increase in PSA concentration is displayed over time from days 1-19. The assay SiMoA LOD was 0.005 pg/mL. For comparison, the LOD of ultrasensitive (3 pg/mL) and standard ELISA (100 pg/mL) are shown. Each sample was measured in triplicate. Error bars represent the standard deviation between triplicate measurements for each individual sample.

As previously mentioned, a luciferase tagged LNCaP cell line (luc-LNCaP) was constructed. The primary reason for using this cell line was to determine that the only source of PSA within the mouse serum was from the primary tumor and not from distant metastases. A total of three male NOD/SCID mice were inoculated with 1×10^6 LNCaP cells and observed for several weeks. An additional three mice were inoculated with Matrigel as a control. After 8 weeks, two of the mice inoculated with the luc-LNCaP cells developed tumors that were approximately 1 cm in diameter. This experiment demonstrates that the mice in the previous study very likely did not have sufficient time to develop tumors and would have if the experiment had been conducted over a longer period of time. One of the mice inoculated with the luc-LNCaP cells did not develop any tumors and none of the control mice developed tumors. Bioluminescence imaging was used to visualize the presence of the tumors within the mice as well as to determine whether or not there were any metastatic regions (**Figure 3.7**). No metastases were present in either of the two mice that had tumors, indicating that the only source of PSA within the mouse serum is from the resected tumors.

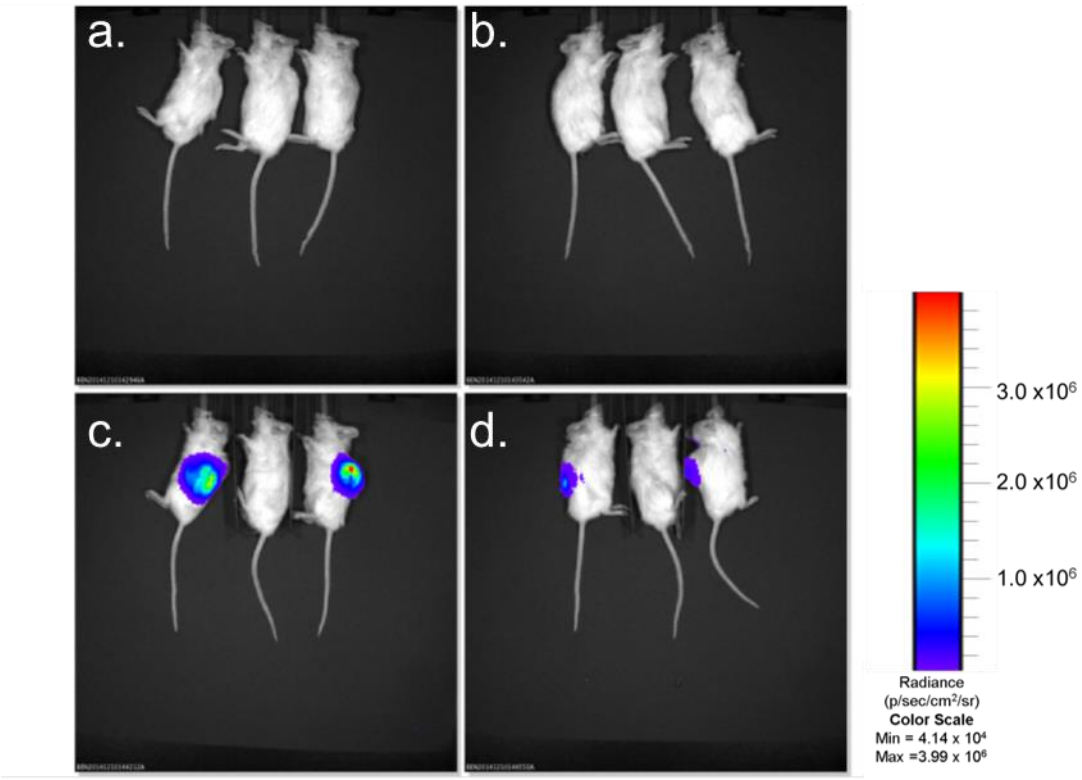


Figure 3.7 Bioluminescence images of three NOD/SCID mice inoculated with $1x10^6$ luc-LNCaP cells. Also shown are three control mice injected with Matrigel and cell growth media after 8 weeks. The control mice showed no signs of tumor formation in either the left (**a**) or right (**b**) views; however, the mice injected with the luc-LNCaP cells had 1 cm tumors that were clearly visible from both the left (**c**) and right (**d**) views of the animal. No evidence of metastasis is present among any of the three mice. The middle mouse did not grow a tumor.

3.2.2 Development of PSA/LNCaP Mouse Model with Low Cell Inoculation

Since the goal of this work was to monitor PSA levels in serum over time prior to tumor formation, preferably over multiple weeks, it was vital that the inoculation concentration of cancer cells was not too high. An inoculation of 1×10^6 LNCaP cells, as previously described, can still be considered relatively high when factoring the overall volume of blood in an average mouse (approximately 2 mL). Thus, experiments were conducted to monitor tumor

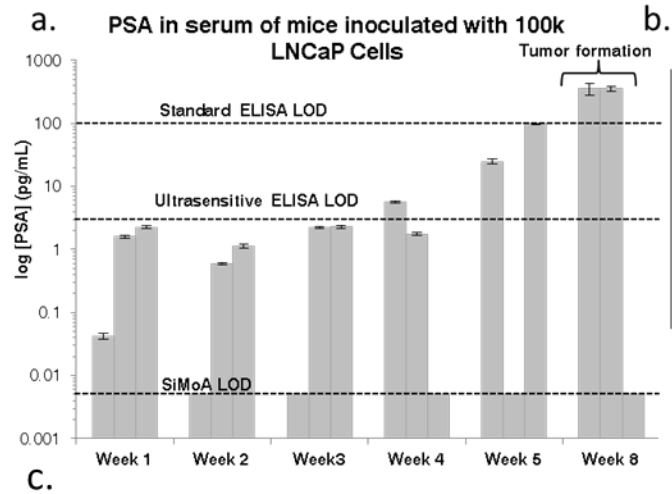
progression in mice using significantly reduced cell concentrations of 100,000 (100k) and 10,000 (10k) cells. To my knowledge, this is the first example of a mouse xenograft model using such a low concentration of LNCaP cells.

3.2.2a Inoculation with 100k Cells

Eighteen NOD/SCID mice were subcutaneously injected with 100,000 LNCaP cells. Terminal bleeds were performed on three mice per week for a total of five weeks. The remaining three mice were sacrificed when palpable tumors formed (week 8). Replicates were used for each time point due to restrictions on the volume of whole blood that can be drawn from each mouse per week as well as the dead volume of the assay plates used for analysis, as described in the Materials and Methods section. **Figure 3.8a-b** illustrates the increase in PSA concentration in the serum of mice over time, where each time point represents a terminal bleed from an individual mouse. **Figure 3.8c** demonstrates that the increase in PSA over time is exponential. Serum from three mice was measured for each time point and triplicate measurements were made for each sample.

Serum from six healthy male mice was tested and all samples had undetectable levels of PSA (not shown). This result is expected since mice do not express human PSA. Despite decreasing the number of cells used to inoculate the mice, the concentration of PSA in the serum of the majority of mice sacrificed after 1 to 5 weeks was measurable at values well above the SiMoA detection limit. All values measured with SiMoA prior to tumor formation were at or below

the detection limit of commercially available ELISA kits. All measurements taken from weeks 1 to 3 were only measurable using SiMoA. Ultrasensitive ELISA was able to measure one sample in week 4 and both of the samples that were measurable by SiMoA from week 5. Two out of three mice developed large tumors (>8 mm in diameter) after eight weeks and had significantly elevated serum PSA levels (355±2 pg/mL) compared to the rest of the mice. Notably, the mouse that did not develop a tumor after eight weeks also had undetectable levels of PSA. The mice with palpable tumors were the only samples with serum PSA concentrations that would have been easily detected using standard ELISA.



b.

PSA in Serum of Mice Inoculated with 100k LNCaP Cells

Average Mouse Serum [PSA] (pg/mL)			
	Sample 1	Sample 2	Sample 3
Week 1	0.04 ± 0.01	1.59 ± 0.08	2.25 ± 0.11
Week 2	Below LOD	0.59 ± 0.02	1.12 ± 0.08
Week 3	Below LOD	2.21 ± 0.06	2.27± 0.10
Week 4	5.67 ± 0.14	1.75 ± 0.08	Below LOD
Week 5	24.9 ± 2.0	Below LOD	98.2 ± 2.1
Week 8	354 ± 70	357 ± 30	Below LOD

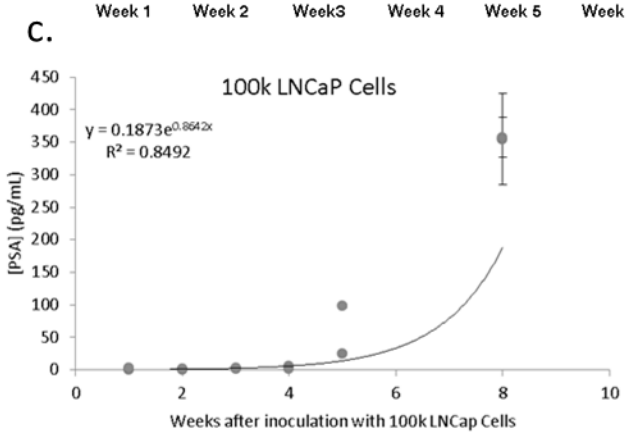


Figure 3.8 PSA in serum of mice inoculated with 100k LNCaP cells. **a)** Bar graph of measured PSA concentrations in the serum of mice inoculated with 100k LNCaP cells on a log scale. All measurements were taken using serum from

terminal bleeds of individual mice. Each sample was measured in triplicate. Error bars represent the standard deviation between triplicate measurements for each individual sample. An exponential increase in PSA concentration is observed over time. Three mice were examined at each time point, but several samples were below the 0.005 pg/mL detection limit for the assay. All samples with values below the LOD are plotted on the LOD line. LODs for standard ELISA and ultra-sensitive ELISA are shown for comparison at 100 pg/mL and 3 pg/mL. **b)** Tabulated values for each measurement. **c)** Scatter plots showing the exponential increase of PSA over time.

3.2.2b Inoculation of 10k Cells

In order to assess the sensitivity limits of SiMoA with our mouse model, the cell inoculate was further reduced to 10,000 LNCaP cells—*over 100 times lower than the typical dose used for inducing tumors.* A similar experimental approach was used, where 18 NOD/SCID mice were subcutaneously injected with 10,000 LNCaP cells and weekly terminal bleeds were performed. Small tumors (>3 mm in diameter) were present at week 8 in all three mice that were not sacrificed earlier; however, one mouse developed thymic lymphoma and it was not possible to obtain a serum sample (week 8, sample 2). **Figure 3.9a-b** depicts the increase in PSA concentration over time, with a large increase at week eight (average of 82.4±4.7 pg/mL PSA), where palpable tumor formation occurred. As demonstrated in **Figure 3.9c**, the increase in PSA over time is exponential. Several mice exhibited PSA concentrations below the SiMoA LOD, but measurements for at least two mice were recorded for each time point in triplicate. All measured PSA concentrations, including those with tumors, were below the detection limit of standard ELISA, as depicted on the graph, and only the samples containing tumors were detectable using ultra-sensitive ELISA.

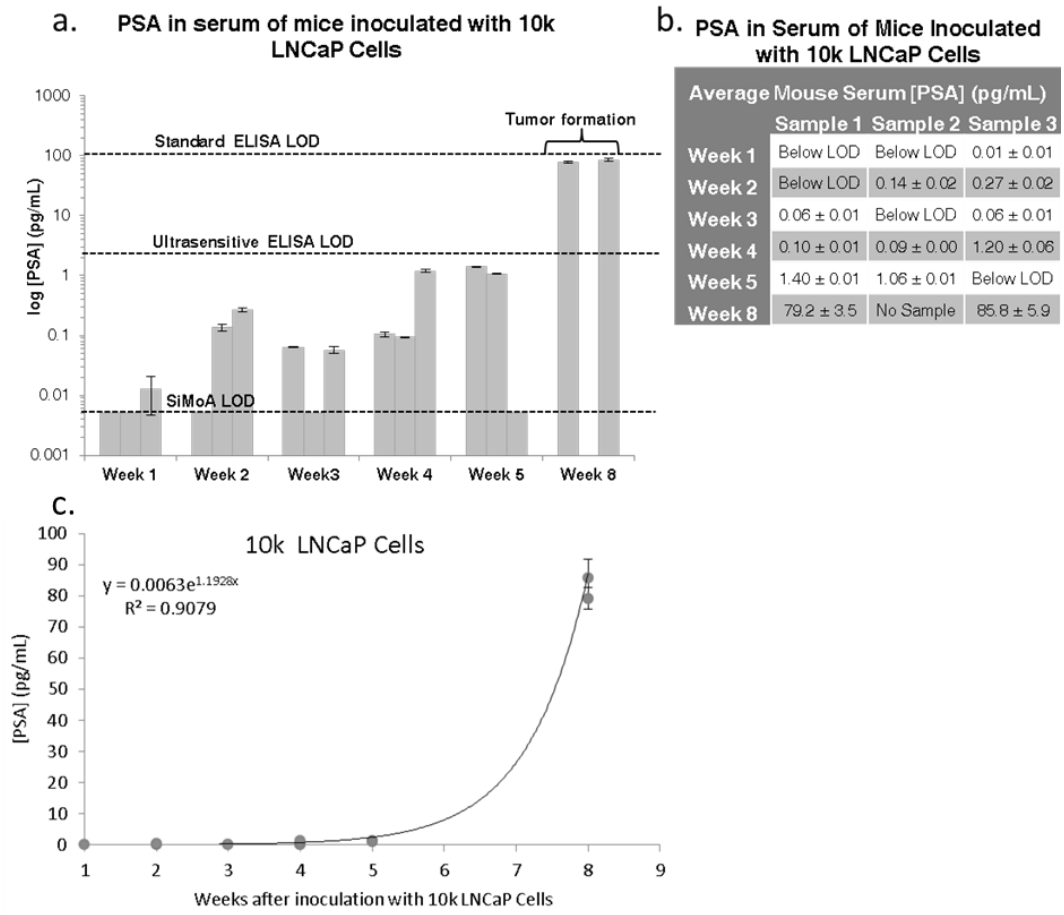


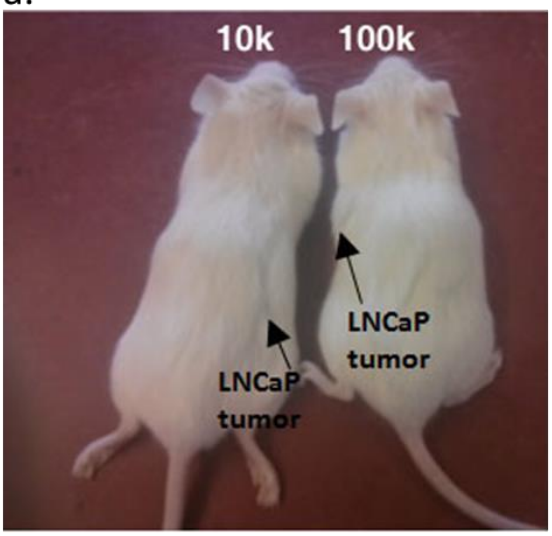
Figure 3.9 PSA in serum of mice inoculated with 10k LNCaP cells. **a)** Bar graph of measured PSA concentrations in the serum of mice inoculated with 10k LNCaP cells on a log scale. All measurements were taken using serum from terminal bleeds of individual mice. Each sample was measured in triplicate. Error bars represent the standard deviation between triplicate measurements for each individual sample. An exponential increase in PSA concentration is observed over time. Three mice were examined at each time point, but several samples were below the 0.005 pg/mL detection limit of the assay. LODs for standard ELISA and ultra-sensitive ELISA are shown for comparison at 100 pg/mL and 3 pg/mL. **b)** Tabulated values for each measurement. **c)** Scatter plots showing the exponential increase of PSA over time.

3.2.2c Tumor Characterization

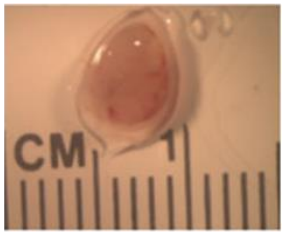
As previously mentioned, mice from both the 10k and 100k cell cohorts developed tumors within eight weeks of inoculation. Images highlighting the

location of tumors *in vivo* are shown in **Figure 3.10a**. The tumors were removed for characterization and images of representative resected tumors are shown in **Figure 3.10b-c**. Tables including tumor diameter, weight, volume, and serum PSA concentration are shown in **Figure 3.10d-e**. The average PSA levels for each cohort were very similar, with an average of 82.4±4.7 pg/mL and 355±2 pg/mL for the 10k and 100k cohorts measured at week 8, respectively.

a.



b.



c.

d. **Mice Inoculated with 10k LNCaP Cells**

Mouse	Tumor Diameter (mm)	Tumor Weight (g)	Tumor Volume (mm^3)	Serum [PSA] (pg/mL)
1	3.5	0.07	22.4	79.2±3.5
2	1	0.03	0.5	NA
3	3.9	0.05	31	85.8±5.9

e. **Mice Inoculated with 100k LNCaP Cells**

Mouse	Tumor Diameter (mm)	Tumor Weight (g)	Tumor Volume (mm^3)	Serum [PSA] (pg/mL)
1	8.2	0.28	288.5	354±70
2	9.2	0.27	407.5	357±30
3	NA	NA	NA	Below LOD

Figure 3.10 Characterization of tumors from mice inoculated with 100k and 10k LNCaP cells. **a)** Representative photographs of mice inoculated with 10k and

100k LNCaP cells. Photographs were taken after 8 weeks with tumor locations indicated by arrows. **b-c)** Photographs showing representative tumors removed from mice inoculated with 10k and 100k LNCaP cells respectively, after 8 weeks at 8x magnification. **d-e)** Tables describing data pertaining to the tumors found in the 10k and 100k mouse cohorts, respectively, as well as the concentrations of PSA within the serum.

The tumor volume, however, was more variable between the samples. For example, the tumor volume in mice 1 and 2 from the 100k cohort were 288.5 and 407.5 mm^3, respectively, despite having very similar concentrations of PSA. The discrepancy between tumor volume and the amount of PSA produced from the tumor can possibly be attributed to necrosis within the tumor, or dead tumor cells. Necrotic cells would be expected to produce lower amounts of PSA. Although it is difficult to characterize exactly how much necrosis was present in each tumor, **Figure 3.11** shows a representative photo from a 100k tumor with H&E staining. Another possible explanation is restricted diffusion from the center of a tumor. Such restriction would prevent PSA from accessing the vasculature and would reduce the amount detectable in the blood.

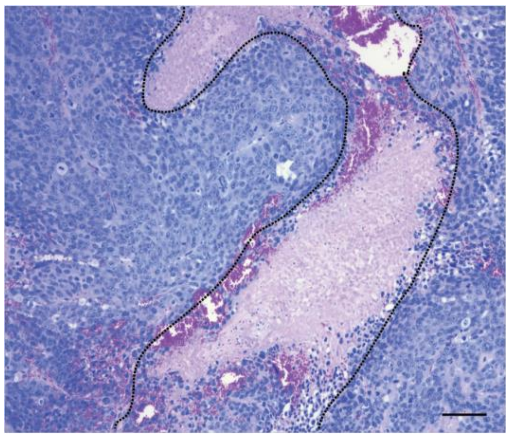


Figure 3.11 Representative photo of H&E stained tumor from mice inoculated with 100k LNCaP cells. The necrotic area is highlighted via the dotted line. Image

taken at 100× magnification. Scale bar is 100 μm. The presence and extent of necrosis in the tumors from each mouse may lead to varying PSA concentrations in the respective serum.

Finally, all tumors were preserved in formalin after removal. Tumor sections were stained with both hematoxylin and eosin (H&E) as well as for PSA (**Figure 3.12**). The PSA stain confirmed that the tumor cells were PSA positive.

