Early Detection and Treatment of Breast Cancer
by Random Peptide Array in neuN Transgenic Mouse Model

by

Hu Duan

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

Approved June 2015 by the:
Graduate Supervisory Committee:

Stephen Albert Johnston, Chair
Leland Hartwell
Valentin Dinu
Yung Chang

ARIZONA STATE UNIVERSITY
August 2015
ABSTRACT

Breast cancer is the most common cancer and currently the second leading cause of death among women in the United States. Patients’ five-year relative survival rate decreases from 99% to 25% when breast cancer is diagnosed late. Immune checkpoint blockage has shown to be a promising therapy to improve patients’ outcome in many other cancers. However, due to the lack of early diagnosis, the treatment is normally given in the later stages. An early diagnosis system for breast cancer could potentially revolutionize current treatment strategies, improve patients’ outcomes and even eradicate the disease. The current breast cancer diagnostic methods cannot meet this demand. A simple, effective, noninvasive and inexpensive early diagnostic technology is needed. Immunosignature technology leverages the power of the immune system to find cancer early. Antibodies targeting tumor antigens in the blood are probed on a high-throughput random peptide array and generate a specific binding pattern called the immunosignature.

In this dissertation, I propose a scenario for using immunosignature technology to detect breast cancer early and to implement an early treatment strategy by using the PD-L1 immune checkpoint inhibitor. I develop a methodology to describe the early diagnosis and treatment of breast cancer in a FVB/N neuN breast cancer mouse model. By comparing FVB/N neuN transgenic mice and age-matched wild type controls, I have found and validated specific immunosignatures at multiple time points before tumors are palpable. Immunosignatures change along with tumor development. Using a late-stage
immunosignature to predict early samples, or vice versa, cannot achieve high
prediction performance. By using the immunosignature of early breast cancer, I
show that at the time of diagnosis, early treatment with the checkpoint blockade,
anti-PD-L1, inhibits tumor growth in FVB/N neuN transgenic mouse model. The
mRNA analysis of the PD-L1 level in mice mammary glands suggests that it is
more effective to have treatment early.

Novel discoveries are changing understanding of breast cancer and
improving strategies in clinical treatment. Researchers and healthcare
professionals are actively working in the early diagnosis and early treatment fields.
This dissertation provides a step along the road for better diagnosis and treatment
of breast cancer.
DEDICATION

I dedicate this dissertation to my wife Xinyao for her unconditional love and support, and my parents who accompany and help me during these days.

And

To my son Roger who is the best gift for me before my graduation.
ACKNOWLEDGEMENTS

Six years of my PhD studies have led to this dissertation. I have experienced a lot during these years, and many people have influenced me positively. I want to offer my most sincere thanks to all the people who helped make me who I have become during these six years.

Special thanks to my advisor, Dr. Stephen Albert Johnston. Dr. Johnston has inspired me and patiently taught me the most during my PhD studies. When I faced a complex problem, he always inspired and guided me to break it into small steps and push my thoughts further. One of his favorite expressions is “Keep it simple, stupid”—the KISS principle—I cannot tell express how much I benefited from this advice. I also learned a very important way to make decisions when the choices are controversial. Dr. Johnston fully supported me in my dissertation project through both his efforts to prepare me intellectually and his willingness to offer me specific scientific advice. Dr. Leland Hartwell has been more than a committee member for me. Besides his insightful comments on my dissertation, what I learned most came from his spirit. His constant eagerness to learn knowledge in a new field, his integrity and his modesty set an excellent example for me to follow. I have also consulted with him on my career development and shared many thoughts about my life and dreams. Dr. Hartwell always encouraged and advised me, which made me more confident in pursuing my dream. Every time I talked with Dr. Valentin Dinu, he provided me with new knowledge on bioinformatics analysis. Although Dr. Dinu was mainly at the Mayo Clinic, and it was a long way for him to come to Biodesign, he was always very willing to meet
me with great patience and kindness. I learned immunology from Dr. Yung Chang when I took her classes in 2010. I would like to thank her for sharing her knowledge of immunology and for supporting my dissertation project.

I also want to thank Dr. Kathy Sykes, Dr. Sid Hecht and Dr. Tim Karr, who were on my comprehensive exam committee. They helped and supported me in becoming a PhD candidate. During my PhD studies, it was a pleasure to work with the people in the Center for Innovations in Medicine in the Biodesign Institute at ASU. I am grateful to be a member of the Biological Design Graduate Program, which provides funding to support me. I thank Dr. Joann Williams, Maria Hanlin, Laura Hawes and Lauren Dempsey, who guided me in each step toward graduation and provided managerial support for my PhD program.

With regard to my dissertation studies, I would like to specially thank the following individuals. Dr. Luhui Shen was the major collaborator in my dissertation project. Dr. Bart Legutki was my first guide when I did my rotation in CIM back in 2009. John Lainson taught me a lot about microarray production and automatic robot systems, and we had many amusing talks that made life at CIM more interesting. I shared an office with Dr. Josh Richer for two years, and we spent quite a lot time together discussing each other’s experiment. The many arrays that were run in the past six years would not have been possible without the support of Dr. Zbigniew Cichacz and the Peptide Array Core team.

I would like to give special thanks to my friends Andy Ge, Jie Wang, Xiaofang Bian and Xiaobo Yu. They provided me with tremendous emotional support and have encouraged me all along the way.
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. xii

LIST OF FIGURES ............................................................................................................... xiii

CHAPTER

1 INTRODUCTION ............................................................................................................. 1

1.1 Breast Cancer .............................................................................................................. 1
  1.1.1 Overview of Cancer .............................................................................................. 1
  1.1.2 Overview of Breast Cancer .................................................................................. 2
  1.1.3 Overview of Transgenic Mouse Model ................................................................. 4

1.2 Current Breast Cancer Diagnosis and Biomarkers ................................................. 6
  1.2.1 Statistics Preparation for Diagnosis and Biomarker ........................................... 6
  1.2.2 Current Clinical Diagnostic Methods of Breast Cancer .................................... 7
    1.2.2.1 Breast Self-Examination (BSE) ...................................................................... 7
    1.2.2.2 Clinical Breast Examination (CBE) ............................................................... 8
    1.2.2.3 Mammography .............................................................................................. 8
    1.2.2.4 Adjunctive Testing: Ultrasound Imaging and MRI ....................................... 10
    1.2.2.5 Molecular Diagnosis Platform ..................................................................... 11
  1.2.3 Benefit of Early Detection and Treatment ......................................................... 12
    1.2.3.1 Early Diagnosis of Breast Cancer Improve Survival Rate 13
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2.3.2</td>
<td>Proposed Early Diagnosis and Treatment Scenario</td>
<td>14</td>
</tr>
<tr>
<td>1.3</td>
<td>Early Detection of Breast Cancer by Antibodies</td>
<td>15</td>
</tr>
<tr>
<td>1.3.1</td>
<td>Overview of Blood Based Breast Cancer Biomarkers</td>
<td>15</td>
</tr>
<tr>
<td>1.3.2</td>
<td>Challenges in Finding Early Breast Cancer Biomarker</td>
<td>17</td>
</tr>
<tr>
<td>1.3.3</td>
<td>Cancer Antibodies as Biomarkers</td>
<td>19</td>
</tr>
<tr>
<td>1.3.3.1</td>
<td>Patients can Develop Autoantibodies Against Tumor Early</td>
<td>19</td>
</tr>
<tr>
<td>1.3.3.2</td>
<td>Advantage of Using Antibodies as Biomarker</td>
<td>21</td>
</tr>
<tr>
<td>1.3.3.3</td>
<td>Breast Cancer Associated Antibodies</td>
<td>24</td>
</tr>
<tr>
<td>1.3.4</td>
<td>Emerging Technologies for Profiling Antibody Responses</td>
<td>25</td>
</tr>
<tr>
<td>1.3.4.1</td>
<td>SEREX</td>
<td>25</td>
</tr>
<tr>
<td>1.3.4.2</td>
<td>Peptide and Peptoid Microarrays</td>
<td>27</td>
</tr>
<tr>
<td>1.3.4.3</td>
<td>SERPA</td>
<td>28</td>
</tr>
<tr>
<td>1.3.4.4</td>
<td>Protein microarray</td>
<td>29</td>
</tr>
<tr>
<td>1.4</td>
<td>Immunosignature</td>
<td>31</td>
</tr>
<tr>
<td>1.4.1</td>
<td>Definition of Immunosignature</td>
<td>31</td>
</tr>
<tr>
<td>1.4.1.1</td>
<td>Building a Complex Chemical Surface</td>
<td>32</td>
</tr>
<tr>
<td>1.4.1.2</td>
<td>Probing Sera</td>
<td>33</td>
</tr>
<tr>
<td>1.4.1.3</td>
<td>Analysis of Binding Pattern</td>
<td>33</td>
</tr>
<tr>
<td>1.4.2</td>
<td>Advantage of Immunosignature</td>
<td>34</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>1.4.3 Published Studies with Immunosignature</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>1.5 Breast Cancer Treatment</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>1.5.1 Clinical Treatment of Breast Cancer</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>1.5.1.1 Surgery and Radiotherapy</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>1.5.1.2 Cancer Drugs</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>1.5.1.2.1 Chemotherapy</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>1.5.1.2.2 Hormone Therapy</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>1.5.1.2.3 Targeted Therapy</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>1.5.2 Emerging Strategy of Immune Checkpoint Blockade</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>2 Early Detection Of Breast Cancer</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>2.2 Results</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>2.2.1 HER2 Transgenic Mouse Model</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>2.2.2 Scheme of Experiment Design</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>2.2.3 Influence of Genetic Background on Immunosignature</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>2.2.4 Pre-tumor Immunosignature Prior to Tumor Palpation</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>2.2.5 Earliest Time Point of the Detectable Immunosignature</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>2.2.6 Time Course of Immunosignatures During Tumor Progression</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>2.2.7 Cross-stage Prediction</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>CHAPTER</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>2.3 Methods</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>2.3.1 Mice and Sample Collection</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>2.3.2 Preparation of Microarray</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>2.3.3 Probing Serum Antibodies on Peptide Microarrays</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>2.3.4 Microarray Data Processing</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>2.3.5 Statistical Analysis</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>2.4 Discussion</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>2.5 Supplements</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>2.5.1 Selection of multiple batches of array</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>2.5.2 Removal of Batch Effect</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>3 A MODEL FOR THE EARLY DETECTION AND TREATMENT OF CANCER</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>3.1 Abstract</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>3.3 Results:</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>3.3.1 Early Detection of Breast Cancer and PD-L1 Treatment</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>3.3.2 Inhibition of Tumor Growth Rate by Early Treatment</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>3.3.3 RNA Expression of PD-L1 in Mammary Glands</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>3.3.4 Changes of Immunosignatures After PD-L1 Treatment</td>
<td>82</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 3.4 Methods ................................................................. 84