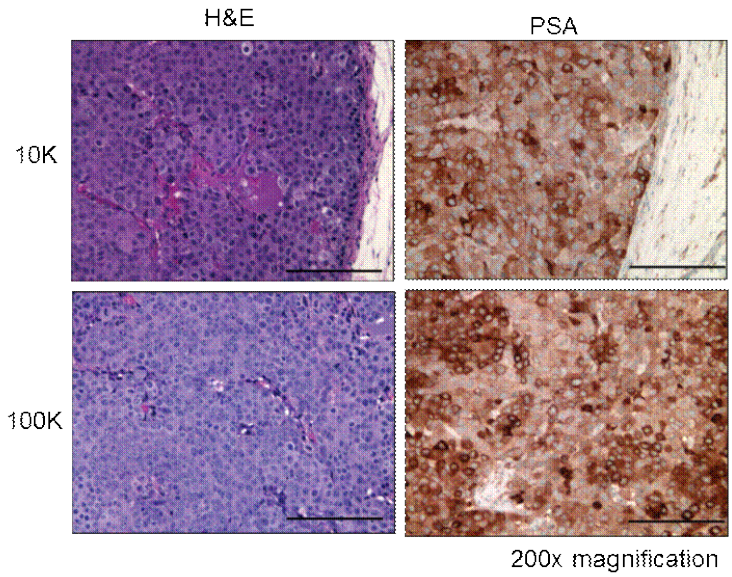


Figure 3.12 H&E and PSA staining of tumor samples from mice inoculated with both 10k and 100k LNCaP cells. PSA staining confirms the presence of PSA in the cells of the resected tumors. Scale bars are100 μm.

3.3 Discussion

SiMoA technology has the potential to revolutionize cancer diagnostics and therapeutics by non-invasively detecting cancer biomarkers in serum earlier than current methods. This chapter describes a proof-of-concept study where prostate cancer was induced in a mouse model at very low cell inoculums to

demonstrate the utility of SiMoA as an early cancer detection tool. Increasing levels of PSA were measured in the serum of mice as a sign of tumor progression, since the only source of increasing levels of secreted human PSA in the mouse model is the proliferation of the PSA-secreting cancer cells. Several mice developed tumors, indicating that SiMoA can be used to monitor ultra-low levels of biomarkers in serum prior to the formation of palpable tumors.

The initial mouse model utilizing female mice illustrates the ability of SiMoA to track PSA increasing within mouse serum at both high and low PSA concentrations. LNCaP cells are androgen sensitive and while the rate of tumor growth is independent of hormone status, tumor development has been shown to directly correlate to the androgen levels present within the host's serum.[20] For this reason, male mice develop LNCaP tumors more rapidly and at a higher frequency than female mice, which explains the lack of tumor growth in half of the cohort receiving the same 3.0×10^6 cell inoculation in the first study.

The female mice that received inoculations of 3.0×10^6 and 1.5×10^6 cells but did not develop tumors still had measurable increases in PSA within their serum, which means that eventually a tumor might have developed if they were allowed to sit for a longer period of time. However, the inoculation concentration was so high that SiMoA would not be necessary for the measurements if a tumor did develop since these samples were already above the LOD of standard ELISA. Despite the high concentrations of PSA in these mice, the experiment did provide valuable insight for the design of future experiments. For example, after only two weeks, the serum from the female mouse that was inoculated with 1.5×10^6 cells

had a PSA concentration of 5.88±0.65 pg/mL which is 17 times less than the standard ELISA LOD of 100 pg/mL, but is still almost 100 times greater than the SiMoA LOD calculated for that day (0.06 pg/mL). This feat was accomplished while only using half of the inoculum of cells recommended to cause tumors to form after a period of eight weeks. Even though this mouse did not form a tumor, the ability to track PSA within this mouse at concentrations below the LOD of traditional ELISA was an important stepping stone towards the overall goal.

It was also established that, using male NOD/SCID mice, SiMoA can be used to measure PSA in mice as soon as 3 days after inoculation with 1×10^6 LNCaP cells. Despite the fact that tumors also did not grow in this model, a general increase in PSA concentration was measured over time for 19 days while still maintaining concentrations below the LOD of standard ELISA. The volume of serum used in this and the previous time course study was extremely limited, approximately 30-50 μL of serum per sample. This required dilutions to 5-10% for replicate measurements and would have simply been impossible for standard ELISA, which typically requires volumes upwards of 100 μL. SiMoA therefore demonstrated its ability to use low volumes of precious samples and still make accurate and sensitive measurements beyond the capabilities of traditional methods. In addition, studies using luc-LNCaP cells and bioluminescence imaging demonstrated that the tumors grown in the mice did not metastasize after long periods of time.

In the case of mice inoculated with 100k LNCaP cells, the PSA concentration after one week ranged from 0.04 to 2.25 pg/mL, levels much lower

than the detection capabilities of conventional ELISA. In comparison, mice inoculated with only 10k LNCaP cells resulted in even lower PSA concentrations, which were at or below the SiMoA LOD after week one. However, by week two, PSA measurements for the 10k cohort surpassed the SiMoA LOD with obtained values of 0.20±0.09 pg/mL, while those in the 100k cohort yielded values of 1.71±0.37 pg/mL after two weeks. The first samples with PSA concentrations high enough to be measured by conventional ELISA were collected eight weeks after inoculation; once palpable tumors formed. Although different mice were used for each measurement, the increasing trends of PSA concentration over the short time courses of these experiments within the majority of samples show that it is possible to monitor tumor markers during the initial stages of tumor formation in low volumes of serum using SiMoA. This sensitivity should enable the ability to detect the earliest stages of cancer, before palpable or visible tumors are detectable. Due to the limited number of cells used for inoculation, PSA was not detectable in all mouse replicates for either the 10k or the 100k model. Injecting mice with lower cell inoculums leads to issues regarding clonal heterogeneity, where the likelihood of having cancer stem cells or clones that can form tumors is reduced. The presence of low numbers of cancer stem cells is also the likely reason why only two of the 100k mice developed tumors. This process has been described previously using LNCaP cells in the work of Wan and coworkers.[21] In addition, all mice are biologically unique and variations are expected between individuals.

The PSA concentrations that were detected in mice after tumor formation seemed to reasonably correlate with the number of cells used for inoculation in each mouse cohort. However, the PSA concentrations do not scale as may be expected (i.e. the concentration of PSA from mice with 100k tumors was not 10x higher than the 10k tumors). This result may be due to the fact that the tumors were encapsulated, thus hindering the secretion of PSA from the tumors into the bloodstream. Also, the presence of varying amounts of necrotic cells within the tumors could also result in different secretion rates of PSA. A larger sample size will be needed to further investigate this correlation.

The data presented here also demonstrate that, by utilizing SiMoA, significantly fewer LNCaP cells are required for PSA to be detected in serum. This methodology can be particularly useful when examining the growth of tumors in mice over time and at lower cell concentrations, where high cell concentrations were previously required. The technique could thus be utilized for models to study early tumor development, either for primary tumors or for relapse after treatment in more complex models. In addition, biomarker levels can potentially be monitored after chemotherapy to ascertain the efficacy of treatment. This work represents a proof-of-concept study using PSA; however, this work can easily be extended to use any other protein biomarker or cell line of interest, which would further advance the field of cancer diagnostics and early detection.

Through the creation of a low cell inoculum mouse model, the work outlined in this chapter successfully demonstrated that SiMoA can be used to measure the biomarker PSA within the serum of mice that ultimately developed

palpable tumors. The sensitive detection of circulating protein biomarkers can not only enable earlier detection of disease, but can also unmask the body's unique underlying chemistry and offer more dynamic information that imaging techniques are incapable of providing. This work shows significant promise for the use of SiMoA in the field of oncology and early cancer detection as a non-invasive approach to early tumor detection.

3.4 Materials and Methods

The human prostate cancer cell line LNCaP (ATCC CRL-1740) was provided by Dr. Charlotte Kuperwasser at Tufts Medical School. Cells were grown in RPMI 1640 media (Gibco) containing 10% fetal bovine serum and 1% antibiotic/antimycotic (Gibco). LNCaP cells were grown at 37°C with 5% CO_2 and were passaged for less than 3 weeks. The cells tested negative for mycoplasma (MilliPROBE; Millipore) and were authenticated by ATCC/Promega.

For injections, LNCaP cells were trypsinized (0.05%; Gibco) and counted using trypan blue to identify viable cells. Cells were resuspended in 50% v/v in Matrigel (BD Biosciences) and RPMI culture media for injections.

For initial time course mouse studies, eight week old NOD/SCID (Jackson Laboratories) mice were anesthetized with isofluorane, and 30ul of resuspended LNCaP cells were subcutaneously injected in a 1:3 ratio of matrigel to media. Approximately 100ul of blood was drawn from the submandibular vein and collected into SST Microtainer tubes (Becton Dixon and Co, Franklin Lakes

NJ). At end stage, mice were anesthetized isofluorane and blood was drawn via cardiac puncture followed by euthanasia by $C0_2$ asphyxiation. For 100k and 10k mouse studies, eight week old male NOD/SCID were subcutaneously injected with 100 μL of resuspended cells. Mice were given food and water *ad libitum*. Whole blood was collected from mice under terminal anesthesia via cardiac puncture and transferred to SST Microtainer tubes. The care of animals and all animal procedures were conducted in accordance with a protocol approved by the Tufts University Institutional Animal Care and Use Committee (IACUC).

Whole blood samples were allowed to clot on ice for 10 min. followed by centrifugation at 1,500 × g for 10 min at 4°C. Serum was removed and immediately snap frozen. All serum samples were stored at -80°C prior to use. In initial studies using female and male mice with inoculations of 3.0, 1.5, and 4.0×10^6 cells, samples were diluted to final concentrations of either 5 or 10% serum in 5 mM EDTA/PBS with 10 g/mL TruBlock. For SiMoA analysis, serum samples were diluted to a final concentration of 25% serum in PSA Diluent (Quanterix Corp.) before being loaded onto an automated HD-1 analyzer (Quanterix Corp.). Serum from healthy male mice was used as control and was also diluted to 25% in PSA diluent and treated similarly. Tumor tissue was preserved in formalin and sections were stained with hematoxylin and eosin (H&E) and prostate specific antigen (PSA; Tufts Histology Core).

Due to restrictions imposed by the Tufts University IACUC, the maximum volume of whole blood that can be drawn from the NOD/SCID mice used is 7.5% of the animal's total circulating blood, which is equivalent to ~82-105 μL per

week. From this volume of whole blood, the maximum volume of serum obtained is at most 50 μL, which is equivalent to the dead volume of the plates used for these experiments. Due to these limitations, multiple mice were sacrificed per time point in order to obtain biological replicates and still obtain valuable information.

Luc-LNCaP cells were constructed as follows. Lentiviral particles were generated by co-transfection of the pLenti-PGKV5LucNeo construct (Addgene) with pCMV-VSVG, expressing the vesicular stomatitis virus glycoprotein and the packaging construct pCMVΔR8.2Δvpr into 293T cells with FuGENE 6 transfection reagent (Promega). Lentivirus-containing supernatant from the transfected 293T cells was filtered through a 0.45μm syringe filter and used to directly infect subconfluent LNCaP cells in the presence of 5μg/mL protamine sulfate (Sigma). LNCaP-luc cells with lentiviral integration were selected with 750 μg/mL hygromycin.

To generate tumors, $1x10^6$ LNCaP-luc cells were injected subcutaneously in 100 μl of a 1:1 mixture of Matrigel and cell growth media into male 8 week old NOD/SCID mice. Control mice were injected with 100 μl of a 1:1 mixture of Matrigel and cell growth media only. When the tumors reached a diameter of 1 cm (after 8 weeks), all mice received a 100 μl intraperitoneal injection of 15 mg/mL luciferin. Five minutes after treatment, luminescence was quantified using an IVIS 200 Imager (Perkin Elmer). Bioluminescence was analyzed using Living Image software (Caliper Life Sciences).

SiMoA PSA Assay

All single molecule measurements were taken using a HD-1 Analyzer (Quanterix Corp.). All HD-1 consumables, including wash buffers, sample diluent, assay discs, 96-well plates, sealing oil, cuvettes, and PSA reagents, were purchased from Quanterix Corp. The SiMoA platform has been described previously in the literature and in Chapter 2. [9, 22] The PSA SiMoA assay consists of three-steps: 1) 15 minute target incubation with capture beads 2) 5 minute incubation with detection antibody, and 3) 5 minute SBG incubation.

3.5 References

1. Etzioni, R. et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. *Cancer Causes Control* **19**, 175-181 (2008).
2. Berry, D.A. et al. Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer. *N. Engl. J. Med.* **353**, 1784-1792 (2005).
3. Shapira, I. et al. Circulating biomarkers for detection of ovarian cancer and predicting cancer outcomes. *Br J Cancer* **110**, 976-983 (2014).
4. Edwards, B.K. et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* **116**, 544-573 (2010).
5. Paoletti, C. & Hayes, D.F. Molecular Testing in Breast Cancer. *Annu. Rev. Med.* **65**, 95-110 (2014).
6. Nolen, B.M. et al. Prediagnostic Serum Biomarkers as Early Detection Tools for Pancreatic Cancer in a Large Prospective Cohort Study. *PLoS ONE* **9**, e94928 (2014).
7. Paik, S. et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *N. Engl. J. Med.* **351**, 2817-2826 (2004).
8. Anderson, N.L. & Anderson, N.G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Molecular & Cellular Proteomics* **1**, 845-867 (2002).
9. Rissin, D.M. et al. Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nat. Biotechnol.* **28**, 595-599 (2010).
10. Prostate Cancer: More Choices in PSA Tests.Siemens AG.September 2, 2014.http://www.healthcare.siemens.com/clinical-specialities/oncology/laboratory-diagnostics-in-oncology/prostate-cancer/laboratorians/
11. Ferguson, R.A., Yu, H., Kalyvas, M., Zammit, S. & Diamandis, E.P. Ultrasensitive detection of prostate-specific antigen by a time-resolved immunofluorometric assay and the Immulite immunochemiluminescent third-generation assay: potential applications in prostate and breast cancers. *Clin. Chem.* **42**, 675-684 (1996).
12. Wilson, D.H. et al. Fifth-Generation Digital Immunoassay for Prostate-Specific Antigen by Single Molecule Array Technology. *Clin. Chem.* **57**, 1712-1721 (2011).
13. Liu, D. et al. Glucose Oxidase-Catalyzed Growth of Gold Nanoparticles Enables Quantitative Detection of Attomolar Cancer Biomarkers. *Anal. Chem.* **86**, 5800-5806 (2014).
14. Malhotra, R. et al. Ultrasensitive Detection of Cancer Biomarkers in the Clinic by Use of a Nanostructured Microfluidic Array. *Anal. Chem.* **84**, 6249-6255 (2012).
15. Li, F. et al. An ultrasensitive sandwich-type electrochemical immunosensor based on signal amplification strategy of gold nanoparticles functionalized magnetic multi-walled carbon nanotubes loaded with lead ions. *Biosensors and Bioelectronics* **68**, 626-632 (2015).
16. Zhang, J. et al. Novel Signal-Enhancing Immunoassay for Ultrasensitive Biomarker Detection Based on Laser-Induced Fluorescence. *Anal. Chem.* **87**, 2959-2965 (2015).

17. Song, L. et al. Single molecule measurements of tumor necrosis factor α and interleukin-6 in the plasma of patients with Crohn's disease. *J. Immunol. Methods* **372**, 177-186 (2011).
18. Warren, A.D. et al. Disease Detection by Ultrasensitive Quantification of Microdosed Synthetic Urinary Biomarkers. *J. Am. Chem. Soc.* **136**, 13709-13714 (2014).
19. Horoszewicz, J.S. et al. LNCaP Model of Human Prostatic Carcinoma. *Cancer Res.* **43**, 1809-1818 (1983).
20. Horoszewicz, J.S. et al. LNCaP Model of Human Prostatic Carcinoma. *Cancer Res.* **43**, 1809-1818 (1983).
21. Wan, X.S., Zhou, Z., Steele, V., Kopelovich, L. & Kennedy, A.R. Establishment and characterization of sublines of LNCaP human prostate cancer cells. *Oncology Reports* **10**, 1569-1575 (2003).
22. Kan, C.W. et al. Isolation and detection of single molecules on paramagnetic beads using sequential fluid flows in microfabricated polymer array assemblies. *Lab on a Chip* **12**, 977-985 (2012).

Chapter 4

The Development of Single Molecule Protein Assays for Early Breast Cancer Detection

4.1 Introduction

It has been demonstrated that the early detection of breast cancer is correlated to both improved patient survival and improved prognosis.[1-3] Mammography is not an ideal method for breast cancer screening and early detection because it is not specific enough to distinguish between the benign and malignant masses it detects and it still misses approximately 20% of breast cancers.[4, 5] In addition, the process for determining whether or not a tumor found via mammography screening often requires a follow-up biopsy to determine malignancy, which is invasive. Testing protein biomarkers in serum offers a significantly less invasive method for early detection screening and monitoring; however, the only biomarkers recommended for analysis in blood for clinical use in breast cancer patients are CA 15-3, CA 27.29, and CEA.[6] Two of the most important features of a sensitive biomarker are that it must be both specific and selective for cancer, ideally for a specific type of tumor. Unfortunately, due to an inherent lack of specificity and sensitivity associated with these biomarkers, they are useful primarily for monitoring only the later stages of disease and not for detection of primary disease or relapse.[7-10]

This chapter describes the use of highly sensitive SiMoA technology to develop a panel of three breast cancer biomarkers that can be measured in serum at extremely low concentrations. Our hypothesis is that breast cancers at the earliest stages consist of tumors or small groups of cancerous cells that are not detectable by mammography or other methodologies, release specific proteins into the bloodstream that would only be detectable by an ultra-sensitive method.[11]

The design of SiMoA assays for such protein biomarkers may enable the detection of early stage cancer. By studying multiple biomarkers in a panel, a signature can be extrapolated from the resulting responses, which can be correlated to the patient's disease state. The ability to measure the concentration of breast cancer biomarkers within serum at levels below the concentrations detectable by traditional ELISA may allow for both earlier detection and disease relapse monitoring in a non-invasive manner, reducing or eliminating the reliance on mammography for screening. In addition, since more information regarding the chemistry and biology of the cancer can be obtained from a protein signature, this methodology may also improve prediction of patient outcome as well as therapeutic efficacy by gaining the ability to monitor a patient's response to treatment. Through biomarker selection, assay development, and the testing of both healthy and breast cancer patient serum, this preliminary work demonstrates the significant potential for SiMoA in the field of early breast cancer diagnostics.

This work specifically focuses on the development of three biomarkers: ERα, PR, and cyclin-dependent kinase inhibitor 2D (CDKN2D). This chapter explains the role of these biomarkers in breast cancer and describes the development of their respective SiMoA assays. First, each marker is assessed individually and comparisons are made in both healthy and breast cancer samples. Next, the development of a supervised classification model is described that uses the response of all three biomarkers as well as patient age as input variables for the accurate prediction of early stage cancer from healthy donors.

4.2 Biomarker Selection

The first goal of this project was to select protein biomarkers that are specific for breast cancer and are secreted into the blood stream so that a simple blood test can be implemented. Other criteria used were that the expression level of the biomarker changes in the early stages of cancer, the biomarker is not detectable in high levels in healthy serum, and that antibodies are readily available to create the SiMoA assay. A list of approximately 25 different biomarkers was established through consultation with experts in the field. Although some biomarkers may be indicative of multiple different types of cancer, the use of a biomarker panel should provide sufficient specificity.[12] From the list of 25 established biomarkers for this study, this work describes the successful development and implementation of three biomarkers: ERα, PR, which were recommended by Dr. Marsha Moses (Harvard Medical School/Boston Children's Hospital), and CDKN2D, which was suggested by Dr. Akhilesh Pandey (Johns Hopkins University School of Medicine). The importance of these three biomarkers in the body as well as in breast cancer is described below.