3.4.1 Mice and Sample Collection ............................................. 84

3.4.2 Binding of Sera to the Peptide Microarrays ....................... 84

3.4.3 PD-L1 Treatment Regime ................................................. 85

3.4.4 RNA Expression in Mammary Gland .............................. 86

3.4.5 Statistical Analysis ....................................................... 87

3.5 Discussion ........................................................................ 87

REFERENCES .......................................................................... 93

APPENDIX

A DETECTION OF OSTEOSARCOMA RECURRENCE IN CLINICAL CANINE SAMPLES BY IMMUNOSIGNATURE ...................... 118

AP1.1 Introduction .................................................................... 119

AP1.1.1 Background of Osteosarcoma ...................................... 119

AP1.1.2 Current Diagnosis and Treatment of OSA: ..................... 120

AP1.1.3 Recurrence of OSA ..................................................... 120

AP1.2 Result and Discussion .................................................. 121

AP1.2.1 Canine Osteosarcoma Clinical Sample Summary .......... 121

AP1.2.2 Correlation Analysis Of Overall Antibody During OSA .... 122

AP1.2.3 Comparison of Immunosignature at Multiple Time Points ... 124
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP1.2.4 Conclusion and Future Direction</td>
<td>127</td>
</tr>
<tr>
<td>AP1.3 Methods</td>
<td>128</td>
</tr>
<tr>
<td>AP1.3.1 Sample Summary</td>
<td>128</td>
</tr>
<tr>
<td>AP1.3.2 Binding of Sera to the Peptide Microarrays</td>
<td>129</td>
</tr>
<tr>
<td>AP1.3.3 Statistical Analysis</td>
<td>129</td>
</tr>
<tr>
<td>B CHARACTERIZATION OF ANTIBODY PROFILE DURING CANCER DEVELOPMENT</td>
<td>131</td>
</tr>
<tr>
<td>AP 2.1 Quantitative Characterization Of Antibody Profiles During Tumor Development</td>
<td>132</td>
</tr>
<tr>
<td>AP2.1.1 Introduction</td>
<td>132</td>
</tr>
<tr>
<td>AP2.1.2: Result and Discussion</td>
<td>133</td>
</tr>
<tr>
<td>AP2.2 A Time Series Study of Antibody Profile Along Tumor Development</td>
<td>138</td>
</tr>
<tr>
<td>AP2.2.1 Introduction</td>
<td>138</td>
</tr>
<tr>
<td>AP2.2.2 Result and Discussion</td>
<td>139</td>
</tr>
<tr>
<td>AP2.3 Robustness of immunosignature</td>
<td>146</td>
</tr>
<tr>
<td>AP2.3.1 Introduction</td>
<td>146</td>
</tr>
<tr>
<td>AP2.3.2 Result and discussion</td>
<td>147</td>
</tr>
<tr>
<td>C PERMISSIONS</td>
<td>150</td>
</tr>
<tr>
<td>BIO SKETCH</td>
<td>152</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1: Cross-Stage Prediction Performance in the Training Set</td>
<td>60</td>
</tr>
<tr>
<td>2.2: Correlation Analysis for Different Printing Batches of Slides</td>
<td>68</td>
</tr>
<tr>
<td>2.3: Statistic Characterization of Different Batches of Slides</td>
<td>69</td>
</tr>
<tr>
<td>3.1: The Timeframe for Occurrence of Signatures, Palpable Tumor and</td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>78</td>
</tr>
<tr>
<td>AP1.1: Canine Osteosarcoma Clinical Sample Summary</td>
<td>121</td>
</tr>
<tr>
<td>AP1.2: Correlation Analysis of Overall Antibody Profile</td>
<td>123</td>
</tr>
<tr>
<td>AP1.3: Distribution of Fold Change in Selected Peptides</td>
<td>123</td>
</tr>
<tr>
<td>AP1.4: Proportion of Overlapped Peptides</td>
<td>126</td>
</tr>
<tr>
<td>AP2.1: Statistic characterization of All Time Points for Each Mouse</td>
<td>136</td>
</tr>
<tr>
<td>AP2.2: Sample Information for Tg8-1-1 and WT2-3</td>
<td>140</td>
</tr>
<tr>
<td>AP2.3: Sample Information for Tg8-3-9 and WT1-3</td>
<td>141</td>
</tr>
<tr>
<td>AP2.4: Sample Information for Robustness Test</td>
<td>147</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Biology of Breast Tissue</td>
<td>2</td>
</tr>
<tr>
<td>1.2: Illustration of Mammogram and MRI</td>
<td>7</td>
</tr>
<tr>
<td>1.3: Five-Year Relative Survival Rates (%) by Stage in US</td>
<td>13</td>
</tr>
<tr>
<td>1.4: Distribution of Human Plasma Protein</td>
<td>22</td>
</tr>
<tr>
<td>1.5: Principle of Immunosignature</td>
<td>32</td>
</tr>
<tr>
<td>1.6: Introduction of CTLA-4 and PD-1 Treatment</td>
<td>43</td>
</tr>
<tr>
<td>1.7: Biology and Mechanism of Anti-PD-L1 Antibody Treatment</td>
<td>44</td>
</tr>
<tr>
<td>2.1: Tumor Free and Growth Curve of FVB/N NeuN Mouse Model</td>
<td>49</td>
</tr>
<tr>
<td>2.2: Scheme of Experiment Design</td>
<td>50</td>
</tr>
<tr>
<td>2.3: Influence of Genetic Background on Immunosignature</td>
<td>52</td>
</tr>
<tr>
<td>2.4: Immunosignature Can Show a Separation Between Transgenic Mice and Age Matched Controls Two Weeks Before First Palpable Tumor</td>
<td>53</td>
</tr>
<tr>
<td>2.5: Heatmap of Immunosignatures of Different Time Points in Tumor Development</td>
<td>54</td>
</tr>
<tr>
<td>2.6: Time Course of Immunosignatures During Tumor Progression in FVB/N NeuN Mice</td>
<td>56</td>
</tr>
<tr>
<td>2.7: Venn Diagram of Different Immunosignatures Along Tumor Development</td>
<td>58</td>
</tr>
<tr>
<td>2.8: Numbers of Selected Peptides Overlap Between Two Adjacent Stages During Tumor Development</td>
<td>58</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>2.9: Numbers of Selected Peptides Overlap Between Earliest Tumor Signature with Other Stages Signatures</td>
<td>59</td>
</tr>
<tr>
<td>2.10: Principle Component Analysis of Different Batches</td>
<td>70</td>
</tr>
<tr>
<td>2.11: Correlation of Technical Replicates Before and After SVA Adjusted and PCA Analysis of Batch Effects</td>
<td>71</td>
</tr>
<tr>
<td>2.12: Correlation of Technical Replicates Before and After Combat Adjusted and PCA Analysis of Batch Effects</td>
<td>72</td>
</tr>
<tr>
<td>2.13: Comparison of Different Batch Effect Removal Methods</td>
<td>73</td>
</tr>
<tr>
<td>3.1: Cartoon of Experimental Design Outline</td>
<td>78</td>
</tr>
<tr>
<td>3.2: Tumor Growth Characteristics in FVB/N NeuN α-PD-L1 Treatment Experiment</td>
<td>80</td>
</tr>
<tr>
<td>3.3: RT-PCR Analysis of MMTV Promotor and PDL1 Expression in Established Tumors and Adjacent Normal Mammary Glands</td>
<td>81</td>
</tr>
<tr>
<td>3.4: Heatmap Showing PD-L1 Generated Immunosignature</td>
<td>83</td>
</tr>
<tr>
<td>AP1.1: Proportion of Overlapped Peptides Shared Among Different Time Points</td>
<td>125</td>
</tr>
<tr>
<td>AP1.2: Intensity Plot of Significant Peptides Between Primary and Recurrent OSA</td>
<td>126</td>
</tr>
<tr>
<td>AP2.1: Summary of Mice Samples</td>
<td>134</td>
</tr>
<tr>
<td>AP2.2: Tumor Growth Curve for Tg8_1-1 and Tg8_3-9 Mice</td>
<td>134</td>
</tr>
<tr>
<td>AP2.3: AbStat Time Course for Tg8-1-1 and WT2-3 Mice</td>
<td>135</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>AP2.4: AbStat Time Course for Tg8-3-9 and WT1-3 Mice</td>
<td>136</td>
</tr>
<tr>
<td>AP2.5: Entropy Time Course for Transgenic and Wild Type Mice</td>
<td>137</td>
</tr>
<tr>
<td>AP2.6: Concept of Autocorrelation Function and Durbin-Watson Test</td>
<td>142</td>
</tr>
<tr>
<td>AP2.7: Self-Organizing Map to Cluster 249 Autocorrelated Peptides in Tg8-3-6</td>
<td>143</td>
</tr>
<tr>
<td>AP2.8: Self-Organizing Map to Cluster 849 Autocorrelated Peptides in Tg8-3-6</td>
<td>144</td>
</tr>
<tr>
<td>AP2.9: Intensity of a Cluster of 24 Autocorrelated Peptides Correlated with Tumor Development</td>
<td>145</td>
</tr>
<tr>
<td>AP2.10: Concept of Robustness in Immunosignature</td>
<td>147</td>
</tr>
<tr>
<td>AP2.11: Correlation of Selected Peptides in Transgenic (Tg) and Wild-Type (WT) Control Mice</td>
<td>148</td>
</tr>
<tr>
<td>AP2.12: Correlation of All Peptides on the 10K Array</td>
<td>149</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

1.1 Breast Cancer

1.1.1 Overview of Cancer

Cancer is a group of diseases characterized by the uncontrolled growth and spread of abnormal cells [1]. Hanahan and Weinberg propose the following hallmarks of cancer: self-sufficiency in growth signals; evasion of growth suppressors; limitless replicative ability; resistance to apoptosis; sustained angiogenesis; activation of invasion and metastasis; evasion of immune destruction; and deregulation of energy metabolism [2, 3]. Genome instability and mutation and tumor-promoting inflammation are enabling characteristics of cancer cells [3]. Cancer is the leading cause of morbidity and mortality in the world. In 2012, there were approximately 14 million new cancer cases, and cancer accounted for 8.23 million deaths [4]. Worldwide in 2012, the five most common cancers in women were breast, colorectal, lung, cervical, and stomach. For men, they were lung, prostate, colorectal, stomach and liver cancer. In the year of 2015, cancer accounted for 589,430 deaths in US, and 78% of those diagnosed cancer are in people 55 years or older. In the next two decades, the number of new cancer cases worldwide is expected to increase by more than 70% [4]. Many factors can increase cancer risk, such as genetic background, viral infections, obesity, low fruit and vegetable intake, use of tobacco and alcohol and others [1, 4].
1.1.2 Overview of Breast Cancer

Most of following information is from the SEER Cancer Statistics Review and the American Cancer Society [5, 6].

Breast cancer is a type of malignant tumor initiated in breast. A normal breast is made of lobules to secret breast milk, ducts that carry milk from lobules to the nipple, connective tissue, blood vessels and lymph nodes (Figure 1.1). The lymph system of the breast is the primary way that tumor cells spread outside the breast. It includes lymph nodes and vessels that carry lymph. When breast cancer cells spread into the lymph system, cancer cells will reside in the local lymph nodes or later spread to other organs through metastasis.

Figure 1. 1 Biology of Breast Tissue
Figure adapted from American Cancer Society with permission [6].
There are several major types of breast cancer. When abnormal cells originate in the ducts, the cancers are ductal. According to whether or not abnormal cells invade the walls of ducts and spread to nearby lymph nodes and other breast tissues, ductal carcinoma is either ductal carcinoma in situ (DCIS) or invasive ductal carcinoma (IDC). IDC is the most common type of breast cancer. Different from DCIS and IDC, invasive lobular carcinoma (ILC) start in the cells lining the lobules and become invasive. Inflammatory breast cancer (IBC) is a rare type of invasive breast cancer. Patients with IBC typically have breast skin that is red, thick and pitted in texture. IBC patients may also feel their breast become bigger, hard or itchy. IBC does not present with abnormal lump and is likely to be missed by a mammogram.

Cancer stages describe the extent of a cancer in a patient. In general, cancer stages can be divided into localized when cancer cells are only found in breast; regional when cancer cells spread into regional lymph nodes; or metastatic when the cancer cells spread to distant organs.

Breast cancer is the most common cancer and the second leading cause of death among women in the United States. In 2012, nearly three million women were living with breast cancer in the US. It is estimated that, in 2015, there will be 231,840 new cases of invasive breast cancer (14% of all new cancer cases) and 40,290 breast cancer deaths (6.8% of all cancer deaths) in the United States. 12.3% of women (one in eight) will be diagnosed with breast cancer during their lifetime. The five-year relative survival rate is used to calculate the proportion of patients expected to be alive five years after diagnosis compared with a general population.
of the same age, race and sex that haven’t been diagnosed with cancer. Based on data from the Surveillance, Epidemiology and End Results Program (SEER) 2005- 2011, breast cancer has a five-year relative survival rate of 89.4%. If we consider different stages of breast cancer, localized breast cancer account for 61% of new breast cancer cases, while regional and distant account for 32% and 6% of new cases, respectively.

According to American Cancer Society data, in the year of 2015, 61% of breast cancer cases are detected at a localized stage, with a five-year relative survival rate of 99% [1]. However, when tumors spread to nearby lymph nodes or other tissues, the survival rate decreases to 85% [1]. If the breast cancer is in a more advanced stage, where tumors are found in lymph nodes around the collarbone or in distant organs, the survival rate falls to 25% [1]. Cancer can reoccur well after the 5 year time period. Women who have had breast cancer live in the fear of recurrence all the time.

### 1.1.3 Overview of Transgenic Mouse Model

The mouse has played an important role in studying the basic biology of breast cancer. With more understanding of the molecular basis of breast cancer development, different oncogenes are engineered into mice to manifest overexpression of a specific oncogene or depletion of a specific tumor suppressor gene in mouse germ line cells. Over 100 different genetically engineered—also called transgenic—mouse models have been constructed and studied, including, HER2, p53, c-myc, TGFα, Cyclin D1 and many others [7-10]. The use of transgenic mouse models allows researchers to study breast cancer samples that
are difficult to access in clinical trials such as the Women's Health Initiative (WHI) study and the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer-screening trial. Multiple time points of samples that cover different stages of tumor development can be collected in a well-controlled manner. This creates a huge advantage in the study of early-stage breast cancer, as human early-stage cancer samples are scarce.

In order to correctly represent breast cancer development in humans, the mouse models must have enough similarity with human breast cancer development—in both genetic and protein-based comparisons—so that the discoveries in mice may be applied to human breast cancer diagnostics. Herschkowitz and colleagues have shown that transgenic breast cancer mouse models have significant genetic similarities to human breast cancer. They used six models with different transgenes to represent different human subtypes of breast cancer. Many of the defining characteristics of human breast cancer can be seen in the mouse models [9]. In the area of immune response, Lu and colleagues used a transgenic mouse model with overexpression of the neu oncogene that developed a serum autoantibody repertoire similar to that in cancer patients [11].

In this study, I use the MMTV-neuN mouse model, which is the most extensively studied breast cancer model in prevention studies [8]. The neuN gene is the mouse homolog of the human Erbb2 gene, which is overexpressed in about 20% - 30% of human breast cancers [12]. The neuN gene is overexpressed in mouse mammary glands under the transcriptional control of the mouse mammary tumor virus promoter long terminal repeat (MMTV-LTR) [13]. This mouse model
has been used to test different cancer vaccines to prevent tumor development [14-18]; to find biomarkers for breast cancer [9, 11, 19]; to test different drug treatments [20-23]; and to study basic mechanisms of breast cancer [9, 13, 24-27].

The MMTV-neuN mouse model shares a great similarity with human breast cancer. However, it also has its limits. The cellular morphology is different from that of human breast cancer. This mouse model can represent human ductal carcinoma in situ (DCIS) but not lobular intraepithelial neoplasia. At late stages of tumor development, bone and brain metastases are missing from this model [12]. Overall, transgenic mouse models serve as a relevant, rapid and inexpensive system to investigate breast cancer in humans [8, 10].

1.2 Current Breast Cancer Diagnosis and Biomarkers

1.2.1 Statistics Preparation for Diagnosis and Biomarker

In order to measure the performance of a diagnostic test or a biomarker, sensitivity, specificity and accuracy are often used. In a diagnostic test of a disease, a positive call means that the test score of a sample is positive for the disease. A negative call means that the sample’s test score is negative for the disease. In a test with an overall population of (A+B+C+D), A is the number of positive samples that actually have the disease. B means non-diseased samples that score positive. C means diseased samples that score negative. D is the number of negative samples that do not have the disease.