4.2.1 Estrogen Receptor-*α*

Estrogens are steroid hormones that are vital to many biological processes in the human body.[10,] 17β-estradiol (E_2) is the most abundant estrogen and is vital for the regulation, development, differentiation, and maintenance of male and female sexual and reproductive functions.[13-15] E_2 also plays important regulatory roles in the cardiovascular, musculoskeletal, immune, and central nervous

systems.[13-15] ERα is a nuclear receptor (NR) that is partially responsible for mediation of the biological functions of estrogens and estrogen-like molecules.[16] NRs regulate gene transcription by either directly binding to specific DNA regulatory sequences or by interacting with co-activator and/or co-repressor proteins to regulate the activity of the RNA polymerase complex.[17] It has been demonstrated that both E_2 and ERα play a role in the development of breast cancer.[18] One hypothesis behind their role in breast cancer is that upon binding of E_2 to ERα, the proliferation of mammary cells is stimulated. This process thus increases the number of target cells within the tissue. Increases in both cell division and DNA synthesis can lead to a higher risk of replication errors and possibly mutations.[16]

4.2.2 Progesterone Receptor

Also a steroid hormone, progesterone plays both a central and diverse role in various functions related to female reproduction, including ovulation, neurobehavioral expression associated with sexual responsiveness, and the development of the uterus and mammary glands.[17, 19] PRs are NRs that interact with progesterone to mediate its physiological effects. Studies have shown that progesterone promotes the growth and progression of mammary tumors and may be associated with proliferation, apoptosis, and differentiation.[19] Studies have demonstrated that in the mammary glands of healthy individuals, cells that express PRs are segregated from proliferating cells; however, this separation is

lost in breast tumor cells. It is believed that this change in PR expression may contribute to the abnormal growth of breast cancer cells.[17]

4.2.3 Cyclin-dependent kinase inhibitor 2D

CDKN2D, which is also known as p19-INK4D, is an inhibitor of cyclin-dependent kinases (INK4). Members of the INK4 family bind to cyclin dependent kinases 4 and 6 (CDK4/6) and inhibit their kinase activity by effectively blocking CDK4/6 from interacting with cyclin D.[20] This process stops cells from entering the S phase of the cell cycle, which is when DNA is replicated. CDKN2D also plays a vital role in inducing G1 phase cell cycle arrest during cellular stress as well as in the DNA damage response pathway.[20] In addition to their roles in cell cycle regulation, INK4 proteins have also been associated with apopotosis, DNA repair, senescence, and multistep oncogenesis.[21] In this regard, CDKN2D has been correlated with the proliferation of several different cancers, including breast cancer.[22]

4.3 Assay Development

SiMoA assays were developed for ERα, PR, and CDKN2D by first biotinylating and coupling several antibodies for each biomarker. Pre-biotinylated detection antibodies were also tested. Different combinations of the capture and detector reagents were screened by performing calibrations with high concentrations of target on a microplate reader. Once an antibody pair was established using this method, the assays were transferred to the HD-1 analyzer

and were optimized by testing different assay step procedures (e.g. 3 step vs. 2 step), trying multiple concentrations of detector and SβG, and by modifying buffers used in the protocol. The final assay conditions for each biomarker are listed in the Material and Methods section below. Representative calibration curves for each biomarker are in shown in **Figure 4.1**.

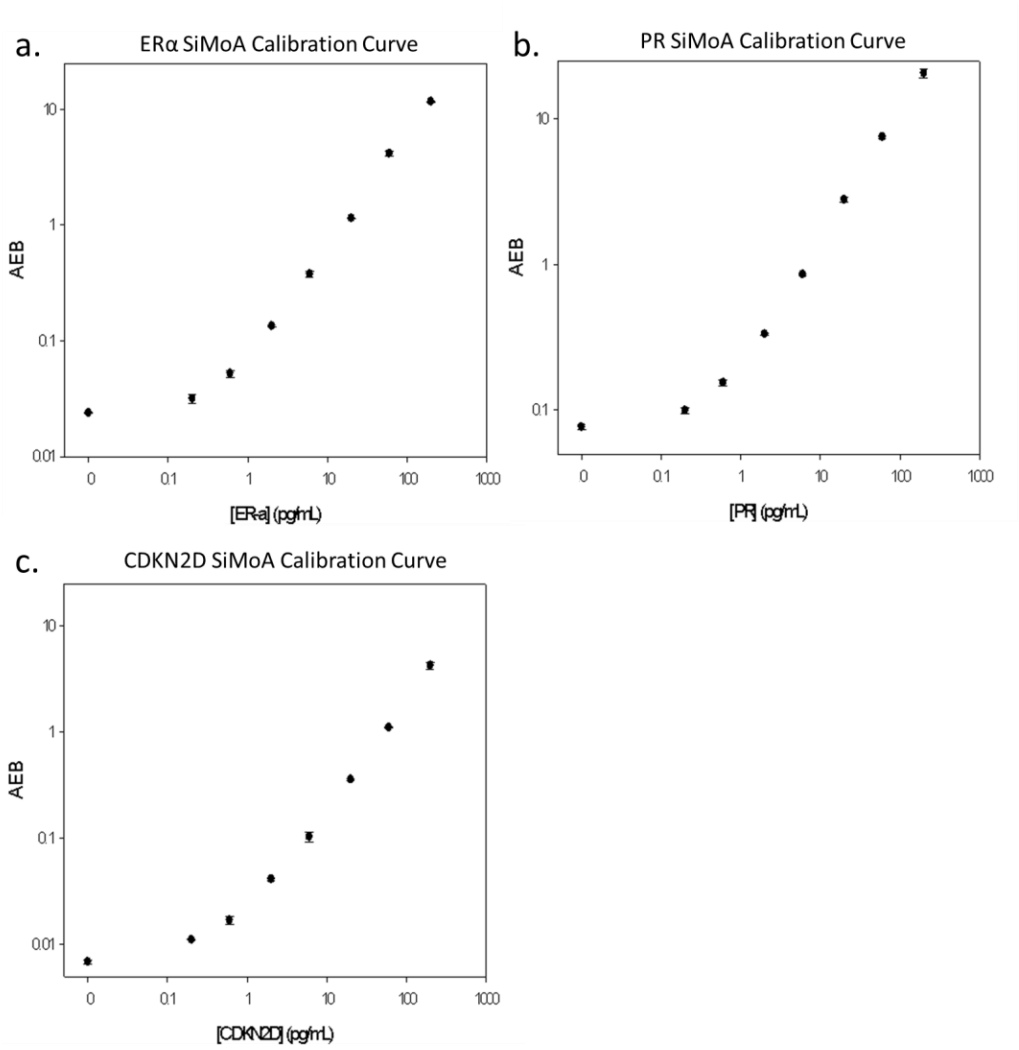


Figure 4.1 Representative SiMoA calibration curves for ERα, PR, and CDKN2D. LODs for these assays are 0.18, 0.22 and 0.12 pg/mL, respectively.

The resulting SiMoA LODs for each biomarker, the lowest available commercial LODs, and the relative fold improvement are presented in **Table 4.1**.

The LOD for the only commercial sandwich ELISA kit for human CDKN2D that was found is reported in **Table 4.1**; however, this kit is currently discontinued by the vendor (AntibodiesONLINE, ABIN812390) and thus it is possible that there are in fact no commercially available kits that can be purchased for this biomarker. Literature searches also did not find any sandwich ELISA data for this biomarker; therefor the information for the discontinued kit is used for comparisons within this chapter. The reported commercial ELISA kits for ERα and PR were both offered from R&D Systems (DYC5715-5 and DYC5415-5, respectively). In comparison to the best available commercial kits, SiMoA for ERα, PR, and CDKN2D was 56, 87, and 75 times more sensitive.

Table 4.1 Summary of SiMoA LODs compared to commercially available ELISA kits. Values are converted to molarity using the molecular weight of the appropriate standard. *The commercial kit for CDKN2D is no longer available and thus SiMoA may be the only ELISA assay for this biomarker.

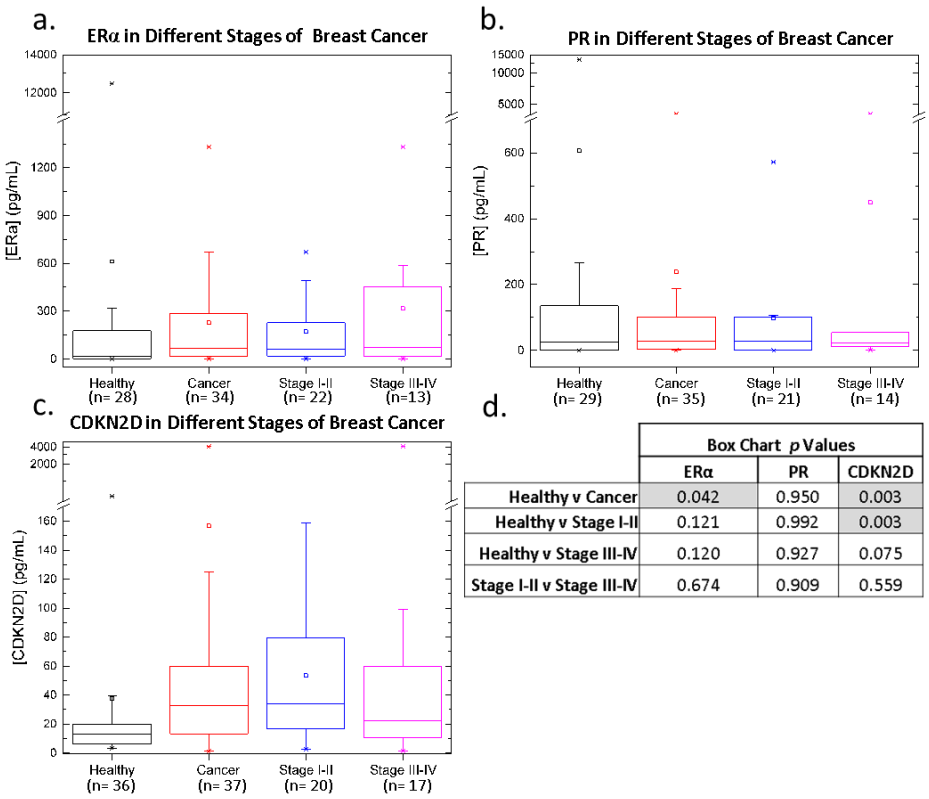
Biomarker	**Commercial ELISA LOD**		**SiMoA LOD**		**Fold Improvement (based on pg/mL)**
	pg/mL	*pM*	*pg/mL*	*fM*	
PR	15.6	0.16	0.18	1.82	87x
ERα	12.5	0.18	0.22	3.32	56x
CDKN2D	9.76*	0.55	0.12	7.42	75x*

4.4 Individual Biomarker Testing in Healthy and Patient Serum

Once the SiMoA assays were developed and optimized, serum samples from both healthy donors and breast cancer patients were tested for each biomarker. A total of 54 breast cancer and 36 healthy samples were purchased

from BioreclamationIVT and used for analysis. Information regarding patient age, cancer stage, and molecular subtype for all samples used within this chapter can be found in the Appendix (**Table A1**).

Due to sample volume constraints, not all of the samples were tested for each biomarker. The responses of each biomarker were first compared to see if there was a difference in the responses between healthy samples and all breast cancer samples. In addition, the breast cancer samples were separated by stage, such that earlier stage samples (I-II) could be compared to later stage samples (III-IV). **Figure 4.2** shows box plots with this analysis for ERα, PR, and CDKN2D. In these plots, all points that gave a response are represented and the number of samples analyzed is indicated. In order to assess the data, statistical comparisons were made using a two-tailed Mann-Whitney U test (**Figure 4.2d**).[23] From this analysis, tests using ERα and CDKN2D were able to differentiate between the groups containing all healthy and all breast cancer samples, with p values of 0.042 and 0.003, respectively, where $p < 0.05$ is significant. In addition, CDKN2D was also able to differentiate between healthy and stage I-II samples, with a p value of 0.003. None of the biomarkers were able to differentiate between stage I-II and stage III-IV breast cancer or between healthy and stage III-IV breast cancer. PR was not able to differentiate between any of the groups. From this analysis, ERα and CDKN2D are the most useful biomarkers in this panel.

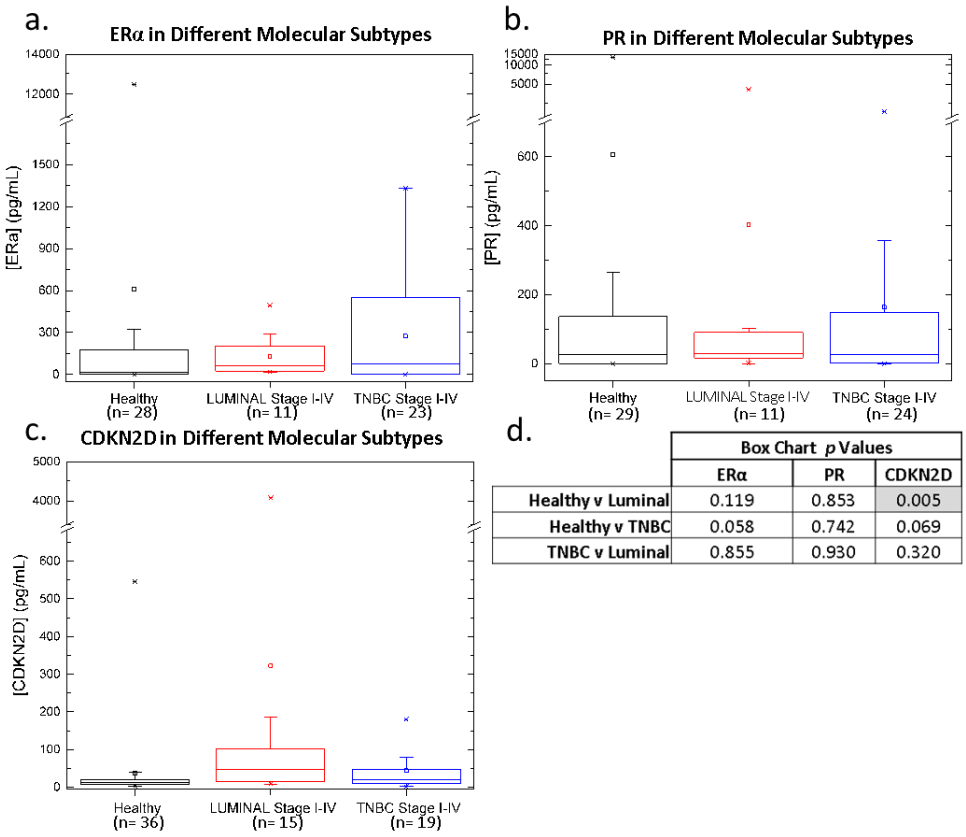


	Box Chart *p* Values		
	ERα	PR	CDKN2D
Healthy v Cancer	0.042	0.950	0.003
Healthy v Stage I-II	0.121	0.992	0.003
Healthy v Stage III-IV	0.120	0.927	0.075
Stage I-II v Stage III-IV	0.674	0.909	0.559

Figure 4.2 Box plots showing the response of **a)** ERα, **b)** PR, and **c)** CDKN2D in healthy and different stage breast cancer serum samples. Breast cancer samples are divided into groups of stage I-II and stage III-IV cancer. Sample numbers are indicated under each plot. Y scale axis is logarithmic after the break. **d)** Calculated *p* values for all data via two-tailed Mann-Whitney U test. Values in grey are statistically different from each other.

Next, each biomarker was assessed for its ability to differentiate between breast cancer samples of different molecular subtypes. Using the same sample data but organized by subtype, **Figure 4.3** shows box plots comparing the responses of ERα, PR, and CDKN2D in healthy, luminal breast cancer, and TNBC, serum. This analysis was performed similarly to that in Figure 4.2 where all data points were included in the plots and were then assessed via a two-tailed Mann-Whitney U test (**Figure 4.3d**). From the plots and the *p* values, it seems that only CDKN2D is able to differentiate between healthy and luminal samples,

with a *p* value of 0.005. CDKN2D and ERα are moderate at differentiating between healthy and TNBC, with *p* values of 0.069 and 0.058, but they are not statistically relevant. Similarly to the classification of stages, PR performed poorly on all accounts and it seems as though ERα and CDKN2D provide the most discriminatory power individually.



Box Chart *p* Values	ERα	PR	CDKN2D
Healthy v Luminal	0.119	0.853	0.005
Healthy v TNBC	0.058	0.742	0.069
TNBC v Luminal	0.855	0.930	0.320

Figure 4.3 Box plots showing the response of **a)** ERα, **b)** PR, and **c)** CDKN2D in healthy and different molecular subtype breast cancer samples. Sample numbers are indicated under each plot. Y scale axis is logarithmic after the break. **d)** Calculated *p* values for all data via two-tailed Mann-Whitney U test. Values in grey are statistically different from each other.

Additionally, once all three biomarkers were assessed individually, they were studied simultaneously to determine if the response fingerprint of all three

biomarkers would result in a more useful assay capable of distinguishing between healthy and cancer patients. In order to accomplish this task, more complex statistical analysis is required to compare the response patterns. The next section contains a short introduction and description of the multivariate techniques utilized in this chapter to accomplish this task.

4.5 Multivariate Analysis

Numerous multivariate statistical methods and algorithms exist for the purpose of processing and interpreting complex data sets. Statistical methods can be categorized as either unsupervised or supervised learning methods. Unsupervised methods do not include a response variable, that is, they do not know the identity of each assigned sample.[24] In contrast, supervised statistical methods have known response variables, enabling cross validating the acquired results. Once a supervised model is calibrated with datasets with known response variables, unknown samples can be tested using the model for prediction.

A commonly used supervised method is partial least squares-discriminant analysis (PLS-DA). Partial least squares (PLS) is a supervised multivariate regression method for the analysis of collinear data.[25] PLS extracts underlying structures within datasets, known as latent variables (LVs), that maximize the covariance between the variables (e.g. biomarker response, age) (X) and a response variable (Y). In PLS-DA, Y is categorical and expresses the class membership of the statistical units. In general, PLS-DA is used to sharpen the

separation between groups of observations. To achieve this goal, LVs are rotated to establish maximum separation among classes.[26]

PLS-DA has been used extensively in metabolomics studies, where the existence of complex datasets containing analytes with varying correlation criteria make data analysis challenging using other multivariate analysis techniques.[27] PLS-DA has also been used in studies for the determination of useful biomarkers and diagnostic factors for breast cancer,[28-30] colorectal cancer[31, 32], glioma disease,[33] and pancreatic cancer.[34] In the following work, PLS-DA models are calculated using the responses of the three biomarkers as well as patient age as inputs to discriminate between healthy and breast cancer serum samples.

4.6 Calculation of PLS-DA Models for Predicting Breast Cancer

When creating a multivariate method that utilizes multiple inputs, it is ideal to have information from each input for every sample in order to create a robust model. As previously mentioned, due to sample volume constraints, not all samples were tested using each biomarker and thus some data points were missing in the panel. One plan to limit the number of samples with missing data was to test multiple samples at the same time. In order to achieve this task, dye encoded beads were coupled to capture antibodies to attempt multiplex assays. Unfortunately, the calibration curves produced using these beads had higher background and therefore increased detection limits. Multiplex beads were not available when the majority of initial samples were tested for ERα and PR, so for

consistency and to achieve the highest sensitivity possible, all samples were tested using a ‘single-plex’ method.

In order to build PLS-DA models, data imputation was necessary to estimate the missing values. Data imputation uses the existing dataset to estimate missing values by assigning a value that statistically resembles the existing data. This process is achieved by assembling statistical models (e.g. Principal Component analysis) for each proposed “new value” iteratively until the final model containing the “new value” resembles the original model with the missing value. Data imputation was utilized as implemented in the PLS_toolbox 7.9.4.