Sensitivity is the measure of the proportion of positive samples in overall samples actually with the disease -A/(A+C). Specificity describes the proportion of negative samples among the samples without the disease –D(B+D). The false
positive rate, which equals 1-specificity, is more frequently used. Accuracy is the proportion of positive and negative samples that are correctly predicted to the actually disease condition in the overall population- \((A+D)/(A+B+C+D)\).

**1.2.2 Current Clinical Diagnostic Methods of Breast Cancer**

The American Cancer Society provides screening guidelines for “early detection of cancer in average-risk asymptomatic people” for 2015[1]. Breast self-examination (BSE), clinical breast examination (CBE) and mammography are recommended for women over age 20 (Figure 1.2). Ultrasound (US) and magnetic resonance imaging (MRI) are also additional diagnostic tests widely used in clinical settings [28].

![Figure 1.2: Illustration of Mammogram and MRI](image)

*Figure 1.2: Illustration of Mammogram and MRI*

Figure adapted from American Cancer Society with permission [6].

**1.2.2.1 Breast Self-examination (BSE)**

Adult women at the age of 20 should begin to do BSE at least once a month and report any new breast symptoms to healthcare professionals. Symptoms include changes in how the breast feels, its appearance, and any
discharge from nipple. Specific symptoms include lumps, nipple tenderness, thickening of the breast, skin texture changes and others. BSE can be done routinely and allows patients to monitor their breasts regularly. Although BSE serves as a good tool to self-monitor for any abnormal breast symptoms, when such symptom occurs, a clinical breast examination should be performed [1, 29].

1.2.2.2 Clinical Breast Examination (CBE)

A clinical breast examination is a clinical exam performed by a healthcare professional that is trained to diagnosis abnormal breast symptoms. The goal of CBE is to check any lumps or physical changes in order to achieve early diagnosis of breast cancer. During the exam, a healthcare professional will check the patient’s breast to discover any unusual texture, lumps, and suspicious areas. If a lump is hard and cannot move easily, further tests are needed.

For women in their 20s and 30s, CBE is recommended to be a part of patient’s periodic health examination at least once every 3 years. For asymptomatic women over 40, CBE is recommended annually. When mammography and other sophisticated tests are not available, CBE is an easy way to perform breast cancer screening [1, 29].

1.2.2.3 Mammography

A mammogram is an x-ray image of the breast used to screen for breast cancer. During a mammogram, a patient’s breasts are compressed between two firm surfaces to spread out the breast tissue. The breasts then are exposed to small dose of iodizing radiation to produce an image of the breast tissue. A trained healthcare professional will examine the image to look for any sign of cancer.
Mammography is recommended annually for women over 40. Women with a high risk of breast cancer should consult with professionals on the advisability of mammography before age 40. The advantage of mammography is that it can often detect a breast lump before the patient can feel the lump. If the mammogram reveals an abnormal area in the breast, additional tests, such as ultrasound or MRI, should be offered, and a biopsy should be performed when test results show that the mass is solid[1, 29]. A mammogram is much more expensive than CBE, and, thus, it not feasible as a routine test for all women.

The estimated sensitivity of mammography ranges from 29-97% with a mean of 77%, and the false-positive rate ranges from 1-29% with a mean of 10% [30]. Several clinical trials showed that mammography screening produced a projected 15-20% reduction in breast-cancer-related mortality [29]. Although mammography has played an important role in breast cancer diagnosis and has helped save lives, it has many critics. Mammography has limited sensitivity and a high false-positive rate [31, 32]. It has been shown to detect only 70% of breast cancers [33]. 54% of all patients that were diagnosed to have breast cancer by screening mammograms, up to 54% are results of overdiagnosis [34]. In addition, mammography tends to have low sensitivity in women with dense breast tissue [35, 36]. Women with aggressive breast cancer, such as triple negative breast cancer [37], and those younger than 50 years old [38] also benefit less [39]. Around 80% of the false positives in mammography are caused by benign growths [40, 41]. Benign growths can be classified as proliferative, hyperplasia and hyperplasia/atypical growths [42, 43]. All three classes are associated with
different breast cancer risks compared to the risk for normal women.

Hyperplasia/atypical growths account for a 4.5-fold increase in the risk for breast cancer [44]. However, mammography does not perform well to differentiate the three classes. Absolute mortality benefit for women screened annually for 10 years is around 1% on average [45].

Another concern about mammography is that the value of mammograms is associated with the experience of physician who interprets them [30]. Physicians who interpret 2500-4000 mammograms annually with a high screening focus have a 50% lower false-positive rate than physicians who interpret 480-750 mammograms annually with a low screening focus[30]. This association between the physician’s experience and the performance of mammography may bias breast cancer diagnosis.

1.2.2.4 Adjunctive Testing: Ultrasound Imaging and Magnetic Resonance Imaging (MRI)

Ultrasound imaging uses high-frequency sound waves to view the breast tissue and can capture images in real time. During an ultrasound exam, a transducer is placed on the skin with a thin layer of gel in between to transmit ultrasound waves from the transducer into the patient’s body. MRI uses magnets and radio waves to produce images of the breast. MRI is the most sensitive imaging method for breast cancer detection [46].

Ultrasound and MRI are adjunctive diagnostic tests used when abnormal areas have been found in the breast. Although current evidence on adjunctive testing is limited, it has proved beneficial in identifying breast cancer in patients.
who have dense breasts, as indicated by mammograms [29]. Most of the cancers detected by adjunctive tests are invasive ductal cancer (IDC) rather than ductal carcinoma in situ (DCIS) [47]. However, due to the high cost of MRI, it typically is used as adjunctive testing when something suspicious has been found, and not for screening of the general population. Also, it is important to note that adjunctive tests can increase the false-positive rate [29].

1.2.2.5 Molecular Diagnosis Platform

Polymerase chain reaction (PCR) and DNA microarray are often used to detect specific mRNA or DNA sequences of breast cancer. The advantage of PCR is that it can perform low-noise amplification even when the signal is low-level. DNA microarray technology can simultaneously detect multiple sequences. The FDA has approved Oncotype DX, HOXB13-IL17BR assays and the MammaPrint assay to measure gene expression of breast cancer [48].

Oncotype DX detects a 21-gene profile by reverse-transcriptase PCR to predict the risk of recurrence in patients taking Tamoxifen. Sixteen of the 21 genes are cancer-related genes with reference genes. A mathematical algorithm is derived from empirical retrospective study to calculate a score to estimate the risk of distant recurrence [49].

MammaPrint is a gene expression assay by Agendia that detects a 70-gene profile by reverse-transcriptase PCR. It used 117 patients with axillary lymph node-negative primary breast cancer to find genes highly correlated with a short interval from primary tumor to distant metastases. MammaPrint profiling is used to identify a sub-population of patients by their prognosis performance [50].
HOXB13-IL17BR assays use complementary DNA microarray technology to detect the expression level of HOXB13/IL17BR genes. It is used to predict the outcome of Tamoxifen in untreated ER-positive/node-negative patients [51].

Current breast cancer diagnostic methods suffer from low sensitivity and high false-positive rates. Breast tumors with high breast density and aggressive growth are not detected during current diagnostic procedures [52, 53]. Moreover, CBE, mammography and adjunctive tests are not effective in the early detection of breast cancer.

1.2.3 Benefit of Early Detection and Treatment

Early detection of cancer identifies tumors before they are diagnosed can bring more opportunities for medical intervention. With early detection, it seems reasonable to expect that systematic therapy involving smaller amounts of less-toxic medications can be initiated earlier for a shorter period of time and be more effective. This could potentially revolutionize current treatment strategies and improve patients’ outcomes—and, perhaps, even eradicate the disease. Take ovarian cancer for example. Patients with stage I ovarian cancer, a 42-mm tumor diameter, on average, have a five-year survival rate of 92%. The five-year survival rate decreases to around 30% when patients are in stage III to stage IV [54]. Thus, researchers and medical professionals have spent considerable effort in seeking a noninvasive, simple and inexpensive means for the early detection of cancer.
1.2.3.1 Early Diagnosis of Breast Cancer Can Improve Survival Rate

In cancer treatment, patients with localized cancer have better survival rates than those with metastatic disease (Figure 1.3) [55]. Failure to detect local cancer at an early stage is the major problem in cancer diagnosis and treatment.

In breast cancer, in the localized stage, tumors have not spread to lymph nodes, nearby structures outside of the breast. According to American Cancer Society projected data for 2015, 61% of breast cancer cases are detected at a localized stage with a five-year relative survival rate of 99% [1]. However, when tumors spread to nearby lymph nodes or other tissues, the survival rate decreases to 85% [1]. If the breast cancer is in a late stage, in which tumors are found in lymph nodes around the collarbone or more distant, the survival rate falls to 25% [1]. Therefore, an early detection system could eliminate the mortality caused by late-stage metastasis and largely improve the five-year survival rate for patients.

We currently lack a screening strategy to make a noninvasive, timely diagnosis.

**Figure 1. 3: Five-year Relative Survival Rates (%) by Stage at Diagnosis in US from 2004-2010**
Figure adapted from American Cancer Society with permission [1].
1.2.3.2 Proposed Early Diagnosis and Treatment Scenario

In order to be successful, an early diagnosis system should have the following merits. First, it should have high sensitivity and a low false-positive rate for patients in the early stage of breast cancer. One of the main challenges with mammography is its high false-positive rate. The proposed early detection system is a screening system for the general population or the sub-population of patients that have found an abnormality during self-exam. Second, the cost of diagnosis by this system should be low so patients can be examined routinely. Although MRI has the highest sensitivity to detect small tumors, its high cost prevents patients from having MRIs frequently. The cost of an early detection system should be less than the cost of clinical breast examination. Another advantage of routine examinations is that they can increase prediction accuracy [56]. For example, a test with 95% accuracy per test will be have 99.75% accuracy after two tests and 99.99% after three tests. Routine testing allows establishing a baseline which should also decrease false positives. In order to allow for routine examinations, the system needs to use non-invasive techniques to screen asymptomatic populations. Blood, saliva or other fluids could be good candidates to be used in the test.

I propose a scenario for early detection of breast cancer through routine examinations. Patients will perform breast self-examination to find any lump and use an early detection system to monitor any aberration. If the system was simple and inexpensive, patients could get the diagnosis result well before lump was detected. Further testing by mammography, MRI, ultrasound and others will be
performed when the test shows an increased risk of breast cancer. Early surgery and therapy will be used to kill the cancer using a low-dose, non-toxic treatment.

Some people may argue that the early diagnosis of breast cancer will increase the financial burden on patients and society. Their logic assumes that if a test has 99% specificity, and over 100 million US women are over age 20[57], annual detection will lead to one million false-positive cases. If the cost of a follow-up mammogram for every false positive is around $100, this will lead to 100 million dollars in unnecessary costs per year. However, this description totally ignores the increased sensitivity of the early diagnosis system. Also, new therapies may appear with the early detection system. The regimen of drugs targeting early-stage cancer could be different from the current ones targeting later-stage cancer. Because the target for early diagnosis is the general population, the market for drugs to treat early breast cancers will increase dramatically and, thus, reduce the costs of drug development and manufacture. This has already been seen in many drugs; a prime example is aspirin, first marketed by Bayer in 1932 and, today, a very low-cost widely-used drug.

1.3 Early Detection of Breast Cancer by Antibodies

1.3.1 Overview of Blood Based Breast Cancer Biomarkers

Current molecular approaches for detecting breast cancer can fall into two categories: 1. direct detection of the tumor or molecular tumor cells shed; 2. detection of the patient’s immune response to the tumor.
Biomarkers that detect the tumor or the cells it sheds include proteins, miRNAs, circulating tumor cells (CTC), aberrant glycosylation and others. Biomarkers that detect the immune response of patients include antibodies that will be discussed in the following section.

CA 15-3 and CA27.29 are biomarkers to detect the circulating MUC-1 antigen in the blood. They are different epitopes on the same MUC-1 antigen. A high CA 15.3 level is correlated with a large tumor size and presence of lymph node metastases [58]. These biomarkers have been shown to have prognostic value in early-stage breast cancer [58]. According to the American Society of Clinical Oncology guidelines, CA 15-3 and CA27.29 are not recommend for screening, diagnosis and staging of primary and recurrent breast cancer because their role in early-stage breast cancer and treatment decisions is not clear [59].

HER2 is a member of the epidermal growth factor receptor (EGFR) family 70. 15% -30% of diagnosed breast cancers have an overexpression of HER2 [60]. It is recommended that HER2 expression be evaluated in primary invasive breast cancer from the time of diagnosis to recurrence [60],[61]. It can help to guide the use of trastuzumab in the adjuvant setting [62].

Circulating tumor cells (CTC) are tumor cells present in the blood that represent a specific cancer type. CTC may predict the presence of micrometastasis or aggressive primary breast cancer [63]. However, the low concentration of CTCs (1 in 10^6~10^7 leukocytes) [63, 64] makes it difficult to be captured by immunomagnetic beads or to use RT-PCR to amplify signals [65-67]. The CTC assay is not recommended in the guidelines for diagnosing breast cancer.
Changes in circulating microRNAs (miRNAs) have been found to be significantly related to early and minimally invasive breast cancer. A panel of 31 candidate miRNAs has shown significant differences between 20 women with early-stage breast cancer and 20 control patients [68]. Another study, using a panel of seven miRNA candidates in 148 patients with breast cancer and 44 controls, showed significantly altered miRNA expression levels [69]. Tumor-specific miRNAs are tissue-specific, dysregulated in cancer, highly abundant and stable in the blood [70, 71]. miRNAs could serve as a good candidate for non-invasive biomarkers.

Aberrant glycans and glycoforms of proteins could also be a source of breast cancer biomarkers to monitor disease progression. Abd Hamid and colleagues discovered that a trisialylated triantennary glycan containing alpha1, 3-linked fucose with a twofold increase in breast cancer patients compared with controls [72].

Although early diagnosis of breast cancer has many merits, only limited biomarkers have been discovered and widely used in the clinical setting. Methods to identify novel breast cancer biomarkers are still urgently needed. Each approach has its strengths and weaknesses. No single technology can provide the ideal detection performance with high sensitivity and specificity.

1.3.2 Challenges in Finding Early Breast Cancer Biomarker

In order to be useful in early detection of breast cancer, a biomarker should have high sensitivity and a low false-positive rate, and it should show evidence of value in determining treatment after the early detection. The
biomarker should also be stable and robust and available in a non-invasive test. However, with thousands of candidate biomarkers identified, only a few are approved each year and widely used in the clinic [29, 59]. Clearly, searches for biomarkers for early detection have not yet been generally successful [73-75]. Researchers in early detection of breast cancer face several challenges, including over-fitting of the algorithm by small sample size, unstandardized sample preparation and specimen annotation [76, 77].

In early detection studies, one problem is that most biomarkers are discovered from diagnosed cancers and then are used in an attempt to diagnose early cancer. From the biology of cancer, tumors change dynamically at different stages. Samples at the time of diagnosis are most likely to be irrelevant for early-stage samples. Antigens in early high-risk lesions may have a different expression panel from later stages.

The reason that many researchers still use samples with diagnosed tumors is that samples from the same patients before and after tumor diagnosis are hard to get. It is difficult to study the correlation of a biomarker with tumor progression and early-stage breast cancer without the proper early-stage samples.

Many clinical studies also have suggested that using biomarkers discovered and widely used in diagnosing cancer perform poorly in pre-diagnostic patients [78-81]. Zhu and colleagues studied 118 patients one to two years before ovarian cancer diagnosis [79]. Among 28 evaluated biomarkers, CA125 achieved the highest score, with a sensitivity of 69.2% and a specificity of 96.6% in diagnosed samples but failed in pre-diagnostic samples. Other models had poorer
performance than CA125 alone. Zhu et al. concluded that “biomarker panels discovered in diagnostic samples may not validate in pre-diagnostic samples.” In the Carotene and Retinol Efficacy Trial, Anderson and colleagues showed they could achieve only limited discriminatory power in pre-diagnostic ovarian specimens and that the power increased as the time of diagnosis approached [78]. In breast cancer, Lu and colleagues used a sample 150 days before diagnosis, and HER2 achieved a ROC score of only 0.63 and p53 with 0.63 [81]. In the Prospect-EPIC study, Opstal-van Winden and colleagues used samples from 68 women with a median of 21.3 months before breast cancer diagnosis and 68 controls. After evaluation of ten breast cancer biomarkers, none of those resulted in correct classification [80]. In these studies, researchers concluded that markers discovered in diagnostic samples may not be validated in pre-diagnostic samples.

1.3.3 Cancer Antibodies as Biomarkers

1.3.3.1 Cancer Patients can Develop Autoantibodies Against Tumor Early

The immune system in our bodies protects us from pathogens and abnormal conditions. It continuously monitors any for foreign invader. B cells from the humoral immune system and T cells from the cellular immune system play important roles in finding non-self-antigens. Antibodies produced by B cells bind to specific antigens and neutralize the foreign invader. After binding, B cells are activated and form a memory response to encounter future events from the same antigen.

T cells are also triggered and help B cells to carry out cytotoxic killing of infected cells. In order to correctly recognize foreign antigens instead of self-
molecules, B cells and T cells are negatively selected to remove any B cells and T cells that can bind to self-proteins.

Cancer cells start as normal cells and enter into atypical growth later. The abnormal changes lead to changes in the abundance of protein expression, different post-translational modification patterns, and other changes. Burnet and Thomas developed an immune surveillance hypothesis in 1971 after Ehrlich proposed that the immune system can recognize and destroy nascent transformed cells in the human body [82]. The formal immune surveillance describes that thymus-dependent immune cells can produce an effective immune response to tumor antigens and protect the host from nascent transformed cells [83-85]. Later, this theory was refined by an immune-editing concept that shows that there is a dynamic interaction between the immune system and the tumor. In the refined theory, immune surveillance inhibits tumor development from tumor initiation, and the tumor evolves through multiple mechanisms to avoid immune system surveillance [86, 87].

A classic example that reflects the role of antibodies early in cancer can be found in the paraneoplastic syndrome (PNS) [88]. PNS is a syndrome that results from a cancer causing an autoimmune effect on the nervous system. When a tumor and nervous tissue share common antigen, then remote pathologic effects will influence the nervous system. Sometimes the symptoms of PNS are detected months to years before the formation of a malignant tumor because tumor cells express neuronal antigens and trigger an anti-tumor immune response [89]. This is further supported by Marcia Wilkinson and colleagues, who
discovered a neuronal antibody in the serum of patients with small-cell lung cancer [88, 90]. After removal of the cancers, PNS is usually cured. These findings reveal that tumors elicit an anti-tumor antibody response at early stages [88, 91].