Although data imputation was necessary due to sample availability constraints in the described work, future work will include more data points so that data imputation will not be necessary. In addition, all samples that yielded responses below the assay LODs were assigned a value of half the reported LOD of the corresponding assay. This was done so that the influence of each tested sample could be considered in the model without significantly biasing the results. A total of four inputs were used to create the PLS-DA model: age, [ERα], [PR], and [CDKN2D]. Age was used because as discussed in Chapter 1, breast cancer risk increases with age. Age is also an ideal variable since it is easily obtained information that can be added into the model for all samples. Models were created without using age as an input and the overall accuracy was slightly lower (See Appendix **Figure A1**). In order to avoid calibration bias, each model was cross-validated to have a better estimate of the classification error. Cross-validation is a common technique in statistical learning and consists on dividing the dataset into

several equal portions.[26] These portions are then used to split the dataset between a calibration set and a testing, or validation, set. In the present work, an n-5 split was used, which results in 80% of the data being used for calibration and 20% for validation. This process is repeated several times until the entire dataset has been used for validation. Initial PLS-DA models showed seven samples with high Q residuals and T^2 values over the 95% confidence limits, which indicates that these samples are outliers from the entire dataset.[26] These samples were excluded from the remaining analysis.

The PLS-DA model was first calibrated to discriminate between healthy and all available breast cancer samples. This model, (Model 1), was calibrated using data from a total of 31 healthy donors ranging in age from 22-73 and 44 breast cancer patients ranging in age from 36-84. The breast cancer samples were a mixture of both luminal and TNBC subtypes and ranged from stages I-IV. **Figure 4.4** shows the resulting receiver operating characteristic (ROC) curves from Model 1. ROC curves are commonly used for describing the utility of a diagnostic test. ROC curves plot the sensitivity, or true positive rate, against 1-specificity, or false positive rate, of a test. The area under the curve (AUC) is a measure of the sensitivity of the test, where 1 is a perfect test and 0.5 is a failed test.[35] Cross validation (CV) was performed as described in the Materials and Methods section. The ROC curves for healthy and breast cancer serum samples both had AUC values of 0.90 with similar CV AUC values of 0.88, which indicates that Model 1 is sensitive. The model was able to correctly identify 90% of the healthy samples and 75% of the breast cancer samples. The overall

accuracy of a PLS-DA model can be calculated by dividing the number of correctly assigned samples over the total number of samples in the model. For Model 1, the overall accuracy was 81%. The precision of a model can also be described. This value relates to the fraction of correctly assigned samples in a class, e.g. healthy, divided by the total number of samples that were assigned to that class in the model. The precision of Model 1 was 72% and 92% for healthy and cancer, respectively. From the above results, it appears that Model 1 is capable of distinguishing healthy serum samples from breast cancer samples of any stage from both luminal and TNBC subtypes.

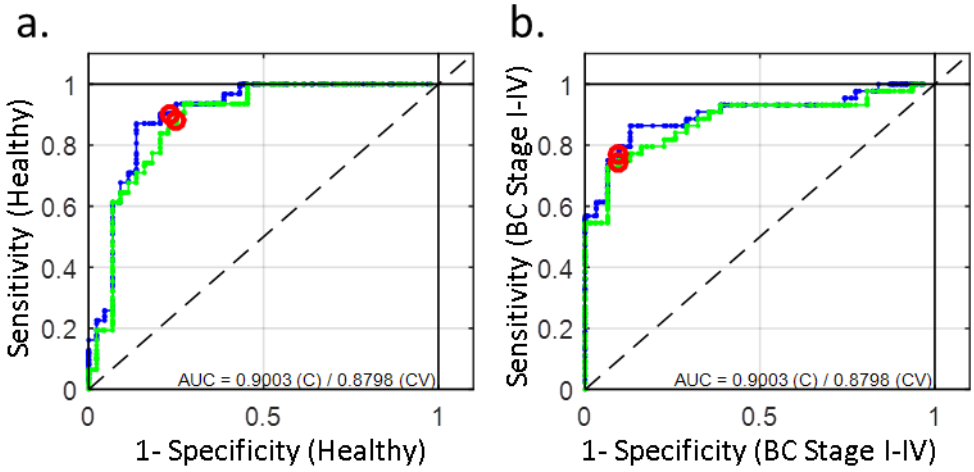


Figure 4.4 ROC curves from PLS-DA analysis of ERα, PR, and CDKN2D in **a)** healthy and **b)** stage I-IV breast cancer serum. Estimated values from calibration are shown in blue and cross validated values in green. The red circles indicate the threshold values (standard and cross-validated) determined by the PLS-DA model on the calibration data.

Next, a second PLS-DA model (Model 2) was calibrated using data collected from both healthy and early stage breast cancer serum samples to see if early stage cancer could be accurately and precisely predicted. For this calibration, data from the same cohort of healthy samples used in Method 1 were

used as well as data from 31 breast cancer serum samples that were reported as either stage I or stage II. The breast cancer samples were a mixture of both luminal and TNBC subtypes and patient ages ranged from 36-84. The resulting ROC curves from this analysis are shown in **Figure 4.5**. The AUC for both healthy and breast cancer samples stages I-II was 0.91 with a CV AUC of 0.89. The true positive rates for healthy and stage I-II breast cancer for this model were 87% and 80%, respectively. The precision of Model 2 was 86% for healthy samples and 80% for breast cancer. The overall accuracy of Model 2 was 84%; slightly better than Model 1.

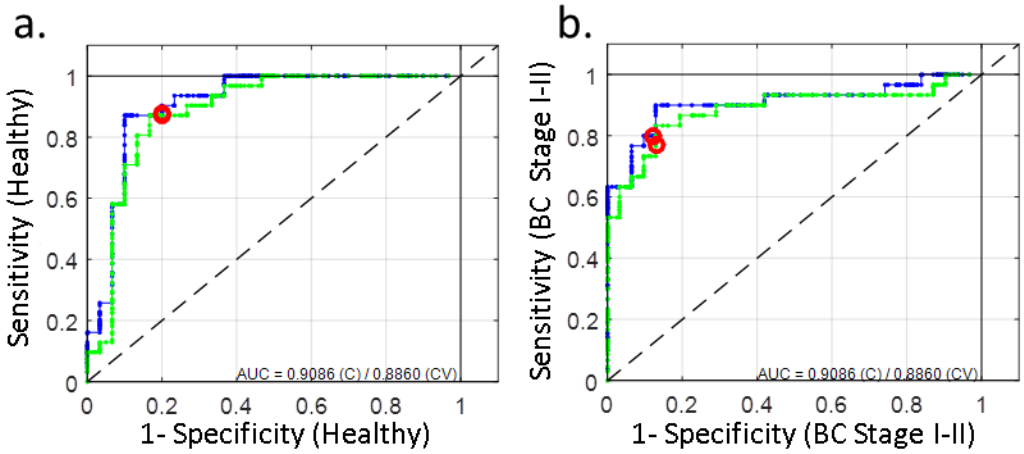


Figure 4.5 ROC curves from PLS-DA analysis of ERα, PR, and CDKN2D in **a)** healthy and **b)** stage I-II breast cancer serum. **c)** Tabulated statistics for Model 2. Estimated values from calibration are shown in blue and cross validated values in green. The red circles indicate the threshold values (standard and cross-validated) determined by the PLS-DA model on the calibration data.

Next, a model was developed to discriminate between healthy serum, early stage, and later stage breast cancer. To calibrate this model (Model 3), data from the same healthy cohort was used along with the data from the same group of

stage I-II samples. In addition, data from 14 stage III-IV breast cancer samples was used. The resulting ROC curves from this analysis are shown in **Figure 4.6**. The AUC values for healthy, stage I-II, and stage III-V were 0.89 (CV = 0.88), 0.78 (CV = 0.76) and 0.69 (CV = 0.63), respectively. Compared to Model 2, the true positive rate for predicting healthy samples and stage I-II samples decreased from 87% to 81% and 80% to 70%, respectively. The precision also decreased for stage I-II samples, falling from 86% to 62% compared to Model 2, but increased by 4% for healthy patients to 86%. Model 3 proved to be both imprecise and inaccurate at predicting stage III-IV cancers, with values of 0% for both. Of the 14 samples in this cohort, ten were assigned as stage I-II and four were assigned as healthy. Due to the inability to correctly assign later stage samples, the overall accuracy of Model 3 was 45%. Model 3 again reiterates that PLS-DA using the given variables of age and [ERα], [PR], and [CDKN2D] in serum is capable of predicting breast cancer, especially earlier stage breast cancer, but may not be able to distinguish between earlier and later stage cancers. Larger sample sizes are required to further assess this model.

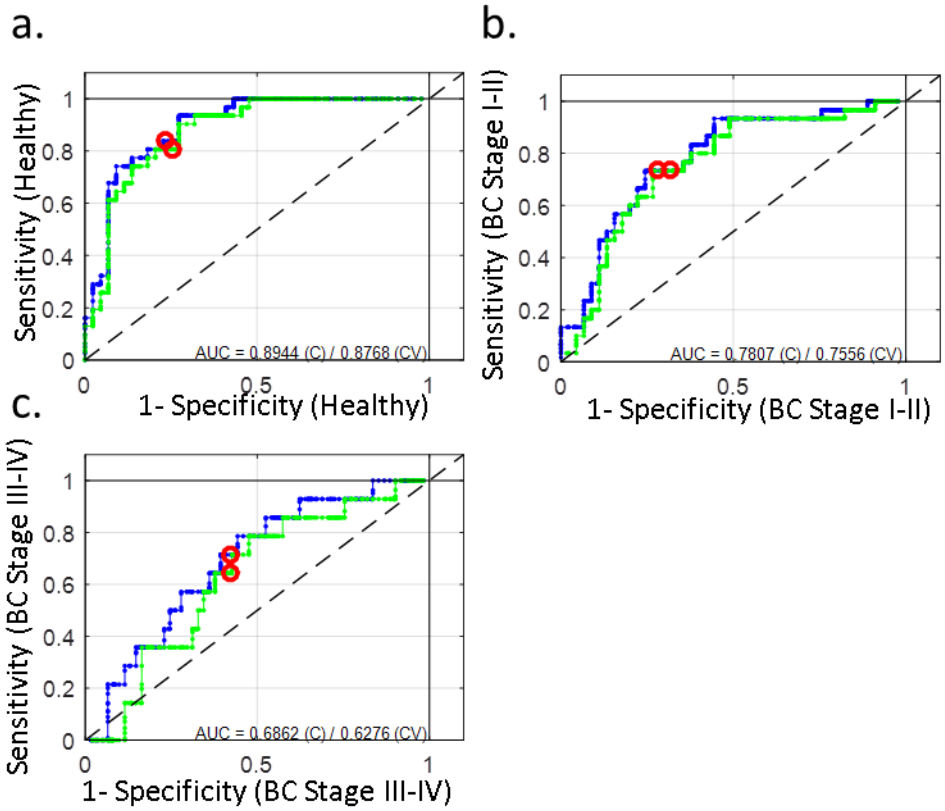


Figure 4.6 ROC curves from PLS-DA analysis of ERα, PR, and CDKN2D in **a)** healthy serum, **b)** stage I-II breast cancer serum and **c)** stage III-IV breast cancer serum. Estimated values from calibration are shown in blue and cross validated values in green. The red circles indicate the threshold values (standard and cross-validated) determined by the PLS-DA model on the calibration data.

In addition to the presence or stage of cancer, another important factor in breast cancer diagnostics is the determination of breast cancer subtype. Different molecular subtypes of breast cancer have very different treatment plans and thus the ability to distinguish which type of cancer a patient has could aid in patient care. Since only serum samples from patients with luminal and TNBC were available, only these two subtypes were investigated in this study. A total of 30 luminal and 14 TNBC serum samples varying from stages I-IV were used to

calibrate the PLS-DA model (Model 4). (**Figure 4.7).** The AUCs from the resulting ROC curves were both 0.73 (CV = 0.65) and the true positive rates of the model were 63% and 78% for luminal and TNBC, respectively. Finally, the precision was 86% for luminal and 50% for TNBC and the overall accuracy of Model 4 was 68%. These results suggest that Model 4 is moderately proficient at distinguishing these two subtypes, but more samples should be tested and possibly more variables introduced (i.e. more biomarkers) to create a more robust model.

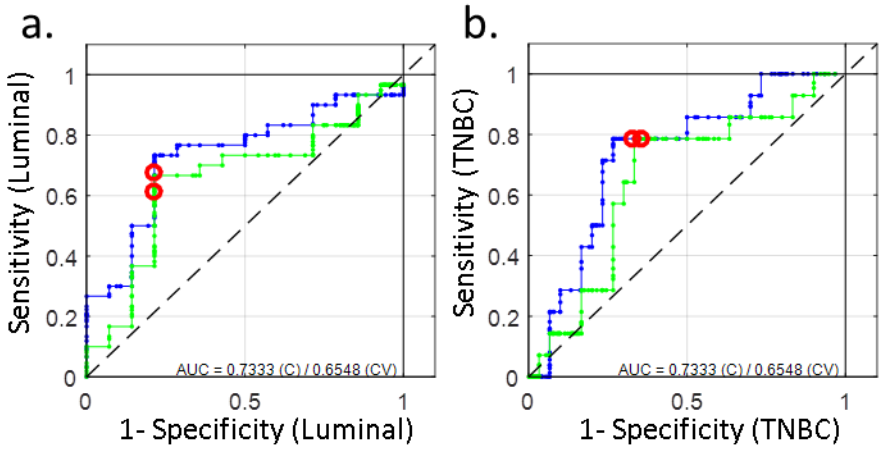


Figure 4.7 ROC curves from PLS-DA analysis of ERα, PR, and CDKN2D in **a)** luminal and **b)** TNBC. Estimated values from calibration are shown in blue and cross validated values in green. The red circles indicate the threshold values (standard and cross-validated) determined by the PLS-DA model on the calibration data.

To summarize the results of the developed models, **Table 4.2** lists the description and results from each model. From this table it is clear that both Models 1 and 2 perform the best. Model 2 performs slightly better, and this result is favorable since the overall goal of this study was to test a biomarker panel to achieve early detection with SiMoA. The following section describes further

analysis using Model 2 to determine the impact of each biomarker within this model.

Table 4.2 Summary of results for all four PLS-DA models

	Description	Precision	True Positives	AUC (CV)	Number of samples	Overall Accuracy
Model 1	Healthy	72%	90%	0.88	31	81%
	BC Stage I-IV	92%	75%	0.88	44	
Model 2	Healthy	82%	87%	0.89	31	84%
	BC Stage I-II	86%	80%	0.89	30	
Model 3	Healthy	86%	81	0.88	31	45%
	BC Stage I-II	62%	70%	0.76	30	
	BC Stage III-IV	0%	0%	0.63	14	
Model 4	TNBC	50%	78%	0.65	14	68%
	Luminal	86%	63%	0.65	30	

In order to determine which variable has the most impact on Model 2, the accuracy of the model was assessed after different variables were removed. A total of 61 samples were used for different model calibrations: 30 stage I-II samples and 31 healthy samples. As previously mentioned, age was implemented as a factor in the PLS-DA model, thus a total of four variables exist: age, ERα, PR, and CDKN2D. **Figure 4.8** shows the results of this analysis in both a table and a bar graph. First, all of the variables are kept in the model for comparison, so the accuracy is 84% as previously reported. In the model where only ERα is removed, the accuracy decreases to 77%. The accuracy decreases similarly when only CDKN2D is removed as well as when ERα and CDKN2D are removed

together. In the model where only PR is removed, the impact on accuracy is slightly less, dropping only to 82%. When only ERα and age are present, the accuracy drops to only 79%, and to 77% when only CDKN2D and age are present. These results suggest that ERα has the greatest impact on the model and that PR has the lowest impact; however, it is clear that the model performs best when all variables are included.

a.

Description	Total Number of Samples	Number of Samples Correctly Assigned	Accuracy
All (ERα, PR, CDKN2D, Age)	61	51	84%
No ERα	61	47	77%
No PR	61	50	82%
No CDKN2D	61	47	77%
Only ERα and Age	61	48	79%
Only PR and Age	61	47	77%
Only CDKN2D and Age	61	47	77%

b.

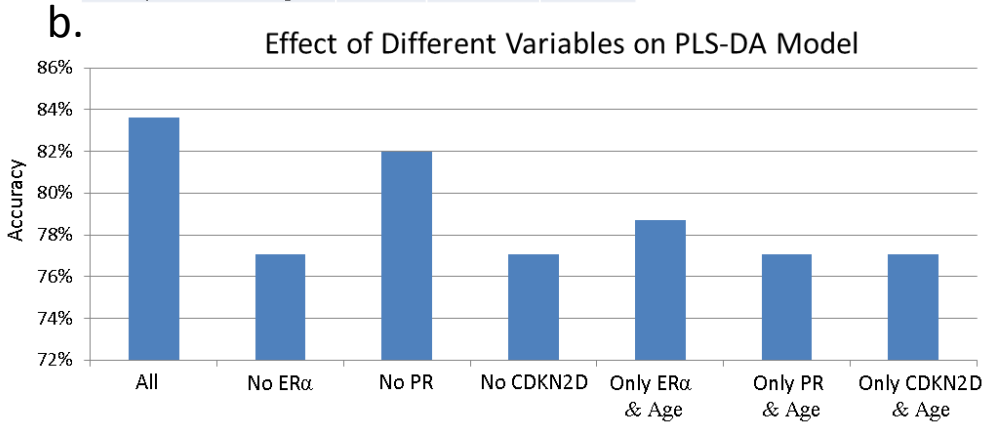


Figure 4.8 PLS-DA models calibrated to determine the impact of each variable in Model 2. **a)** Table describing each model, the samples used and correctly assigned, and the resulting accuracy. **b)** Graphical depiction of each model.

4.7 Discussion

This work presents the use of a sensitive protein biomarker panel consisting of ERα, PR, and CDKN2D to detect early stage breast cancer in serum using SiMoA technology. The SiMoA assays for each biomarker are more

sensitive than available commercial ELISA kits. Additionally, the only commercial ELISA kit for CDKN2D is discontinued, meaning that the SiMoA assay presented in this work may represent not only the most sensitive ELISA assay for this biomarker, but possibly the only one. The increased sensitivity gained through SiMoA was critical for this work since many of the analyzed samples were below the reported LODs of the most sensitive commercial ELISAs both when diluted and when the dilution factor was considered (see Appendix **Figure A2**). The ability to dilute samples and use small volumes but still achieve high sensitivity makes SiMoA ideal for studying precious banked samples or samples with low volume. Although the creation of SiMoA assays for these three biomarkers presents a significant improvement in sensitivity, some samples were still undetectable using SiMoA for both ERα and PR, so further improvements in these assays are desirable.

Each biomarker was individually assessed to determine its ability to distinguish healthy samples from cancer samples. Both ERα and CDKN2D were able to statistically separate healthy and breast cancer samples. In addition, CDKN2D could differentiate between healthy and stage I-II breast cancer as well as between healthy and luminal type breast cancer. Surprisingly, PR alone was unable to distinguish between any of the samples.

Statistical models using PLS-DA were derived from data using both the responses from each biomarker and the age of each patient. Despite the low number of samples available, models were built for a total of four different situations: comparison of healthy vs. breast cancer, healthy vs. early stage (I-II),

healthy vs. early and late stage (III-IV), and luminal vs. TNBC. The developed models demonstrated that the use of ERα, PR, CDKN2D and age could be used to distinguish between both healthy and cancer samples as well as healthy and early stage breast cancer. The model used to distinguish between early stage breast cancer and healthy samples (Model 2) had a false positive rate of 20% and a false negative rate of 13%, which is comparable to the 8-10% false positive rate and ~20% false negative rate from mammography testing.[4, 36] Obtaining such similar false positive and false negative rates to the current gold standard from preliminary data is extremely promising for future work utilizing SiMoA as an early breast cancer detection method.

It was determined that both ERα and CDKN2D contributed the most to the early detection method, while PR contributed the least. ERα and PR are routinely used as tissue biomarkers for establishing breast cancer subtype upon diagnosis. Both ERα and PR have been studied in serum via qPCR as part of the OncotypeDxR biomarker panel,[37] which is approved by the ASCO for predicting breast cancer risk. The bioactivity of ERα has also been studied in serum and has been shown to be related to increased breast cancer risk.[38] To the best of our knowledge, CDKN2D has not been investigated in serum for breast cancer diagnostics, although it has been associated with breast cancer progression. This work demonstrates the first study in which SiMoA is used to detect early stage breast cancer. Although the results are preliminary, it is promising that the use of these three biomarkers along with patient age may offer the ability to non-invasively screen patients for the early detection of breast cancer or to aid in the

diagnosis of women who have had abnormal mammograms and would like an alternative to biopsy.

4.8 Future Directions

Although this work presents very promising results, there is a significant amount of work that still needs to be performed. First, more samples need to be tested in order to further validate the models. All samples used in this study were purchased from Bioreclamation and all patients were already undergoing treatment, such as chemotherapy or hormone therapy, or had surgery. It is very important to obtain samples that are from patients who have not yet undergone therapy or surgery for early detection to eliminate or reduce confounding factors when attempting to create an accurate and robust model. Additionally, it would be interesting to see if the assays and statistical modeling approach proposed in the present work can determine the difference between patients who are healthy as well as those who have benign tumors, indolent tumors, and aggressive tumors.

The addition of more biomarkers to the panel should also make it more robust. Although beyond the scope of this thesis, SiMoA assays have been developed for other biomarkers on the list by other Walt lab collaborators on this project, including cysteine-rich angiogenic inducer 61 (CYR-61), cancer antigen 19-9 (CA19-9), disintegrin and metalloproteinase domain-containing protein 12 (ADAM12), neural precursor cell expressed developmentally down-regulated protein 9 (NEDD9), and HER2. The addition of these and other assays yet to be established offers a significant opportunity for creating a sensitive and accurate

breast cancer panel using SiMoA and may strengthen faults in the current panel, such as the ability to differentiate between different stages and subtypes of breast cancer. The creation of a robust panel may one day enable early detection and may also allow the panel to be utilized for patient monitoring during therapy and remission to improve patient care and prevent relapse.

4.9 Materials and Methods

All serum samples were purchased from BioreclamationIVT. Serum samples were diluted by a factor of four in Sample Diluent that consisted of 10 ug/mL TruBlock (Meridian Life Science, A66800H) 5 mM EDTA (Invitrogen, 15575-020), and PBS (Sigma Aldrich, P5493-1L). All water used was Milli-Q water. All buffers are vacuum filter sterilized prior to use (VWR 28199-774).

SiMoA Assays

SiMoA assays were carried out using procedures and reagents described in Chapter 2 unless otherwise described below.

CDKN2D capture antibody was purchased from LifeSpan Biosciences (LS-C37972) and detection antibody was purchased from Abnova (H00001032-D01P). CDKN2D protein standard was purchased from OriGene (TB314065). (CDKN2D calibrators are prepared in 25% Newborn Calf Serum (Life Technologies, 16010-142), 0.01% Tween-20 (Sigma, P7949), 5 mM EDTA (Invitrogen, 15575-020), 0.15% ProClin300 (Sigma Aldrich, 48914-U), and PBS

(Sigma Aldrich, P5493-1L). 500 pM of SβG and 1 μg/mL of detector were used. A 2-step assay was used on the HD-1 with incubations of 35 and 5 minutes.

ERα capture and detection antibodies and protein standard were purchased in a kit from R&D systems (DYC5715-5). The detection antibody was prebiotinylated. Both the protein standard and detection antibody were resuspended in 1% bovine serum albumin (BSA) (Millipore 82-045-1) and PBS. Calibrators were also diluted in 1% BSA and PBS. A 2-step assay was used on the HD-1 with incubations of 35 and 5 minutes.

PR capture and detection antibodies and protein standard were purchased in a kit from R&D systems (DYC5415-5). The detection antibody was prebiotinylated. Both the protein standard and detection antibody were resuspended in 1% BSA and PBS. Calibrators were also diluted in 1% BSA and PBS. A 2-step assay was used on the HD-1 with incubations of 35 and 5 minutes.

Data Analysis

Box plot analysis was performed in OriginPro 9.1 (OriginLab). PLS-DA was performed in Matlab R2014a (The Mathworks, Inc.) using PLS_Toolbox 7.9.4 (Eigenvector Research). For each PLS-DA model, 80% of the data was randomly used to build a calibration and the remaining 20% was used for cross validation. Iterations of this process were repeated until each data point was used, and then the entire process was repeated a total of four times. Due to sample volume constraints, not all samples were tested using each biomarker and thus some data points were missing in the panel. In order to build PLS-DA models,

data imputation was implemented via PLS_toolbox 7.9.4. HD-1 data analysis was performed using the HD-1 Analyzer (Quanterix Corp.). Assay LODs were calculated by extrapolating the three standard deviations from the blank after fitting to a 4-PL regression.

4.10 References

1. American Cancer Society.Breast cancer survival rates by stage.26 February 2015.Web.6 March 2015.http://www.cancer.org
2. DeSantis, C., Ma, J., Bryan, L. & Jemal, A. Breast cancer statistics, 2013. *CA: A Cancer Journal for Clinicians* **64**, 52-62 (2014).
3. Berry, D.A. et al. Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer. *N. Engl. J. Med.* **353**, 1784-1792 (2005).
4. Taplin, S. et al. Mammography Facility Characteristics Associated With Interpretive Accuracy of Screening Mammography. *Journal of the National Cancer Institute* **100**, 876-887 (2008).
5. American Cancer Society. Breast Cancer Facts & Figures 2013-2014.Atlanta: (2013)
6. Harris, L. et al. American Society of Clinical Oncology 2007 Update of Recommendations for the Use of Tumor Markers in Breast Cancer. *Journal of Clinical Oncology* **25**, 5287-5312 (2007).
7. Duffy, M.J. Serum Tumor Markers in Breast Cancer: Are They of Clinical Value? *Clin. Chem.* **52**, 345-351 (2006).
8. Song, L. et al. Single molecule measurements of tumor necrosis factor α and interleukin-6 in the plasma of patients with Crohn's disease. *J. Immunol. Methods* **372**, 177-186 (2011).
9. Harris, L. et al. American Society of Clinical Oncology 2007 Update of Recommendations for the Use of Tumor Markers in Breast Cancer. *Journal of Clinical Oncology* **25**, 5287-5312 (2007).
10. Anderson, K.S. et al. Application of Protein Microarrays for Multiplexed Detection of Antibodies to Tumor Antigens in Breast Cancer. *Journal of Proteome Research* **7**, 1490-1499 (2008).
11. Anderson, N.L. & Anderson, N.G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Molecular & Cellular Proteomics* **1**, 845-867 (2002).
12. Zhu, C.S. et al. A Framework for Evaluating Biomarkers for Early Detection: Validation of Biomarker Panels for Ovarian Cancer. *Cancer Prevention Research* **4**, 375-383 (2011).
13. Heldring, N. et al. Estrogen Receptors: How Do They Signal and What Are Their Targets. 87. 2007.
14. Gustafsson, J.-Å. What pharmacologists can learn from recent advances in estrogen signalling. *Trends Pharmacol. Sci.* **24**, 479-485 (2003).
15. Hall, J.M., Couse, J.F. & Korach, K.S. The Multifaceted Mechanisms of Estradiol and Estrogen Receptor Signaling. *J. Biol. Chem.* **276**, 36869-36872 (2001).
16. Deroo, B.J. & Korach, K.S. Estrogen receptors and human disease. *The Journal of Clinical Investigation* **116**, 561-570 (2006).
17. Mulac-Jericevic, B. & Conneely, O.M. Reproductive tissue selective actions of progesterone receptors. *Reproduction* **128**, 139-146 (2004).
18. Henderson, B.E. & Feigelson, H.S. Hormonal carcinogenesis. *Carcinogenesis* **21**, 427-433 (2000).
19. Kougioumtzi, A., Tsaparas, P. & Magklara, A. Deep Sequencing Reveals New Aspects of Progesterone Receptor Signaling in Breast Cancer Cells. *PLoS ONE* **9**, 1-13 (2014).

20. Felisiak-Golabek, A. et al. p19INK4d mRNA and protein expression as new prognostic factors in ovarian cancer patients. *Cancer Biology & Therapy* **14**, 973-981 (2013).
21. Cánepa, E.T. et al. INK4 proteins, a family of mammalian CDK inhibitors with novel biological functions. *IUBMB Life* **59**, 419-426 (2007).
22. Denoyelle, C. et al. Molecular mechanism of the anti-cancer activity of cerivastatin, an inhibitor of HMG-CoA reductase, on aggressive human breast cancer cells. *Cellular Signalling* **15**, 327-338 (2003).
23. Corder, G.W. & Foreman, D.I. Nonparametric Statistics: A Step-by-Step Approach. 2nd. Wiley, Hoboken: 2014.
24. James, G., Witten, D., Hastie, T. & Tibshirani, R. An Introduction to Statistical Learning with Applictions in R. Springer, New York: 2013.
25. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109-130 (2001).
26. Barker, M. & Rayens, W. Partial Least Squares for Discrimination. *J. Chemom.* **17**, 166-173 (2003).
27. Szymańska, E., Saccenti, E., Smilde, A.K. & Westerhuis, J.A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics : Official journal of the Metabolomic Society* **8**, 3-16 (2012).
28. Asiago, V.M. et al. Early Detection of Recurrent Breast Cancer Using Metabolite Profiling. *Cancer Res.* **70**, 8309-8318 (2010).
29. Keun, H.C. et al. Serum Molecular Signatures of Weight Change during Early Breast Cancer Chemotherapy. *Clinical Cancer Research* **15**, 6716-6723 (2009).
30. Cao, M.D. et al. Prognostic value of metabolic response in breast cancer patients receiving neoadjuvant chemotherapy. *BMC Cancer* **12** (2012).
31. Bertuzzi, M. et al. Plasma clusterin as a candidate pre-diagnosis marker of colorectal cancer risk in the Florence cohort of the European Prospective Investigation into Cancer and Nutrition: a pilot study. *BMC Cancer* **15** (2015).
32. Farshidfar, F. et al. Serum metabolomic profile as a means to distinguish stage of colorectal cancer. *Genome Med* **4**, 1-13 (2012).
33. Li, Z. et al. Analysis of the raw serum peptidomic pattern in glioma patients. *Clin. Chim. Acta* **425**, 221-226 (2013).
34. Nishiumi, S. et al. Serum metabolomics as a novel diagnostic approach for pancreatic cancer. *Metabolomics* **6**, 518-528 (2010).
35. Hanley, J.A. & McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29-36 (1982).
36. American Cancer Society.American Cancer Society Guidelines for the Early Detection of Cancer.29 October 2014.04 April 2015.http://www.cancer.org/
37. Paik, S. et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *N. Engl. J. Med.* **351**, 2817-2826 (2004).
38. Widschwendter, M. et al. Serum oestrogen receptor [alpha] and [beta] bioactivity are independently associated with breast cancer: a proof of principle study. *Br J Cancer* **101**, 160-165 (2009).

Chapter 5

Protein Counting in Single Cancer Cells

5.1 Introduction

The cell is the basic unit of biology and protein expression drives cellular function. In order to fully understand how cell heterogeneity affects population behavior and biological function, it is essential to study protein expression within single cells. Previous chapters have focused on traditional uses of SiMoA for applications in early cancer detection. This chapter describes the importance of single cell analysis and single molecule protein detection within single cells. Also discussed is the adaptation of the SiMoA platform to enable single cell studies. Finally, I describe how the single cell SiMoA platform is used to study the distribution of PSA expression within single cells from two related cancer cell lines. Using this single molecule counting method to count the proteins in single cells enables the quantification of phenotypic responses with single cell resolution, which is evidenced by our ability to track a 30-fold difference in PSA concentration between two cell lines as a result of genetic drift. Single cell SiMoA introduces a straightforward process that is capable of detecting both high and low protein expression levels in single cells. This technique could be vital for understanding fundamental biology, such as molecular mechanisms, pathways, and cell heterogeneity and may eventually enable both earlier disease detection and targeted therapy.[1]

5.1.1 Motivation for Single Cell Studies

Cellular processes are stochastic and gene expression can vary widely between individual cells, spanning several orders of magnitude.[2-4] Cell development and activity are dictated by protein expression. Changes in protein concentration can affect the cell phenotype, resulting in dramatic consequences for processes such as cell growth, metabolism, and disease progression.[1, 5, 6] Most studies of cellular biochemistry are based on bulk measurements of many cells and such ensemble experiments can only yield averages that may not be indicative of the actual population distributions present at the single cell level, as summarized by DiCarlo *et al.* in **Figure 5.1**.[7, 8] For example, single-cell time-responses can be used to study the dynamics of how genetically identical cells respond to stimuli over time. As depicted in **Figure 5.1a**, if only a few cells respond strongly to an applied stimulus and bulk techniques are used, then the resulting bulk average gene expression of the population masks the response from the cells of interest (**Figure 5.1a)**.[8] In addition, cell populations that demonstrate out-of-phase dynamics (**Figure 5.1b**) or respond to stimuli via bimodal expression (**Figure 5.1c**) will provide uninformative bulk averages.[8] Another important example where bulk measurements mask single cell characteristics is gene expression studies containing rare cell populations (**Figure 5.1d**).[8]

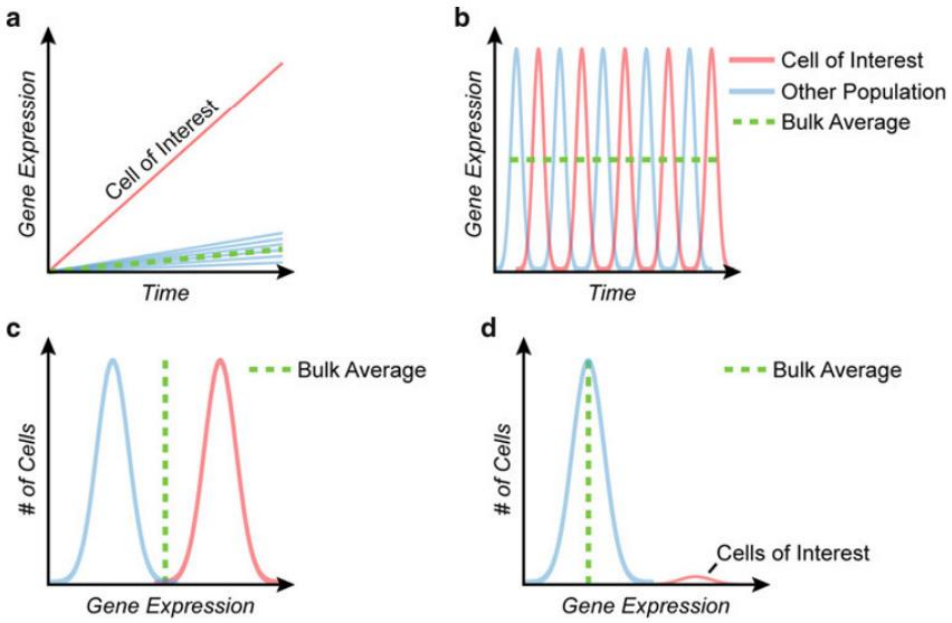


Figure 5.1 Description of bulk averaging vs. single cell analysis in gene expression. **a)** Bulk averaging of rate measurements mask the difference in amplitude between the desired cells of interest and the majority population. **b** Gene expression information from cells exhibiting out-of-phase dynamics is averaged in overall bulk populations. Bulk end-point assays do not accurately represent subpopulations within both **c)** multimodal distributions and **d)** rare cell populations (Reprinted with permission from Reference 8).

It is well known that the protein expression of a specific gene varies from cell to cell.[2] In addition to the above examples, single cell analysis also enables the ability to track protein expression in single cells to study cellular functions and pathways. However, this feat requires methodologies sensitive enough to detect low numbers of protein molecules with a wide dynamic range to distinguish unique cells and quantify population distributions;[1] endeavors that are not possible using traditional ensemble methodologies.[8] There have been numerous single cell studies that measure mRNA using the transcriptome as a surrogate for the proteome.[9-11] While these measurements are useful, others have

shown that the amount of mRNA does not correlate with the amount of protein expressed.[4] The genome provides a map for protein synthesis, but knowing the genome or the transcriptome does not directly correlate with knowing the proteome due to the stochastic nature of biological processes and other factors, such as the cellular microenvironment.[4, 12] Studying protein expression at the single cell level can yield insight regarding cellular functions and pathways, enabling the study of cell-to-cell variations and stochasticity.[13, 14] Therefore, in order to truly understand the complexities of many biological processes, protein expression must be characterized at the single cell level. The following section describes current progress in the field of single cell proteomic studies.

5.1.2 Current Technologies for Single Cell Proteomic Studies

Flow cytometry is an extremely high throughput technique that is most commonly used for protein analysis in single cells.[1] This technique utilizes fluorescently labeled antibodies to label and then simultaneously detect as many as 18 different proteins.[15-18] The invention of Fluorescence-Activated Cell Sorting (FACS),[19] a type of flow cytometry that analyzes and separates cells from heterogeneous cell mixtures based on differential fluorescent labeling, has made a significant impact on the field of single cell analysis. The throughput of flow cytometers and cell sorters is high – up to thousands of single cells per second.[18] Although flow cytometry has laid much of the groundwork for single cell protein analysis due to its high throughput and multiplexing capabilities,[8] the resulting

signals from each cell represents the total fluorescence for each label used to identify the cell. Furthermore, no temporal information can be obtained.

Western blots, another common method for protein quantification, have recently been adapted for single cell analysis.[20] Single-cell Western blots (scWesterns) enable the simultaneous analysis of approximately 2,000 individual cells in less than 4 hours by implementing a 30-μm-thick photoactive polyacrylamide gel patterned with thousands of 20 × 30 μm (w×d) wells.[20] Although this technique has vastly improved upon both traditional Western blotting and other recent advances, such as microfluidic Western blotting[21], the resulting blots in scWesterns are quantified based on the bulk fluorescence response for the protein in question from each cell. These techniques acquire the total fluorescence response for the protein in question for each interrogated cell. As opposed to integrating the protein concentration from single cells, the ability to count single protein molecules ithin single cells would further strengthen the ability to study basic cellular functions as well as more complex cellular systems.

5.1.3 Single Molecule Protein Detection in Single Cells

Recent advances in single molecule and single cell detection have pushed the boundaries of both biological and chemical detection limits. Several studies have successfully overcome experimental limitations in sensitivities to achieve single molecule protein detection at the single cell level.[13, 14, 22-24] However, many of the aforementioned single cell single molecule protein analysis methods

involve complicated experimental designs, require external cellular stimulation, genetic modification, sample amplification, and/or arduous algorithms[13, 14, 22, 23]. A few notable and recent advances in the field of single molecule protein detection in single cells are described below.

Xie *et al*. notably reported the quantification of both protein and mRNA within single *E.coli* cells.[4] This work noted the lack of correlation between the number of protein and mRNA molecules for any given gene, underscoring the importance of protein analysis.[4] The assay was performed in a microfluidic device where the protein copy number is obtained via fluorescence imaging of YFP-fusion library strains. Fluorescence in situ hybridization (FISH) was used to measure single molecule mRNA expression.[4] Despite the groundbreaking impact this work has had on the field, the requirement to create genetically modified fusion libraries makes it impractical for clinical studies and difficult to implement into other assay designs.

Zare and coworkers presented a microfluidic chip capable of counting proteins at low copy numbers within single cells.[13, 25] This methodology involves trapping individual cells using a system of valves followed by flowing lysis buffer and a subsequent labeling reagent. A custom optical set-up was used to count the number of fluorescent bursts generated as molecules flowed through a small detection volume. Despite the authors' achievement of obtaining low protein counts from single cells, a complicated algorithm is required to compensate for false negatives in samples containing too many molecules per frame. In addition, the described technique is only able to analyze up to three cells at a time due to

the chip design, making it extremely low throughput. Finally, this method requires genetic modification for non-fluorescent samples.