More research has shown that autoantibodies can be found several months to years before the appearance of symptoms of lung cancer [92], esophageal squamous cell carcinoma [92], prostate cancer [93], breast cancer [94], and others [91, 95, 96].

B cells have been detected in invasive breast cancer [97], a phenomenon called immune infiltration. This shows that the immune system closely monitors cancer cells. In 2005, p53 antibodies were detected in 12 of 49 individuals who developed cancer later, compared with four out of 54 individuals who did not [98]. Autoantibodies against HER2 have also been detected in early-stage breast cancer patients [99]. Evidence of autoantibodies to tumor-associated antigens (TAAs) precedes manifestations of diagnosed breast cancer, making autoantibodies a source of biomarkers to detect breast cancer early.

1.3.3.2 Advantage of Using Antibodies as Biomarker

Although the discovery of biomarkers for early diagnosis of breast cancer holds great promise, the low concentrations of biomarkers in blood impede the development of effective tests. Current plasma protein biomarkers in the human proteome organization (HUPO) project are in the range of ug/ml to mg/ml (Figure 1.4). When using proteins and miRNA secreted from tumors or circulating cells as biomarkers to detect early-stage cancer, they are 10-fold too low to be detected
by state-of-the-art technology [100-102]. At the time of diagnosis by current biomarker technology, presumably a tumor would be larger than 1mm$^3$ in size, with more than three million cells, and would have been developing in body for several years. It is not feasible to detect such a low signal without any amplification technology to extract and amplify the tumor-specific signal. Tumor-specific antibodies are the perfect biomarker candidate to amplify tumor-specific signals [103].

Figure 1. 4: Distribution of Human Plasma Protein. Figure adapted from Anderson et al. with permission [101].
After activation, a B cell can produce 300,000-1,200,000 antibodies per hour [104, 105] and replicates every 70 hours [106]. It has a lifespan of up to 4.5 months [107, 108]. B cells towards a tumor antigen can amplify the production of tumor-specific antibodies up to ~10-fold\textsuperscript{11} in one week [109, 110]. Memory B cells that generated those antibodies can exist years after the immunogenic event occurs.

Antibodies are also stable and resistant to common types of proteolysis. Other proteins and small molecules will be either rapidly degraded or cleared from the blood. Antibodies have a half-life of over seven days in the blood. Even after separation from the blood, antibodies are stable for several years [111, 112]. This means that we can store achieved blood samples in sera or a standard filter spot [113]. This allows researchers to use stored, historical samples to find antibody biomarkers [114].

Antibodies are also easy to detect. Each subtype of human immunoglobulin shares the same Fc structure on heavy chains. Commercial antibodies specifically binding to each subtype of human immunoglobulin are widely available with high specificity and affinity. This means that researchers can use the same secondary antibody to detect different tumor-specific antibodies (primary antibody), thus largely reducing the complexity of developing a biomarker.
1.3.3.3 Breast Cancer Associated Antibodies

Much effort has gone into finding antibody biomarkers to diagnosis breast cancer. ELISA, protein array and other technologies have been used to study single biomarkers and a panel of multiple biomarkers.

Early in 1997, Disis and colleagues found the presence of the HER2 antibody in 12 of 107 breast cancer patients and none among the 200 normal controls. They showed that nine of 44 patients with HER2-positive tumors had antibodies against HER2 [115]. Autoantibodies against HER2 have been detected in early-stage breast cancer patients [99]. Later, in 2005, p53 antibodies were detected in 12 of 49 individuals who developed cancer later, compared with four out of 54 individuals without cancer [98]. This study, for the first time, showed the relationship between p53 autoantibodies and the subsequent development of malignancy. A panel of multiple antibody biomarkers was investigated to increase prediction performance. Chapman and colleagues reported a panel of six antigens (p53, c-myc, HER2, NY-ESO-1, BRCA2 and MUC1) to distinguish patients’ samples of primary breast cancer from normal and ductal carcinoma in situ. Antibody response was observed in at least one of six antigens in 64% of primary breast cancer samples and 45% of DCIS samples with 85% specificity [116]. Anderson and colleagues used a protein microarray to discover 28 tumor antigens that can distinguish invasive breast cancer (stages 1-3) from benign breast disease [95]. Desmetz and colleagues focused on autoantibody biomarkers in in-situ carcinoma (CIS) in younger women under age 50) with breast cancer. They used a panel of antibodies against PPIA, PRDX2, FKBP52, HSP60 and MUC1 to study
60 primary breast cancer patients, 82 CIS patients and 93 healthy controls. The panel of biomarkers significantly distinguished primary breast cancer with AUC of 0.73 and CIS with AUC of 0.80 from healthy controls [117]. Five biomarkers (GAL3, PAK2, PHB2, RACK1 and RUVBL1) were investigated in DCIS and node-negative early-stage breast cancer samples. The five markers significantly distinguished healthy controls from early-stage cancer with AUC of 0.81 and DCIS with AUC of 0.85 [118].

1.3.4 Emerging Technologies for Profiling Antibody Responses

Humoral response can provide us with important information about a person’s disease progression and overall health status. In order to get information about humoral response, we need high-throughput methods to profile an antibody repertoire in a patient at a given time. In order to do that, we need a large library of binding ligands to represent possible antigens. The ligand could be the real antigen, a peptide sequence representing the epitope, or a mimotope sharing the same binding with a disease-specific antibody. Several technologies have been proposed in the past and are summarized below [119, 120].

1.3.4.1 SEREX

The serological screening of recombinant cDNA library using phage display (SEREX) is as follows: lambda phages are used to display proteins from a recombinant cDNA library of interests and probed with sera of interest to select antigens with high binding interactions with antibodies in the sera. Michael Pfreundschuh and colleagues [121] first introduced SEREX to isolate tumor antigens that trigger a high-titer humoral response in cancer patients. More than
2700 immunogenic antigens have been identified by SEREX [120, 122, 123]. The methodology of SEREX includes three main steps: building a cDNA library; probing with sera; and iteratively selecting sub-clones of phages.

Building the cDNA library: The cDNA library is generated by reverse transcription of RNA from tissues or cell lines of interest. After the reverse transcription, all cDNA is inserted into constructed lambda phage vectors. After plating phages and transfer onto a nitrocellulose membrane, proteins will be expressed in an E. coli expression system.

Probing with sera: Phage clones will be probed with diluted serum samples from cancer patients and proper controls. High-binding clones will be detected by an antibody-coupled enzymatic reaction and selected to be sequenced. Normally, candidates will be validated by independent methods such as ELISA or microarray.

SEREX has been used to identify tumor-associated antigens (TAA) NY-BR-1 to NY-BR-7 [124] and p33ING1[125] in breast cancer; NY-CO-37 and NY-CO-38 in colon cancer [126]; NY-ESO-1 and SSX2 in esophageal cancer and melanoma [127, 128]; and NY-ESO-1, LAGE-1, and XAGE-1[129] in prostate cancer.

Although SEREX has been used to discover thousands of tumor-associated antigens, it has internal limitations. First, a large serum volume is required to screen for specific phage clones due to the extensive screening. However, many historical sample databases have only limited aliquots of each patient’s sample. Second, SEREX is biased to high-expression mRNA—a limit
that is internally inherent from its beginning. As we recall from its definition, SEREX aims to select antigens with high-binding interactions with antibodies in the sera. Although the cDNA library could have as many as $10^6$ copies of phages, the constructed cDNA library will overrepresent the reverse transcriptions of those mRNA with a high expression level. When cancer-associated antigens are expressed at a lower level, those that are important indicators in early-stage will be substantially diluted and not even be represented during the screening. Third, because SEREX uses prokaryotic expression library, the structure of the antigens may not be correct—e.g., unnatural folding and post-translational modification. This will lead SEREX to miss some TAA and to an increased false-positive rate. Finally, the decrease in amount of an antibody will not be picked up by this system, which could also be as biomarker for diagnosis.

1.3.4.2 Peptide and Peptoid Microarrays

Peptide microarrays are comparable to SEREX in that they both use a large library of short peptide sequences. However, peptide microarrays overcome the bias of high-expression mRNA and the large volume of serum required in SEREX. The core of peptide microarrays is a high-throughput way to display a large library of peptides with different sequences. Peptides can be chemically synthesized in large quantities in advance, and pure products are then linked to microarrays. This gives the peptide library a long shelf life. Sigma-Genosys provides a synthesis platform to allow the rapid parallel synthesis of custom libraries. Peptides can also be directly generated in situ[130]. More than 330,000 peptides can be generated on silicon wafers per assay, using
semiconductor manufacturing technology. Diluted serum samples will be probed on peptide microarrays, and the signal will be collected by labeled secondary antibodies with dye.

Peptide microarrays have been used to study cell interaction[131], binding sequences of antibodies [132], enzymes [133, 134], proteins [135], and DNA and small molecules [136-138]. One important focus of peptide microarrays is on epitope mapping. Tiled peptide sequences with partial overlaps reveal the epitope of a monoclonal antibody [139]. The immunosignature section of this paper will include a more-detailed analysis of peptide microarrays.