Klug has recently used two different single molecule techniques for quantifying proteins in single cells. The first method involves incorporating a microfluidic microarray with TIRF detection. The platform functions within the mass sensing regime such that the majority of analyte is detected, enabling a detection limit of 21 molecules while maintaining a dynamic range greater than three orders of magnitude.[22] Although this detection limit is impressive, the microfluidic device requires nanoliter sized volumes, which requires complicated preparation. Unfortunately, the use of these devices is not high throughput. Isolation of single cells is performed via optical trapping, which further complicates the set-up. In addition, imaging on the chip occurs via a 'detect and bleach method' where the fluorescent signal from the sample is detected and followed by subsequent photobleaching until all of the molecules present in the sample are counted. This methodology is time consuming and requires complicated downstream data analysis to reduce bias.

In the second study by Klug and coworkers, they describe the first implementation of single molecule protein detection in single cells within droplets.[24] In this study, they used a simple device to dispense droplets onto coverslips spotted with a specific antibody via a micro-contact arrayer. Single cells were loaded into the droplets via micromanipulation followed by optical lysis. The number of proteins bound to the antibody spot was then monitored over time via single molecule TIRF microscopy. This technique was only sufficient at

quantifying protein copy numbers from 10^6 to 10^8, meaning that lower abundant proteins could not be measured using this technique.

The above techniques and methodologies briefly describe the advances in the field of single molecule proteomic studies in single cells. However, as mentioned, there are several shortcomings in terms of throughput, design limitations, and sensitivity. Therefore, there is a need for a robust platform that can sensitively analyze protein content within single cells. The following section describes the implementation of the previously described SiMoA platform for single cell analysis in an effort to overcome these disadvantages.

5.2 Development of Single Cell SiMoA Platform

This section demonstrates that SiMoA technology can be employed to fully quantify protein expression in single prostate cancer cells. This approach, based on single molecule counting techniques, is straightforward and ultrasensitive – eliminating time-consuming microchip assembly and any reliance on complicated algorithms, amplification steps, external stimulation, or genetic engineering, all of which can introduce bias. Previous work has shown that SiMoA can dramatically improve detection limits compared to traditional ELISA[26], exhibit a wide dynamic range[27], and enable multiplexed protein analysis[28], making SiMoA an attractive technology for single cell protein studies.

In order to modify the SiMoA process for single cell analysis, cells first needed to be isolated. Cell isolation was performed using standard laboratory

equipment and no expensive or intricate set-ups were required. Since LNCaP cells are adherent, cells were first suspended in media and washed 3x in DPBS via centrifugation. The cells were then counted and diluted to a concentration of ~2 x 10^3 cells/mL in DPBS. Cells were isolated by transferring 1 µL of the washed cell solution into the cap of a flat, optically clear PCR tube (**Figure 5.2a**). The presence of only one cell inside of the droplet was visually validated using a microscope. Representative images of an isolated cell are shown in **Figures 5.2b-c**. Once it was verified that only a single cell was present, the bottom of the PCR tube was placed on the cap for either immediate use or storage. It is important to note that no labeling was necessary to visualize the individual cells.

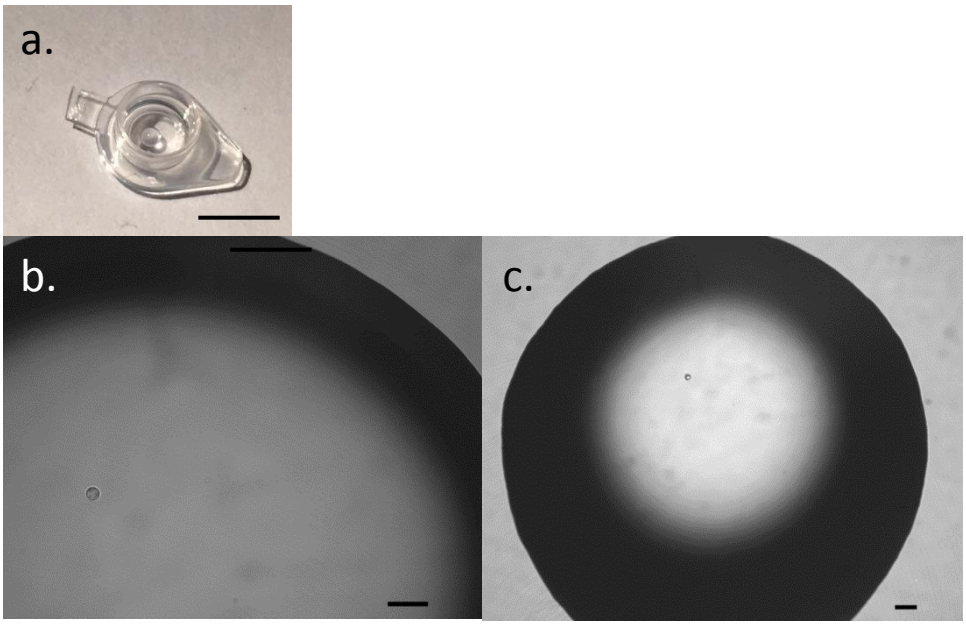


Figure 5.2 Single cell isolation. **a)** Photograph of 1 µL droplet inside optically clear PCR cap. Scale bar is 5 mm. Representative bright field images of a single LNCaPB cell inside a 1 µL droplet visualized at **b)** 10× and **c)** 5× magnification. Scale bars are 50 µm.

Lysis was performed by adding 64 µL of lysis buffer to the PCR tubes containing isolated cells. The lysate was then transferred to a 96-well PCR plate containing 75 µL of diluent for a total assay volume of 140 µL. As described in Chapter 2, the HD-1 analyzer is a fully automated system and once all reagents are loaded and the experiment is programmed, the instrument completes the entire single molecule analysis process. **Figure 5.3** outlines the single cell SiMoA process. First, single cells are isolated as described above. The cell lysate sample is then incubated with capture beads, biotinylated detection antibody, and SBG with wash cycles occurring between each step. The beads are then resuspended in RGP and loaded onto the SiMoA disc arrays, sealed with oil, and imaged.

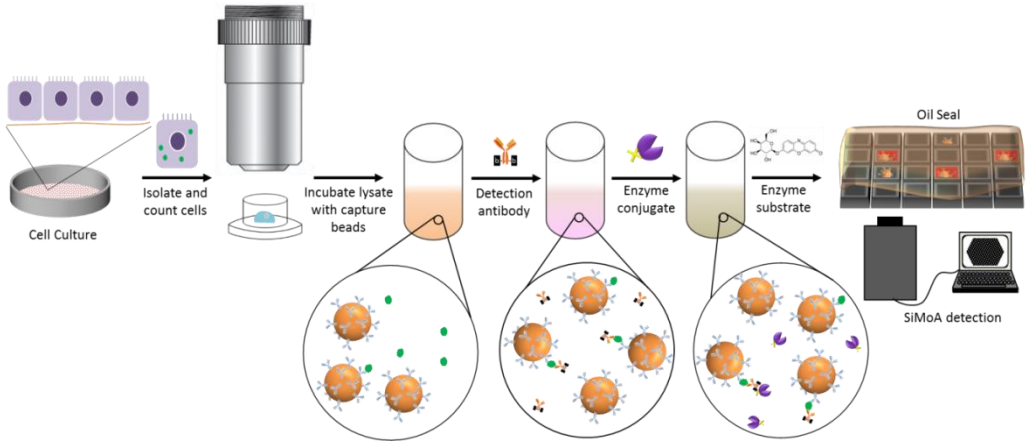


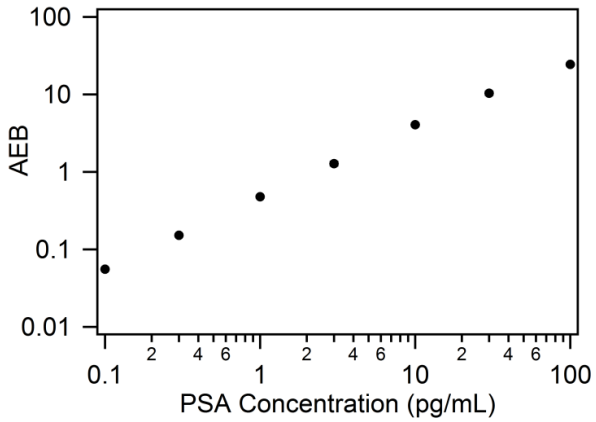
Figure 5.3 Experimental scheme for single cell SiMoA analysis. Cells were isolated manually by visual inspection under a microscope. The cells were then lysed and loaded into the SiMoA HD-1 analyzer, which performed the subsequent incubations with capture beads, detection antibody, and enzyme conjugate. After forming the enzyme-labeled immunocomplex on the beads, enzyme substrate was added, the beads were loaded into an array of wells, and the wells were sealed for imaging. (Figure by Stephanie R. Walter)

.

In the following study, SiMoA is employed to determine PSA expression in single LNCaP cells. A representative PSA calibration curve is shown in **Figure**

5.4. The LOD for PSA in this assay was 0.0043 ± 0.0022 pg/mL, which equates to ~12,000 PSA molecules in 140 μL.

a.



b.

PSA conc. (pg/mL)	Average AEB	Standard deviation
0	0.0068	0.0005
0.1	0.056	0.002
0.3	0.152	0.005
1	0.478	0.006
3	1.28	0.07
10	4.1	0.1
30	10.4	0.5
100	24.4	0.6

Figure 5.4 PSA calibration curve. **a)** SiMoA calibration curve for PSA plotted on a log-log scale. Error bars are contained within data points and represent n=3 measurements. The assay LOD of 0.0043 ± 0.0022 pg/mL was calculated by extrapolating the background PSA concentration plus 3 standard deviations of the background using a 4-parameter logistic fit. **b)** Tabulated values for typical PSA calibration curve.

5.3 Counting PSA molecules in LNCaP Cells

In this proof-of-concept study, PSA is quantified in two related prostate cancer cell lines: a low passage LNCaP cell line ($LNCaP_A$) and an over sub-cultured LNCaP cell line that has undergone genetic drift ($LNCaP_B$). These two LNCaP cell lines represent models for high and low protein expression, both of which are easily detected using SiMoA and require minimal sample preparation. Previous research has demonstrated that highly cultured LNCaP cells secrete significantly altered concentrations of PSA[29]. Using SiMoA, we measure the impact of genetic drift on protein expression at the single cell level and show that

$LNCaP_B$ cells have significantly depressed PSA expression compared to $LNCaP_A$. These results indicate that genetic instability in cancer cells affect protein expression, and by extension, cancer progression, and also highlight the necessity of cell culture authentication. Importantly, we establish SiMoA as a unique and facile approach to count protein molecules with single cell resolution and statistics not previously possible to reveal unbiased phenotypic information.

5.3.1 Cell Line Verification

$LNCaP_A$ was purchased directly from ATCC, while $LNCaP_B$ was obtained through a collaborator. The authenticity of $LNCaP_A$ was certified by ATCC upon purchase. Short tandem repeat (STR) profiling of $LNCaP_B$ was carried out using the Promega Cell Line Authentication Sample Collection Kit. The STR profiles, compared in **Table 5.1**, show that $LNCaP_B$ exhibits an 88% match to $LNCaP_A$. The apparent 12% genetic drift in $LNCaP_B$ is attributed to the extensive sub-culturing of this cell line and the genetic instability of cancer cells[29]. In the following sections, we quantify how a 12% genetic drift in highly sub-cultured cells alters the PSA expression in single cells.

Table 5.1. STR Profile Comparison of $LNCaP_A$ and $LNCaP_B$ cells

Loci	$LNCaP_A$	$LNCaP_B$
D5S818	11, 12	11, 12
D13S317	10, 12	10, 13
D7S820	9.1,10.3	8, 8.1, 9, 9.1
D16S539	11	11,12
vWA	16, 18	16,18
THO1	9	9
AMEL	X,Y	X, Y
TPOX	8, 9	8, 9
CSF1PO	10, 11	10, 12
% match to ATCC cat# CRL-1740	100	88

5.3.2 SiMoA Analysis and Comparison of PSA in Single $LNCaP_A$ and $LNCaP_B$ Cells

The PSA content of a total of 124 individual $LNCaP_A$ and 68 $LNCaP_B$ cells were measured using the SiMoA platform. **Figure 5.5** presents a dot plot that illustrates the range in the number of PSA molecules detected in individual cells from both the $LNCaP_A$ and $LNCaP_B$ cell lines. In both cell lines, the number of PSA molecules spans over two orders of magnitude, reflecting the large degree of cell-to-cell variability within the same homogeneous population. The number of PSA molecules observed in single $LNCaP_A$ cells ranged from $4.34 \times 10^4 – 1.52$

$\times 10^7$, with a mean of 2.15×10^6 molecules per cell. In contrast, the $LNCaP_B$ cells ranged from 7.71×10^3-1.19×10^6 molecules, with a mean of 7.04×10^4 molecules per cell. These numbers equate to an average PSA concentration per cell of 1.79 µM (53.7 µg/mL) and 0.0585 µM (1.76 µg/mL) for $LNCaP_A$ and $LNCaP_B$, respectfully, assuming a cell volume of 2 pL.

A previous study by Pinzani and coworkers using immuno-qPCR and LNCaP cells (diluted to approximately 1 cell/tube) determined the median number of PSA molecules within a single LNCaP cell to be approximately 3.3×10^6 with a range of 2.3×10^6 to 4.3×10^6; however, their sample size was very low with n = 5[30]. Our results from the $LNCaP_A$ cell line agree well with these numbers, but our sample size is significantly higher (n=124), yielding better statistics from the high-throughput capabilities of SiMoA. The average PSA concentration for the $LNCaP_B$ cell line is significantly lower than what Pinzani *et al.* determined, but this discrepancy can be attributed to the extensive subculturing of this cell line.

Figure 5.5 also shows the histogram analysis comparing the distribution of PSA expression across both cell lines. Interestingly, we observe two distinct populations of PSA expression with minimal overlap between the $LNCaP_A$ and $LNCaP_B$ cell lines. The vast difference in PSA content between $LNCaP_A$ and $LNCaP_B$ cell lines illustrates how crucial cell line verification is, and how genetic drift can greatly alter cellular biology and experimental results. Although the two cell lines are closely related, we measure over a 30-fold difference in PSA protein expression. The sensitivity of SiMoA allows for single molecule protein counting

in single cells over a wide range of protein concentrations with minimum sample handling.

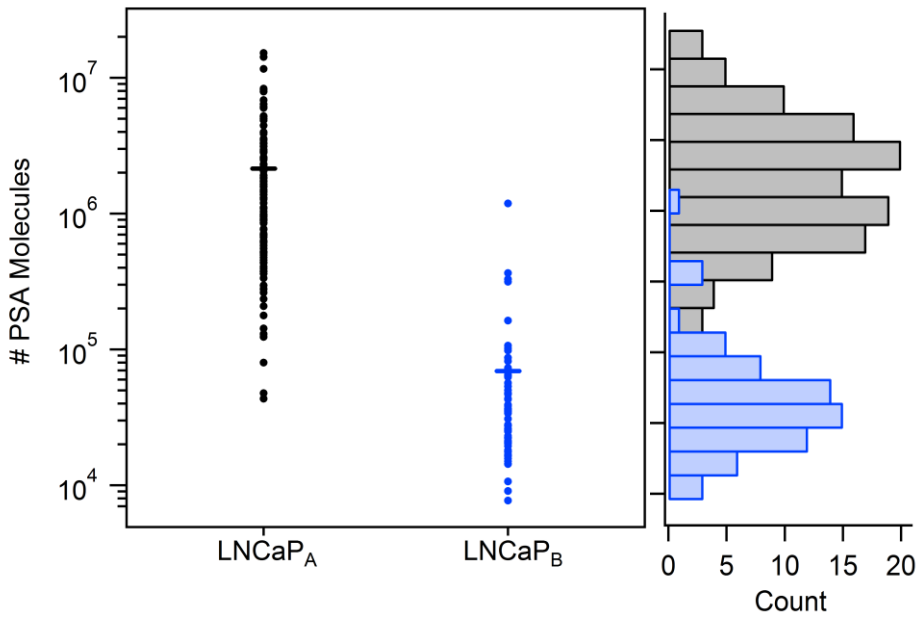


Figure 5.5 Single cell/single molecule analysis of PSA in LNCaP$_A$ and LNCaP$_B$ cells. (Left) Dot plot showing the number of molecules reported in all individual LNCaP$_A$ and LNCaP$_B$ cells. The average of each population is represented with a bar. (Right) Histograms illustrating the log-normal distribution of PSA molecules in individual LNCaP$_A$ (grey) and LNCaP$_B$ (blue) cells.

5.3.3 SiMoA Analysis of PSA in Bulk Cells

For comparison with our single cell analysis, we analyzed low numbers of LNCaP cells to obtain ensemble averages of PSA. In these bulk experiments, cell suspensions were washed, counted, and diluted via serial dilutions to concentrations of $1x10^4$, $5x10^3$, $2.5x10^3$, and $1x10^3$ cells/mL; 10 µL from each prepared concentration (corresponding to 100, 50, 25, and 10 cells, respectively) was lysed and analyzed using SiMoA. **Figure 5.6** shows the average PSA concentration observed for eight replicates containing low cell numbers of

$LNCaP_A$ and $LNCaP_B$ cells. As expected, PSA concentration increases linearly with increasing cell numbers for both $LNCaP_A$ and $LNCaP_B$ (**Figure 5.7** R^2 = 0.968 and 0.986, respectively).

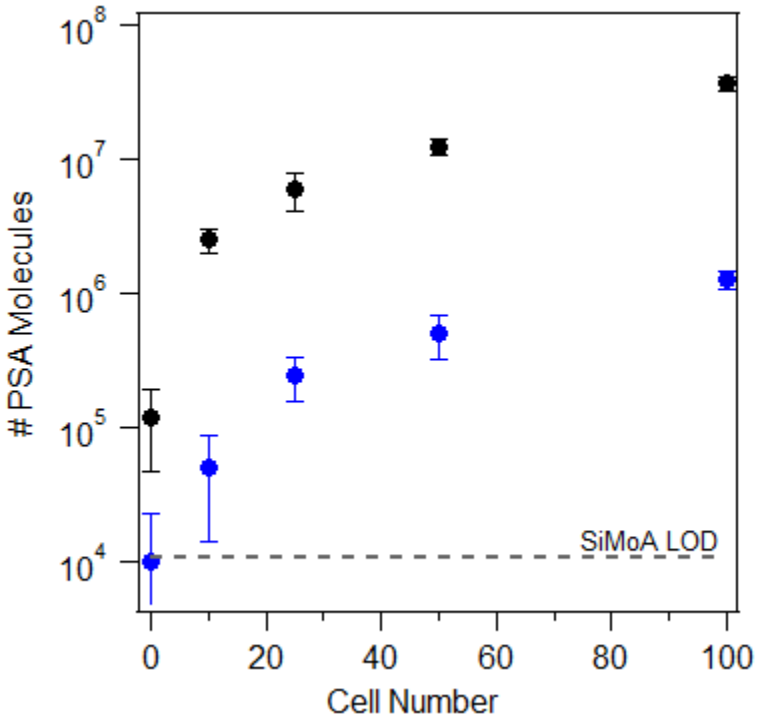


Figure 5.6 PSA in low numbers of LNCaP cells. Plot of PSA content in blank controls and low cell numbers of both $LNCaP_A$ (black) and $LNCaP_B$ (blue) cells. Cell numbers were estimated via serial dilution.

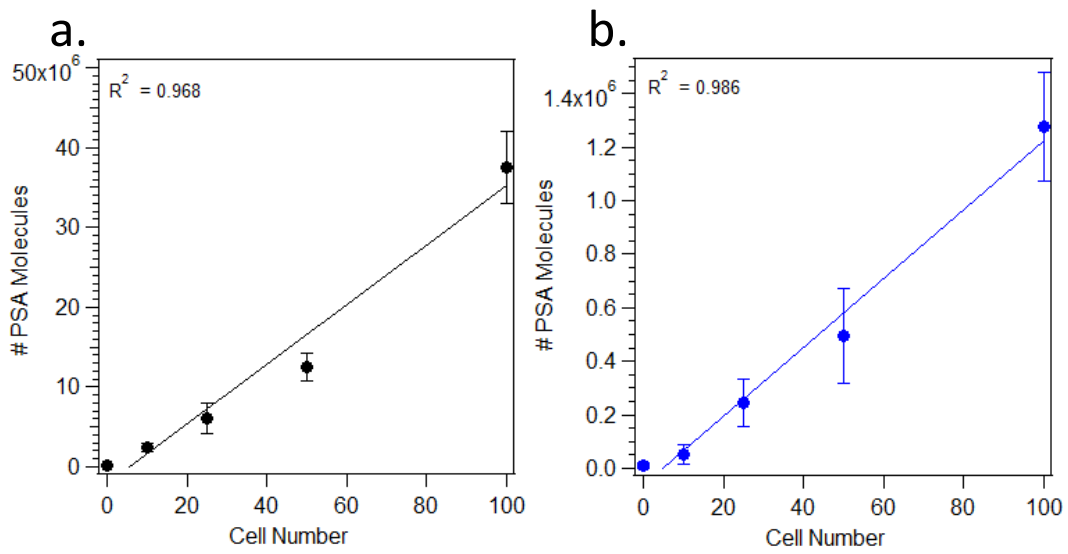


Figure 5.7 Linear fit of PSA content in blank controls and low cell counts of **a)** $LNCaP_A$ and **b)** $LNCaP_B$ cells. Cell numbers were estimated via serial dilution. Error bars represent the standard deviation from eight replicate measurements.

Similar to our findings in single cells, we observe a 30-fold difference in the magnitude of PSA expression between $LNCaP_A$ and $LNCaP_B$ cells. Despite this agreement, we find that dilution-based measurements actually underestimate PSA expression compared to single cell studies. For dilution-based measurements, we observed an average value of 3.75×10^7 PSA molecules for 100 $LNCaP_A$ cells compared to 1.27×10^6 PSA molecules in 100 $LNCaP_B$ cells. Based on these values, one would extrapolate that single $LNCaP_A$ cells contain 3.75×10^5 PSA molecules on average while single $LNCaP_B$ cells contain an average of 1.27×10^4 molecules. However, single cell analysis (*vide supra*) returned average PSA values over five times higher for both cell lines. The likely cause of this discrepancy when measuring protein concentrations using dilution methodologies is that initial cell counts using hemocytometers may be inaccurate.

In addition, cells can stick to tubes and pipettes used in the dilution process, altering the cell count and introducing significant error when cell numbers are low. These experiments further illustrate the value of true single cell protein counts compared to concentrations extrapolated from bulk measurements.

5.4 Discussion

While advances in detecting nucleic acids in single cells have been prominent in the past several years, the work reported here focuses on advancing the important field of single molecule/single cell proteomics. The importance of single cell protein quantification cannot be overstated. Molecular mechanisms, pathways, and cell heterogeneity at the single cell level can be studied to potentially enable both earlier disease detection and targeted therapy by identifying rare cells in a population. Our technique employs a simple isolation scheme that requires only a standard microscope and a commercially available instrument, making it straightforward to translate the approach to other cell lines and proteins. In addition, since SiMoA is an ELISA based technique, there is no risk of bias arising from an amplification step, as is the case with techniques such as immuno-PCR.

To our knowledge, this study represents the first example of protein quantification in single cells using a single molecule counting technology that does not require genetic engineering, fluorescent labeling, or microchip assembly. Advantageously, SiMoA is also commercially available and facilitates high

throughput single cell analysis. To showcase the power of single cell SiMoA technology, we used both a high and low passage prostate cancer cell line of the same origin. The $LNCaP_A$ and $LNCaP_B$ cell lines exhibited a 12% genetic difference according to STR profiling and represent high and low PSA expressing traits, respectively. The variations in PSA content between the $LNCaP_A$ and $LNCaP_B$ cells were measured at the single cell level and the averages between these cell lines were found to differ by 30-fold. The substantial decline in PSA production in the $LNCaP_B$ cell line due to this genetic variation has significant implications for the need to standardize cell lines across scientific studies. Our work demonstrates the range and sensitivity of SiMoA with its capability to count low numbers of protein molecules within individual cells. We have demonstrated that SiMoA can be applied to study molecules that are not highly abundant or that are down regulated within single cells. Single molecule single cell measurements enable the measurement of phenotypic/genetic processes and cellular responses with unprecedented sensitivity and statistics.

Since many molecular pathways involve cascades of molecular events, monitoring multiple proteins simultaneously is necessary to gain a complete picture. Thus, multiplexed analysis over a wide dynamic range is particularly beneficial for single cell protein studies. Although beyond the scope of this work, SiMoA analysis can easily be multiplexed to investigate multiple proteins for cellular dynamics and correlation studies within individual cells. The SiMoA platform for single cell analysis can also be coupled with high throughput cell isolation techniques, such as ensemble-decision aliquot ranking (eDAR) [31, 32], to

concentrate and analyze rare cells, including circulating tumor cells (CTCs). Proteomic analysis of CTCs from clinical samples may be useful for early cancer detection, to identify the CTC phenotype and tissue of origin, and to guide therapy. In conclusion, the work described here represents a sensitive and robust system for the quantification of protein molecules in single cells using single molecule counting. This technique provides an important new tool for the field of single cell analysis.

5.5 Materials and Methods

Materials

The SiMoA HD-1 Analyzer, SiMoA consumables, and PSA assay kits (ref 100683) were purchased from Quanterix Corporation. The PSA assay kit contains magnetic PSA capture beads, biotinylated PSA detection antibody, streptavidin-β-galactosidase (SBG), and resorufin β-D-galactopyranoside (RGP). Free PSA antigen (J63000, 96% free, MW = 30 kDa) was purchased from BiosPacific and diluted in 1% BSA in 1x PBS for calibration standards.

Cell Culture and Isolation

$LNCaP_A$ cells were obtained from ATCC (CRL-1740). $LNCaP_B$ cells were generously donated by the Kuperwasser lab (Tufts University School of Medicine). All cells were cultured in RPMI-1640 medium (A10491-01, Life Technologies) with 10% fetal bovine serum (26140-079, Life Technologies).

Cultures were incubated at 37°C with 5% CO_2. Medium replacement was carried out two to three times per week in a SterilGARD III Advance biosafety cabinet (SG403, The Baker Company).

To isolate cells, culture plates were rinsed with 5 mL DPBS (14190-144, Life Technologies) and 3 mL of trypsin-EDTA (30-2101, ATCC) was added for 4-5 minutes then pipette mixed with 7 mL of complete growth medium to inhibit trypsin and suspend the cells. Cell suspension was centrifuged for 5 min at 130 g. Supernatant was aspirated, then the cell pellet was resuspended in complete growth medium. Cells were stained with Trypan Blue solution (15250-061, Life Technologies), washed 3x in DPBS, counted, and diluted to a concentration of ~2 x 10^3 cells/mL in DPBS. Cells were isolated by transferring 1 µL of the washed cell solution into the cap of a flat, optically clear PCR tube and validating by eye the presence of only one cell inside of the droplet using a microscope. Cells were isolated as quickly as possible following washing steps to reduce any secreted PSA in the bulk cell solution. Isolated cells were lysed and analyzed the same day or stored in PCR tubes at -20°C until use.

For bulk experiments, low cell number samples were prepared via serial dilutions. Cell suspensions were washed, counted with a hemocytometer, and diluted to concentrations of $1x10^4$, $5x10^3$, $2.5x10^3$, and $1x10^3$ cells/mL.

Single Molecule Array Analysis

Isolated single cells were lysed in PCR tubes with 64 µL Lysis Buffer 17 (895943, R&D Systems). For dilution-based bulk samples containing low

numbers of cells, 10 µL from each prepared concentration (corresponding to approximately 100, 50, 25, or 10 cells) was lysed with 55 µL lysis buffer. Lysates were then transferred to a 96-well PCR plate (10011-228, VWR) and diluted with 75 µL diluent (1% bovine serum albumin in 1x PBS). All samples and assay reagents (capture beads, detection antibody, SBG, and RGP) were loaded into the appropriate reagent bays in the HD-1 analyzer.

All single molecule measurements were taken using a HD-1 Analyzer (Quanterix Corp.). All HD-1 consumables, including wash buffers, sample diluent, assay discs, 96-well plates, sealing oil, cuvettes, and PSA reagents, were purchased from Quanterix Corp. The SiMoA platform has been described previously in the literature and in Chapter 2.[26, 33] The PSA SiMoA assay consists of three-steps: 1) 15 minute target incubation with capture beads 2) 5 minute incubation with detection antibody and 3) 5 minute incubation with SBG.

Since PSA is a secreted protein, the following controls were made to account for any PSA that may have secreted into the cell solution after the washing steps. Controls were made by preparing a cell solution of 2×10^3 cells/mL which was left at room temperature for 30 minutes, centrifuging the solution for 5 minutes at 150 rcf to pellet the cells, and removing 1 uL of the supernatant to analyze via SiMoA under the same conditions employed for cell samples. On average, controls for background PSA in the cell suspension supernatant yielded levels of 0.042 ± 0.026 pg/mL PSA for the $LNCaP_A$ cell line (n=16) and 0.005 ± 0.005 pg/mL PSA for the $LNCaP_B$ cell line (n=11).

5.6 References

1. Wu, M. & Singh, A.K. Single-cell protein analysis. *Current Opinion in Biotechnology* **23**, 83-88 (2012).
2. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183-1186 (2002).
3. Newman, J.R.S. et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-846 (2006).
4. Taniguchi, Y. et al. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329**, 533-538 (2010).
5. Irish, J.M., Kotecha, N. & Nolan, G.P. Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nat. Rev. Cancer* **6**, 146-155 (2006).
6. Zenobi, R. Single-Cell Metabolomics: Analytical and Biological Perspectives. *Science* **342** (2013).
7. Toriello, N.M. et al. Integrated microfluidic bioprocessor for single-cell gene expression analysis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20173-20178 (2008).
8. DiCarlo, D., Tse, H. & Gossett, D. in Single-Cell Analysis, Vol. 853. (eds. S. Lindström & H. Andersson-Svahn) 1-10 (Humana Press, 2012).
9. Klein, C.A. et al. Combined transcriptome and genome analysis of single micrometastatic cells. *Nat Biotech* **20**, 387-392 (2002).
10. Thompson, A.M. et al. Self-Digitization Microfluidic Chip for Absolute Quantification of mRNA in Single Cells. *Anal. Chem.* **86**, 12308-12314 (2014).
11. Bengtsson, M., Hemberg, M., Rorsman, P. & Ståhlberg, A. Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. *BMC Molecular Biol* **9**, 1-11 (2008).
12. Zhang, B. et al. Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382-387 (2014).
13. Huang, B. et al. Counting Low-Copy Number Proteins in a Single Cell. *Science* **315**, 81-84 (2007).
14. Cai, L., Friedman, N. & Xie, X.S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358-362 (2006).
15. Perez, O.D. & Nolan, G.P. Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. *Nat Biotech* **20**, 155-162 (2002).
16. De Rosa, S.C., Herzenberg, L.A., Herzenberg, L.A. & Roederer, M. 11-color, 13-parameter flow cytometry: Identification of human naive T cells by phenotype, function, and T-cell receptor diversity. *Nat Med* **7**, 245-248 (2001).
17. Perfetto, S.P., Chattopadhyay, P.K. & Roederer, M. Seventeen-colour flow cytometry: unravelling the immune system. *Nat. Rev. Immunol.* **4**, 648-655 (2004).
18. Kalisky, T. & Quake, S.R. Single-cell genomics. *Nat Meth* **8**, 311-314 (2011).
19. Bonner, W.A., Hulett, H.R., Sweet, R.G. & Herzenberg, L.A. Fluorescence Activated Cell Sorting. *Rev. Sci. Instrum.* **43**, 404-409 (1972).
20. Hughes, A.J. et al. Single-cell western blotting. *Nat Meth* **11**, 749-755 (2014).
21. Hughes, A.J. & Herr, A.E. Microfluidic Western blotting. *PNAS* **109**, 21450-21455 (2012).
22. Burgin, E. et al. Absolute quantification of protein copy number using a single-molecule-sensitive microarray. *Analyst* (2014).

23. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X.S. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. *Science* **311**, 1600-1603 (2006).
24. Salehi-Reyhani, A., Burgin, E., Ces, O., Willison, K.R. & Klug, D.R. Addressable droplet microarrays for single cell protein analysis. *Analyst* (2014).
25. Wu, H., Wheeler, A. & Zare, R.N. Chemical cytometry on a picoliter-scale integrated microfluidic chip. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12809-12813 (2004).
26. Rissin, D.M. et al. Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nat. Biotechnol.* **28**, 595-599 (2010).
27. Rissin, D.M. et al. Simultaneous Detection of Single Molecules and Singulated Ensembles of Molecules Enables Immunoassays with Broad Dynamic Range. *Anal. Chem.* **83**, 2279-2285 (2011).
28. Rissin, D.M. et al. Multiplexed single molecule immunoassays. *Lab on a Chip* **13**, 2902-2911 (2013).
29. Esquenet, M., Swinnen, J.V., Heyns, W. & Verhoeven, G. LNCaP prostatic adenocarcinoma cells derived from low and high passage numbers display divergent responses not only to androgens but also to retinoids. *The Journal of Steroid Biochemistry and Molecular Biology* **62**, 391-399 (1997).
30. Pinzani, P. et al. Prostate-specific antigen mRNA and protein levels in laser microdissected cells of human prostate measured by real-time reverse transcriptase-quantitative polymerase chain reaction and immuno-quantitative polymerase chain reaction. *Human Pathology* **39**, 1474-1482 (2008).
31. Schiro, P.G. et al. Sensitive and High-Throughput Isolation of Rare Cells from Peripheral Blood with Ensemble-Decision Aliquot Ranking. *Angewandte Chemie International Edition* **51**, 4618-4622 (2012).
32. Zhao, M. et al. New Generation of Ensemble-Decision Aliquot Ranking Based on Simplified Microfluidic Components for Large-Capacity Trapping of Circulating Tumor Cells. *Anal. Chem.* (2013).
33. Kan, C.W. et al. Isolation and detection of single molecules on paramagnetic beads using sequential fluid flows in microfabricated polymer array assemblies. *Lab on a Chip* **12**, 977-985 (2012).

Chapter 6

Towards The Development of a Breast Cancer Mouse Model for Early Disease Monitoring

6.1 Introduction

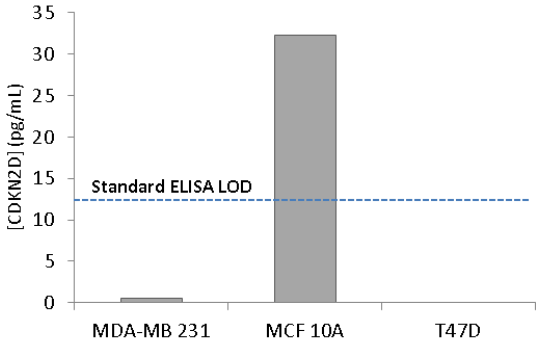
Chapter 3 discussed the development of a mouse model using PSA to track tumor progression in an LNCaP mouse model. Chapter 4 discussed the development of SiMoA assays for ERα, PR, and CDKN2D and their applications to developing a biomarker panel to distinguish healthy serum from breast cancer serum samples. Prior to testing the described biomarkers in human samples, an attempt was made to construct a mouse model to enable the tracking of breast cancer biomarkers in serum with disease progression. The development of mouse models helps gain valuable insight into the usefulness of biomarkers prior to their implementation in human studies, where samples are often limited. This Chapter discusses the use of a mouse model to ascertain the utility of CDKN2D for the early detection of breast cancer.

6.2 Cell Line Determination

The first step in developing an appropriate breast cancer mouse xenograft model involves determining the appropriate breast cancer cell line to use for the study. This choice can typically be determined from the literature and by knowing what subtype of breast cancer typically overexpresses the biomarker of interest. In addition, to see if cells actively secrete the biomarker in question, cell media retrieved from active cell culture can also be measured. If the media contains the protein of interest, in this case CDKN2D, it may also be found in serum samples from mice inoculated with particular cell lines and will therefore give insight as to which cell line to use for further developing a breast cancer mouse model.

Cell media used in the active culture of two different breast cancer cell lines (MDA-MB 231 and T47D) and one non-tumorigenic breast cell line (MCF 10A) were obtained. Two cell lines were ER-/PR- (MDA-MB 231 and MCF 10A) while one was ER+/PR+ (T47D)[1]. The results of the cell culture media test are shown below in **Figure 6.1.** As shown in the figure, the levels of MCF 10A were well above the 10 pg/mL LOD of standard ELISA (Antibodies-online, ABIN812390), while the concentration of MDA-MB 231 was below the ELISA LOD, at only 0.582 pg/mL. Since MCF10A is non-tumorigenic, only MDA-MB231 was used for further development of a potential mouse model.

a.



b.

Cell culture media	[CDKN2D] (pg/mL)	n
MDA-MB 231	0.582	1
MCF 10A	32.3	1
T47D	0	1

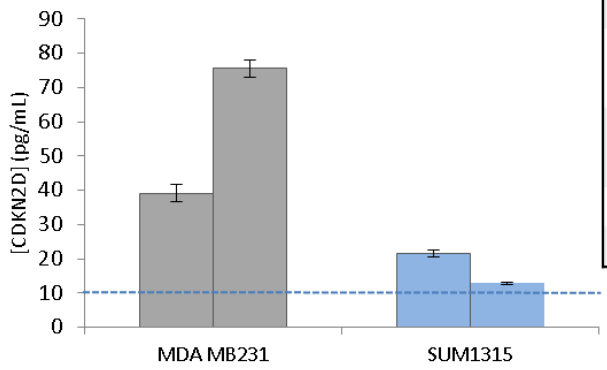
Figure 6.1 CDKN2D in breast cancer cell culture media. **a)** Bar graph of measured CDKN2D concentrations in MDA-MB 231, MCF 10A and T47D cell culture media. The LOD for standard ELISA is 10 pg/mL and is indicated on the graph. The assay LOD for CDKN2D was 0.16 pg/mL. **b)** Table showing concentrations of CDKN2D in each sample.

Once it was determined that MDA-MB 231 cells would likely be useful for a cell model, serum samples from a previously established mouse model trial were obtained. These mouse models were constructed using both MDA-MB 231 and SUM1315 cells. SUM1315, like MDA-MB 231, is also ER-/PR- and was

thought to also be a potential candidate cell line. No cell culture media was available from this cell line for the initial test.

In the models used, 1×10^6 SUM 1315 and MDA MB231 cells had been injected into both 4th mammary glands of 8 week old female NOD/SCID mice for a total inoculation of 2×10^6 cells. Blood was collected from mice on a weekly basis starting two weeks post inoculation and continued until tumors in each mouse reached 1 cm in diameter (approximately 16 weeks). The first samples tested for CDKN2D were the samples collected after 16 weeks, which were endpoint samples. The rationale for this approach is that the endpoint samples would theoretically contain the highest concentration of CDKN2D. If CDKN2D was not detected in these samples, then the remaining time course samples could be used for testing other biomarkers. SiMoA analysis of CDKN2D in endpoint serum of two mice for both MDA MB231 and SUM1315 is shown in **Figure 6.2**. As shown in the bar graph, the measured concentrations of CDKN2D in MDA MB231 inoculated mice were significantly higher than the measured values of SUM1315 inoculated mice and both were above the LOD of standard ELISA. For this reason, only the time course samples from the SUM1315 mouse model were used in further experiments.

a.



b.