Other non-natural molecules can also be used for antibody profiling. One example is peptoid, a small molecule of N-substituted oligoglycines [140]. A peptoid microarray is constructed to display thousands of peptoids reproducibly on a chemically functionalized glass surface [141]. Diluted serum samples are probed on peptoid microarrays, and the signal will be collected by a labeled secondary antibody with dye. Muralidhar Reddy and his colleagues have shown the application of peptoid microarrays in their discovery of two candidate IgG biomarkers for Alzheimer’s disease [140].

1.3.4.3 SERPA

Serological proteomic analysis (SERPA) uses a different approach to overcome problems in peptides or protein synthesis. SERPA takes advantage of two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and mass spectrometry analyses. Instead of using chemically synthesized peptides or expressed proteins to mimic antigens in the patients’ samples, SERPA directly
uses cell lines or tissue lysates. It first separates proteins in the samples by their isoelectric point and then further separates them by molecular mass in 2D-PAGE. The separated proteins are transferred onto membranes and then are probed with patient or control serum samples. The corresponding blot will be extracted from the 2D-PAGE and identified by mass spectrometry. SERPA overcomes some of the major limitations of other platforms, such as how to ensure a correct protein structure. But unlike other methods using microarray, it lacks the ability to perform high-throughput screening [142] and also requires a large number of serum samples [143].

1.3.4.4 Protein microarray

The definition of a protein microarray is the technology that displays an array of proteins on a slide surface in a reproducible and addressable manner. The binding of each protein can be analyzed in a high-throughput way. Roger Ekins first introduced the concept of the protein array in 1989 [144], and MacBeath et al. developed the first mature protein array in 2000 [145]. Heng Zhu and his colleagues also printed protein arrays with 5,800 yeast proteins on a slide in 2002 [146].

Protein microarray can be divided into three categories: analytical protein array, reverse-phase protein array and functional protein array. Analytical protein arrays use antibodies printed on the array surface to capture proteins labeled with fluorescence in cell lysate [147]. Reverse-phase protein arrays, on the contrary, print cell lysate on a glass surface and probe proteins in cell lysate with fluorescence-labeled antibodies of interest. This platform requires high-quality
antibodies to achieve reliable results. Functional protein arrays aim to investigate the biochemistry properties and interactions of proteins with their targets, such as protein, DNA, RNA, drug, and enzyme. In a functional protein array, proteins are synthesized by a protein expression system and then spotted on the array surface or directly synthesized on the array surface. Cell-based expression systems include bacteria, insect cells or yeast, as well as the in-vitro expression system [145, 146, 148-150]. A widely used human protein microarray is Protoarray®, manufactured by Thermo Scientific [151]. In order to represent the correct protein structure, Protoarray uses the Sf9 insect cell line, which is the close to the mammalian expression system.

Protein microarrays have been used extensively to profile humoral response to diseases and to screen for disease-specific autoantibody biomarkers [152-157].

Although thousands of tumor-associated antigens have been discovered using the protein microarray, it also has limitations. First, because proteins need to have the correct structure and be purified to ensure high-quality proteins that can be spotted on arrays, much effort is required to express the proteins correctly and to isolate and purify proteins from the expression system mix [158]. Post-translational modifications may not be represented. Second, after collecting pure proteins, storage is a big issue. Proteins have only a limited shelf life and may aggregate or even become denatured after a few weeks. This will influence any functional study or protein-antibody bindings that require conformational epitopes [159]. A cell-free in situ expression system, Nucleic Acid Programmable Protein
Array (NAPPA), has been proposed to address these two problems [148]. Proteins will first be encoded by plasmids with a fusion tag and spotted on an array surface. After being expressed through a cell-free expression system, proteins will be captured by an affinity reagent in situ. However, some steps will still be required: building a plasmid system; inserting each protein into plasmids; and printing each plasmid to each spot on the microarray. Before probing with sera, proteins on a NAPPA array need to be expressed by an in vitro expression system. In addition, these systems are expensive and do not lead themselves to high-throughput production so are primarily useful as discovery platforms.

1.4 Immunosignature

1.4.1 Definition of Immunosignature

The immunosignature is the binding pattern of a complex mixture of antibodies. Immunosignature technology uses a peptide array with 10,000 spotted peptides selected without bias from a non-biological peptide sequence space (Figure 1.5). It can reflect a disease state or, more generally, the health status of an individual. Immunosignature is a multidimensional reflection of overall humoral immune response. It can serve as a universal diagnostic platform to detect and identify any disease with a significant antibody response.
Left is the scan image of CIM 10K array. Right is the major steps in immunosignature. Diluted Serum samples are probed on arrays. Biotinylated anti-mouse IgG (H+L) antibodies are incubated on arrays next. Bound secondary antibodies are visualized by Alexa Fluor-647 labeled streptavidin.

1.4.1.1 Building a Complex Chemical Surface

At the core of immunosignature technology is an array with 10,000 random-sequence peptides spotted in an addressable, machine-readable fashion with 20-mer peptides.[160]

A random number generator generated the peptides sequences. We designed the algorithm of the random number generator to evenly cover as many combinations of amino acid sequences, potentially epitopes, as possible. These peptide sequences were not based on protein sequences in nature. For the choice of amino acids, we included all natural amino acids except cysteine to synthesize 10,000 peptides. Cysteine is used only on the C-terminus of each peptide to form a chemical link to the activated surface in an oriented manner. The sequence of peptides can be a true epitope of that antibody or a serve as a mimotope of the actual epitopes due to the cross-reactive properties of antibodies [161].
Each peptide was spotted to an array surface in an addressable manner, so the sequence of each spot is known. We spotted 10,000 peptides on an array to cover a substantial portion of the molecular recognition space of all circulating antibodies. The aim of this design is to provide chemical complexity on an array to allow an unbiased display of antibody binding.

1.4.1.2 Probing Sera

To generate an immunosignature, a drop of blood or serum (2ul) is diluted at 1:500 in buffer. At this dilution, essentially only the antibodies from the serum bind the surface. The diluted serum is then applied to the 10,000 peptides array.

After applying the diluted serum on an array, we incubate the array for one hour and then wash it to remove unbound antibodies. Antibodies that are bound to peptides are detected by a secondary antibody and fluorescence dye. The array is then washed, dried and scanned to determine the array’s antibody binding profile. Different secondary antibodies can be used for specific isotypes of interest, which could provide more layers of information about an individual’s immune status.

1.4.1.3 Analysis of Binding Pattern

We read this pattern with our peptide microarray, which captures enough information about a patient that the health status is legible. Once the arrays are probed with patient sera, the immunosignature of each disease state is determined by selecting peptides that show common reactivity among patients with the disease of interest and different reactivity in controls. Selected peptides are used to train classification algorithms to evaluate diagnostic efficacy and classification error.
1.4.2 Advantage of Immunosignature

As mentioned in the previous section, we can find the immunosignature that is unique to each disease and is stable over time [160]. The main reason for this advantage is that antibodies are stable in both circulating blood and stored sera samples. B cells that generated those antibodies can exist years after the immunogenic event has occurred, with a lifespan of up to four-and-a-half months [107, 108]. Moreover, after sera separation, antibodies are stable for several years in solutions [111, 112]. This means that we can use archived blood samples stored in sera or a standard filter spot [113].

The random peptide array is universal. We designed the algorithm of the random number generator to evenly cover as many combinations of amino acid sequences, potentially epitopes, as possible. These peptide sequences were not based on protein sequences in nature. The aim of this design is to provide chemical complexity on an array to allow an unbiased display of antibody binding. In this way, we can use same array to make diagnostic assays of other diseases or any condition that can be reflected in a circulating antibody repertoire.

Immunosignature technology also does not require pre-knowledge of a complex disease. Unlike PCR, ELISA and protein arrays that require isolation of molecular targets to form an assay, we use random peptide sequences instead of known protein or epitope information. We can investigate diseases even if we do not known their antigens. The sequence of a peptide can be a true epitope of that antibody or serve as a mimotope of the actual epitopes due to the cross-reactive properties of antibodies [161]. We have shown that antibodies generated against
multiple types of targets, such as proteins and sugar, can bind peptides on the array [162]. This gives immunosignature the advantage of dealing with an outbreak of an unknown pathogen, and we can use same array for other diseases in different species. Although the peptide sequences are randomly generated, we can “decipher” diseases’ specific immunosignature based on sequencing analysis to guide us find real epitopes at times [163, 164].

Immunosignature technology can be adapted to have multiple antibody isotypes measured simultaneously on an array. This will provide orthogonal measurements of a disease. Studies have shown that multiple isotypes can increase diagnostic performance [19, 109, 110, 113, 160]. We have also routinely detected both IgG and IgM on the same array in other studies.

Because the immunosignature measures a host’s humoral response to a disease, we can evaluate how the host’s immune system responds to that disease. Some patients may have a high titer of protective antibodies and may not need any treatment later. We can use immunosignature technology to identify those patients and devote limited medical resources to the highly vulnerable patients who need treatment.

In addition, the arrays are inexpensive and can be used to assay large numbers of samples quickly.
1.4.3 Published Studies with Immunosignature

Immunosignature has been demonstrated in multiple diseases (infectious and chronic) or immunizations and in animal models (multiple mouse models), dog and human samples. We have demonstrated that the resulting immunosignature is unique for each disease that we tested and is stable over time [110, 160, 165-173].