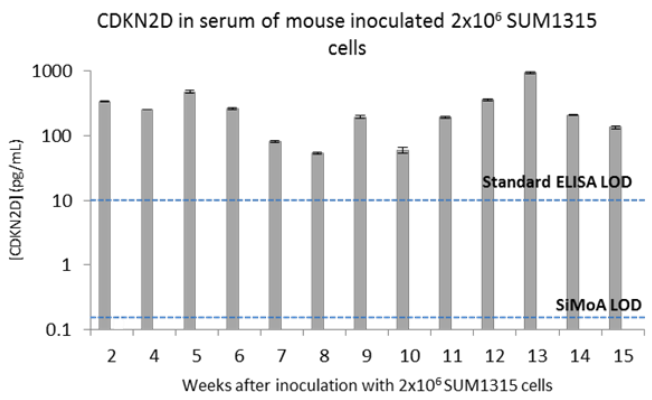
Cell line	[CDKN2D] (pg/mL)	n
MDA MB231	39.2±2.9	3
MDA MB231	75.6±2.6	3
SUM1315	21.7±1.1	2
SUM1315	12.9±0.2	3

Figure 6.2 CDKN2D in MDA MB231 and SUM1315 mouse serum. **a)** Bar graph of measured CDKN2D concentrations in MDA-MB 231 and SUM1315 endpoint serum in two mice per cell line. The assay LOD for CDKN2D was 0.16 pg/mL. **b)** All values were well above the LOD for standard ELISA, which is 10 pg/mL and is indicated on the graph. Table showing concentrations of CDKN2D in each sample.

6.3 SUM1315 Mouse Model

Figure 6.3 shows results for CDKN2D in the serum of one mouse measured each week from 2-15 weeks after inoculation with 2×10^6 SUM1315 cells. As shown in the graph, all values of CDKN2D were well above the calculated SiMoA assay LOD and were all also well above the lowest commercially available CDKN2D ELISA LOD. Despite the fact that this mouse grew a tumor 1 cm in size, no increasing trend in CDKN2D is noticed in the serum over the course of 15 weeks.

a.



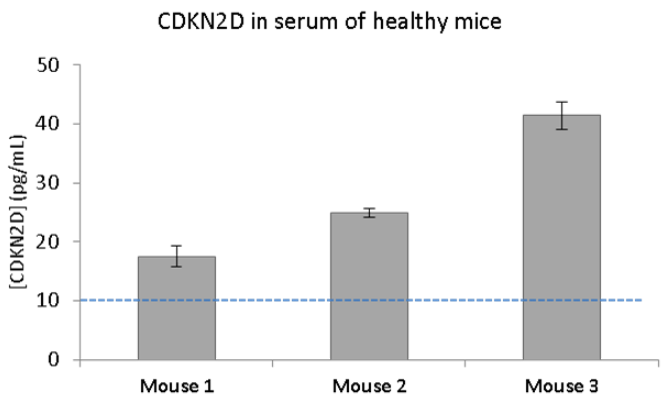
b.

Weeks after inoculation	[CDKN2D] (pg/mL)
2	337±7
4	253±3
5	479±20
6	260±7
7	81.3±2.9
8	53.6±1.1
9	195±9
10	59.5±6.1
11	194±4
12	361±11
13	953±28
14	209±4
15	134±6

Figure 6.3 CDKN2D in SUM1315 mouse model serum. **a)** Bar graph of measured CDKN2D concentrations in one mouse inoculated with 2×10^6 SUM1315 c. The assay LOD for CDKN2D was 0.16 pg/mL. All values were well above the LOD for standard ELISA, which is 10 pg/mL and is indicated on the graph. **b)** Table showing the concentration of CDKN2D in each sample. Error represents the standard deviations between triplicate measurements.

Despite the lack of a correlation in CDKN2D concentration with increased tumor growth, there is still an overall increase in CDKN2D in the tested mouse compared to healthy mice. The serum of three healthy female mice was collected and measured for the presence of CDKN2D as a negative control. **Figure 6.4** shows the results of this screening. While the overall levels are lower than what was shown in the mouse with tumor growth, they were still above the LOD of traditional ELISA.

a.



b.

Weeks after inoculation	[CDKN2D] (pg/mL)
2	337±7
4	253±3
5	479±20
6	260±7
7	81.3±2.9
8	53.6±1.1
9	195±9
10	59.5±6.1
11	194±4
12	361±11
13	953±28
14	209±4
15	134±6

Figure 6.4 CDKN2D in healthy mouse serum. **a)** Bar graph of measured CDKN2D concentrations in three healthy female mice. All values were well above the LOD for standard ELISA, which is 10 pg/mL and is indicated on the graph. **b)** Table showing the concentration of CDKN2D in each sample. Standard deviations represent the variation between triplicate measurements.

Since there was no marked increase in CDKN2D in the mouse studied over time and the healthy mice had levels of CDKN2D that were well above the limit of both SiMoA and traditional ELISA, no further attempts were made to use SUM1315 in a mouse xenograft model to study CDKN2D. The remaining mice from the time course trial can be used to study biomarkers for future studies, if necessary. Since one of the antibodies used in the SiMoA was a mouse monoclonal antibody, it is very likely that there was simply cross reactivity between the human CDKN2D protein produced from the tumor and the mouse CDKN2D present in the host. This cross reactivity is a likely reason for the high background observed in the healthy mice and could also mask the response from

the mouse with the tumor. Further studies investigating this biomarker were used in human serum, as described in Chapter 4.

6.4 Materials and Methods

Cell Culture

Conditioned cell media was provided by Dr. Lisa Arendt in the Kuperwasser lab at Tufts Medical School. Cell media for the following cell lines were as follows: MDA-MB 231, DMEM+/- with 10% FBS; MCF10A, BME; T47D, DMEM+/- with 10% FBS. All media and serum components were purchased from Gibco. Media was removed from active cell culture during feeding, placed in a 15 mL centrifuge tube, and stored at -80°C until used. Media was not diluted prior to SiMoA analysis.

Mouse Xenografts

The human breast cancer cell line SUM1315 was generously donated from Dr. Stephen Ethier (Kramanos Institute, MI) and cultured in the laboratory of Dr. Charlotte Kuperwasser at Tufts Medical School. Cells were grown in Ham's F12 media with 5% calf serum, 5 μg/ml insulin, and 10 ng/ml epidermal growth factor (all from Gibco) at 37°C with 5% CO_2. NOD/SCID mice (Jackson Laboratories) were anesthetized with isofluorane, and 30 ul of $1x10^6$ SUM1315 cells were injected into the inguinal mammary glands in a 1:3 ratio of Matrigel (BD Biosciences) to media. Approximately 100 μl of blood was drawn from the submandibular vein and collected into SST Microtainer tubes (Becton Dixon and

Co, Franklin Lakes NJ) weekly until tumors reached 1 cm in diameter. At end stage, mice were anesthetized isofluorane and blood was drawn via cardiac puncture followed by euthanasia by $C0_2$ asphyxiation. The care of animals and all animal procedures were conducted in accordance with a protocol approved by the Tufts University Institutional Animal Care and Use Committee (IACUC).

Whole blood samples were allowed to clot on ice for 10 min. followed by centrifugation at 1,500 × g for 10 min at 4°C. Serum was then removed and immediately snap frozen. All serum samples were stored at -80°C prior to use. For SiMoA analysis, serum samples were diluted by a factor of four in 5 mM EDTA/PBS with 10 μg/mL TruBlock (Meridian Life Science Inc.) before being loaded onto an automated HD-1 analyzer (Quanterix Corp.).

SiMoA CDKN2D Assay

All single molecule measurements were taken using a HD-1 Analyzer (Quanterix Corp.). All HD-1 consumables, including wash buffers, assay discs, 96-well plates, sealing oil, and cuvettes were purchased from Quanterix Corp. CDKN2D capture and detection antibodies were purchased from LifeSpan Biosciences Inc. and Abnova, respectively. The beads and detector were coupled and biotinylated as previously described in Chapter 4. Calibrators were diluted in PBS containing 25% newborn calf serum (Life Technologies), 5 mM EDTA (Sigma Aldrich), 0.15% ProClin 300 (Sigma Aldrich) and 0.01% Tween-20 (Sigma Aldrich). The detector was diluted to 1 μg/mL in a 1% solution of BSA in

PBS. Streptavidin β-galactosidase (Quanterix) was diluted to a concentration of 500 pM in SBG buffer (Quanterix).

6.5 References

1. Neve, R.M. et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515-527 (2006).

Appendix

This appendix contains supplemental data and information pertaining specifically to Chapter 4. Table A1 contains information regarding all samples used in Chapter 4.

Table A1 Sample ID, stage, patient age, and molecular subtype for all samples used within Chapter 4.

Sample ID	Stage	Age
BRH813984	Healthy	22
BRH813991	Healthy	25
BRH813992	Healthy	32
BRH813985	Healthy	33
BRH813986	Healthy	36
BRH813996	Healthy	38
BRH813993	Healthy	42
BRH813988	Healthy	43
BRH813994	Healthy	45
BRH813995	Healthy	45
BRH813990	Healthy	46
BRH813997	Healthy	47
BRH943393	Healthy	47
BRH813998	Healthy	50
BRH943394	Healthy	50
BRH774364	Healthy	51
BRH943396	Healthy	51
BRH813989	Healthy	52
BRH888416	Healthy	52
BRH943389	Healthy	52
BRH943391	Healthy	52
BRH943397	Healthy	52
BRH943386	Healthy	53
BRH888415	Healthy	54

Sample ID	Stage	Subtype	Age
BRH824420	1	Luminal	36
BRH922995	1	Luminal	38
BRH824412	1	Luminal	52
BRH853583	1	TNBC	53
BRH824417	1	Luminal	65
BRH885851	1	Luminal	69
BRH885852	1	Luminal	70
BRH824415	1	Luminal	71
BRH922993	1	Luminal	71
BRH824413	1	Luminal	77
BRH921366	1	TNBC	82
BRH922994	2	Luminal	43
BRH824419	2	Luminal	44
BRH921364	2	TNBC	49
BRH922986	2	Luminal	51
BRH922989	2	Luminal	51
BRH921371	2	TNBC	54
BRH824418	2	Luminal	56
BRH921360	2	TNBC	58
BRH921363	2	TNBC	59
BRH921357	2	Luminal	61
BRH921367	2	TNBC	62
BRH921370	2	TNBC	63
BRH921368	2	TNBC	67

BRH943383	Healthy	54
BRH943395	Healthy	56
BRH888417	Healthy	57
BRH943388	Healthy	57
BRH943385	Healthy	58
BRH943387	Healthy	58
BRH943390	Healthy	58
BRH943392	Healthy	58
BRH888414	Healthy	70
BRH888412	Healthy	72
BRH943384	Healthy	72
BRH888413	Healthy	73

BRH921369	2	TNBC	69
BRH922998	2	Luminal	69
BRH921359	2	TNBC	74
BRH921361	2	TNBC	74
BRH824416	2	Luminal	79
BRH921362	2	TNBC	79
BRH824414	2	Luminal	84
BRH922991	3	Luminal	47
BRH922997	3	Luminal	47
BRH885847	3	TNBC	53
BRH885848	3	TNBC	53
BRH921372	3	TNBC	54
BRH922988	3	Luminal	55
BRH922992	3	Luminal	64
BRH922990	3	Luminal	68
BRH853584	3	TNBC	70
BRH921358	3	TNBC	73
BRH845596	4	Luminal	39
BRH885844	4	HER2	40
BRH885846	4	HER2	56
BRH845595	4	Luminal	58
BRH921365	4	TNBC	68
BRH885854	4	unknown	71
BRH845593	4	Luminal	71
BRH845599	4	Luminal	72
BRH885853	4	Luminal	75
BRH845592	4	Luminal	75
BRH845597	4	Luminal	75
BRH845598	4	Luminal	75
BRH922987	4	Luminal	68

In the final PLS-DA models described in Chapter 4, age was used as an input variable along with ERα, PR, and CDKN2D . Two PLS-DA models were also made using both healthy vs. all breast cancer samples and healthy vs. stage I-II breast cancer samples without using age. These models are shown in **Figure A1**. The AUC for healthy vs. stage I-IV (**Figure A1 b-c**) was 0.89 and the overall accuracy for this model was 80%. The AUC for healthy vs. stage I-II was 0.90 and the overall accuracy was 84%. These results are very similar to the results described in

Chapter 4 where age was not used. The extra variable was used since it is well established that age is associated with breast cancer risk and it is an easy variable to add into a multivariate technique to potentially gain additional discriminatory power. In this situation, the addition of age to the model only made a very marginal difference.

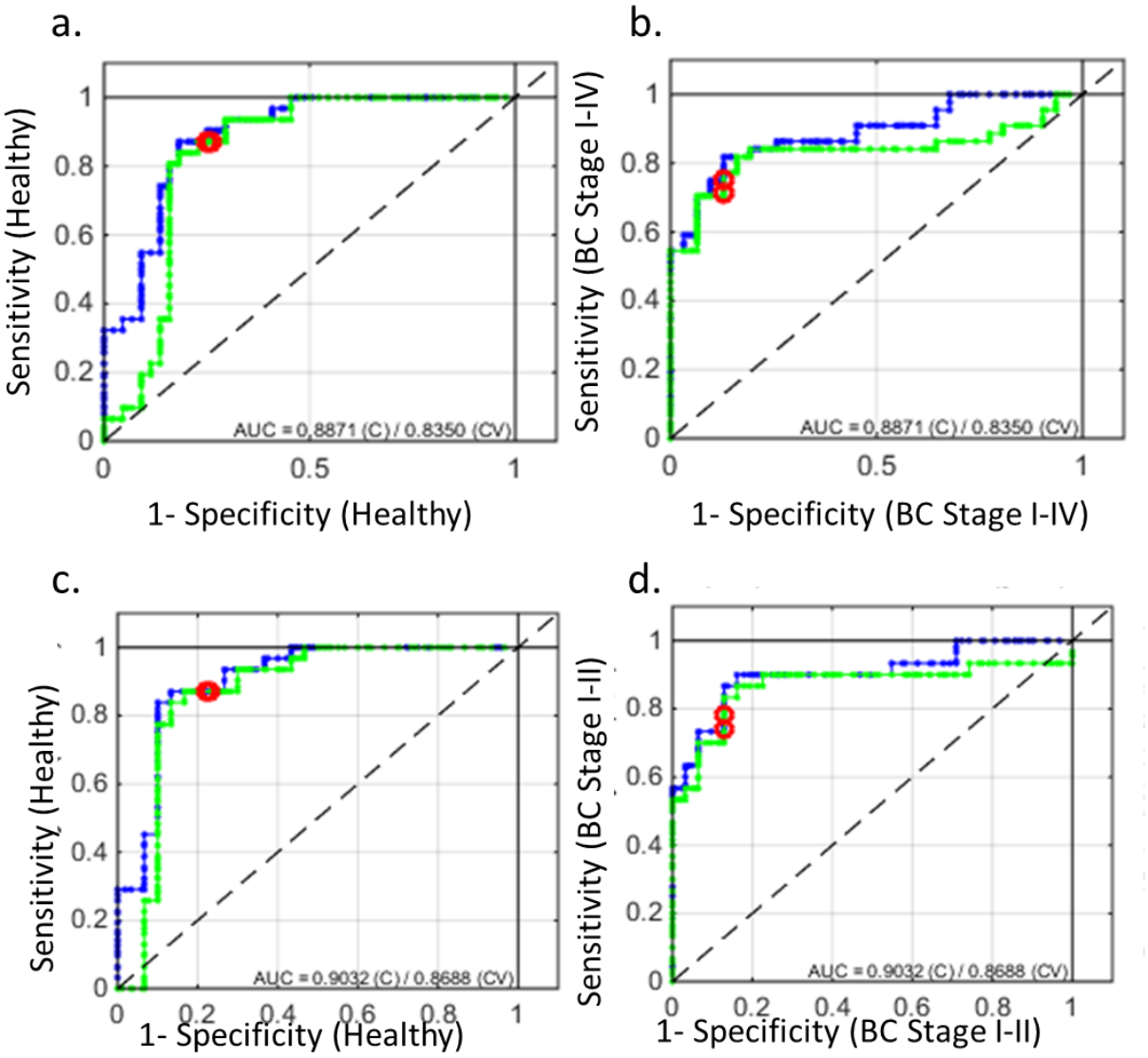


Figure A1 PLS-DA models without age. Two models are shown, with ROC curves from PLS-DA analysis of ERα, PR, and CDKN2D without age in **a)** healthy vs. **b)** stage I-IV breast cancer and **c)** healthy vs. **d)** stage I-IV breast cancer.

All samples were diluted 1:4 in sample diluent and were run on the HD-1. Although some samples had concentrations of tested biomarker that were above the reported commercial ELISA LOD, many were below the ELISA LOD when diluted. Diluting samples is convenient for banked precious samples that are limited in volume. The actual concentration of each biomarker within the samples was also calculated based on the dilution factor. A significant number of samples remained below the LOD of standard ELISA for both ERα and PR, but were detected by SiMoA. In the case of CDKN2D, fewer samples were below the LOD of ELISA; however, the ELISA LOD reported is from a kit that is no longer available and thus essentially only SiMoA is able to detect all of these samples. **Figure A2** shows the response from all data points for each biomarker with the respective ELISA and SiMoA LODs.

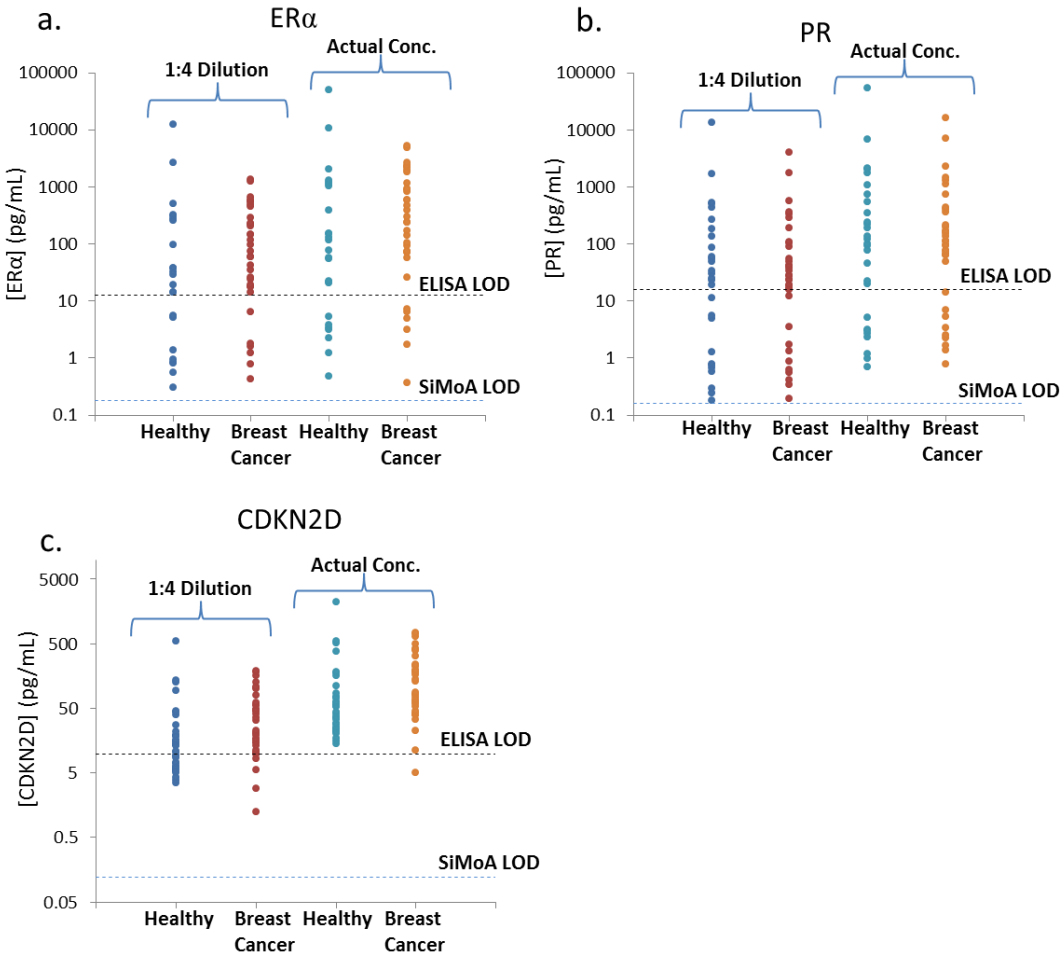


Figure A2 Response of all healthy and breast cancer samples for ERα, PR and CDKN2D. Samples were diluted 1:4 in sample diluent and both the raw values and the actual concentrations with the dilution factor taken into consideration are shown. SiMoA and ELISA LODs are indicated on the dotted lines.