Take valley fever (Coccidioidomycosis) as an example of an infectious disease. We have demonstrated that by using a training set from 55 infected individuals and 55 uninfected individuals, we can achieve 100% prediction accuracy in an independent testing set. The immunosignature can detect the disease in samples from patients at a point in time when the standard ELISA test shows no titer, but the patient later develops valley fever [174].

Also in chronic diseases such as cancers, immunosignature can detect the cancer-specific signature for each type of cancer. We have used a training cohort with five cancers (20 samples for each cancer) and 20 non-cancer samples to generate reference immunosignatures to distinguish each disease. The immunosignatures gave 95% classification accuracy in a blinded test with 120 blinded samples with the same disease structure [175].

In Alzheimer’s disease, Lucas Restrepo has developed immunosignature to detect the disease in both mouse-model and human samples [169].

1.5 Breast Cancer Treatment

1.5.1 Clinical Treatment of Breast Cancer
Current clinical breast cancer treatment includes three main methods: surgery, radiation, and drug treatment, including chemotherapy, hormone therapy, and targeted therapy. More recently, great achievements have been made in cancer immunotherapy, which may become an important part of future cancer treatment. Adjuvant therapy is a way of providing patients with more treatment after surgical removal of cancers. Neoadjuvant therapy is different in that a systemic treatment or radiation is giving before surgery to decrease tumor size. Healthcare professionals and medical researchers have investigated many different strategies for cancer treatment, and a combination of different treatments instead of a single one is the trend in the future.

1.5.1.1 Surgery and Radiotherapy

Removing the tumor through surgery remains the major way that breast cancer is treated. Surgery for cancer has been performed since the early 19th century and was later combined with anesthesia and antisepsis in the following half century. Surgery remained the only option before the appearance of radiation by the middle of the 20th century.

The purpose of surgery is to remove as much of the cancer from a patient as possible. By using imaging technology, more-accurate surgery can be performed to almost completely remove the cancer cells and preserve the healthy tissues. Based on tumor size, extent of spread and the patient’s preference, different surgeries can be performed.

Breast-conserving surgery (BCS): BCS, also called partial mastectomy, removes only a part of the breast, depending on the size and place of the tumor
and others. The removed breast tissue will be examined to see whether the tumor has spread to the edges. If it has, more surgery is needed to completely remove all tumor tissues. BCS is often followed by radiation to kill any remaining tumor cells.

Mastectomy is a surgery that removes the entire breast and, sometimes, nearby tissue. If only the breast is removed, it is called a simple mastectomy. When a simple mastectomy is combined with an axillary lymph node dissection, it becomes a modified radical mastectomy. Axillary lymph nodes will be examined for any tumor cells. A sentinel lymph node biopsy examines the first lymph nodes to which cancer may spread. Studies have shown that for early non-invasive breast cancers, long-term survival is similar between BSC with radiation and mastectomy.

The discovery of X-ray in 1895 by Wilhelm Roentgen made cancer radiation therapy possible. Radiation therapy uses high-energy rays such as x-rays or particles to kill cancer cells and shrink tumor size in the human body. External beam radiation uses a machine outside of the body to emit radiation. Brachytherapy is the placement of radioactive pellets into the breast tissue to kill targeted tumor tissues. Its long-term results may not be as good as those with external beam radiation. With the development of better radiation machines and computer imaging technology, radiotherapy developed significantly during the 20th century. This contributed to a 30% increase in the cancer cure rate in the 1950s, when radiotherapy was combined with surgery [176].
1.5.1.2 Cancer Drugs

Besides surgery and radiotherapy, breast cancer treatment can also use drugs to kill cancer, including chemotherapy, hormone therapy, and targeted therapy. These drugs can be used to prevent cancer or treat diagnosed cancer.

1.5.1.2.1 Chemotherapy

Chemotherapy is the use of cancer-killing drugs. Drugs enter the human body and spread to different organs by blood circulation. Chemotherapy can be used before or after surgery for early-stage breast cancer. When chemotherapy is performed before surgery, it can shrink tumor size so that only BCS is needed instead of mastectomy. It can also decrease the risk of recurrence after surgery. Chemotherapy is given in cycles, with each cycle of a treatment followed by a rest period. For early-stage breast cancer, a normal course of treatment may last for three to six months. However, if necessary, the chemotherapy can be continued as long as it is effective.

After the use of chemotherapy in curing childhood leukemia and advanced Hodgkin’s lymphoma in the 1960s [177, 178], more drugs were discovered to treat major types of cancer. Chemotherapy has greatly improved patients’ outcome. However, chemotherapy has side effects, including menstrual changes, nerve damage, heart damage, nausea, hair loss and others. Because chemotherapy does not specifically target tumor cells, it will also affect normal cells. Diagnostic tests have been developed to determine which patients will benefit most from chemotherapy. MammaPrint is a gene expression assay by Agendia to detect a 70-genes profile by reverse-transcriptase PCR. It uses 117 patients with axillary
lymph node-negative primary breast cancer to find genes highly correlated with a short interval from primary tumor to distant metastases. MammaPrint profiling can help doctors to identify the sub-population of patients that will benefit most from chemotherapy [50].

1.5.1.2.2 Hormone Therapy

Hormone receptors—such as those for estrogen and progesterone—are on the surface of cancer cells. They promote the growth of the cancer. Hormone therapy works by blocking the hormone receptors in the receptor-positive tumor cells. It has been used to reduce the risk of recurrence after surgery and to treat advanced breast cancer. Hormone therapy may include tamoxifen, toremifene, and fulvestrant, which are estrogen blockers. Anastrozole, exemestane, and letrozole are aromatase inhibitors. Hormonal therapy is beneficial for women with early-stage breast cancer that is positive for hormone receptors. A diagnostic assay has been developed to identify the subpopulation of patients that benefit most from hormone therapy. Oncotype DX detects a 21-gene profile by reverse-transcriptase PCR to predict the risk of recurrence in patients taking tamoxifen. Sixteen of the 21 genes are cancer-related genes with reference genes as reference. A mathematical algorithm is derived from an empirical retrospective study to calculate a score to estimate the risk of distant recurrence [49].

1.5.1.2.3 Targeted Therapy

Specific genes targets have been found to promote breast cancer development. New drugs that specifically target those candidates are called targeted therapy. Trastuzumab, lapatinib and pertuzumab are targeted drugs that
are available to treated women with breast cancer overexpressing the growth-promoting protein HER2. Most of these drugs are humanized monoclonal antibody or small-molecule kinase inhibitors that bind to different epitopes on the extracellular domain of HER2 receptors and inhibit its dimerization. Because tumor growth requires blood vessel formation to supply nutrition, antiangiogenesis drugs that prevent blood vessel growth in breast cancer are also tested in clinical trials. Chemotherapy, hormone therapy and targeted therapy have improved many patients’ outcomes. However, the problem is that after the primary tumor has been killed, the remaining sub-population of cancer cells will develop resistance to the initial treatment and can result in recurrence.

1.5.2 Emerging Strategy of Immune Checkpoint Blockade

The immune system in our bodies protects us from pathogens and abnormal conditions. In the immune-editing theory, immune surveillance inhibits tumor development from tumor initiation, and the tumor evolves through multiple mechanisms to avoid immune system surveillance [86, 87]. Much effort has been devoted to leverage immune system to fight against disease.

William Coley performed the first clinical trial of cancer immunotherapy in 1891—Coley’s toxins. It was used to treat erysipelas with live or attenuated bacteria [179]. However, the efficacy of Coley’s toxins was controversial. Other immunotherapy drugs have been developed to target cancer-associated proteins and to induce tumor cell apoptosis, phagocytosis and antibody-dependent cellular cytotoxicity. These will trigger the adaptive immune system to target cancer [180]. Immunotherapy using immune checkpoint inhibitors has been promising in recent
years. Immune checkpoint inhibitor therapy is based on the biology of immune suppression of the immune system by tumor cells. One mechanism of immune-editing and surveillance is that tumor cells can suppress immune surveillance. Tumors can express immune suppression factors that inhibit immune infiltration and recruit myeloid-derived suppressor cells (MDSCs) and Regulatory T cells (Tregs) to the local tumor environment to inhibit the normal function of immune cells[181-186]. This will result in apoptosis of activated anti-tumor immune cells and induce tolerance.

Among those, Tregs play an important role in the battle between tumor cells and the immune system [187, 188]. Tregs constitutively express CTLA-4 [189, 190], which bind to CD80 and CD86 on the surface of antigen-presenting cells (APCs). The binding of CTLA-4/CD80 or CD86 blocks the co-stimulatory signal transduction and results in immunosuppression (Figure 1.6) [191-193]. Inhibitory receptor programmed cell death 1 (PD-1) contributes to immune tolerance of self-antigens and is expressed in many tumor-infiltrating lymphocytes, such as natural killer cells, dendritic cells, activated monocytes, B cells and T cells [194].
Figure 1.6: Introduction of CTLA-4 and PD-1 Treatment
Figure adapted with permission from Merelli et al [195].

The binding of PD-1/PD-L1 is an important mechanism to convey an inhibitory signal to T cells and let tumor cells evade the immune system (Figure 1.7) [196, 197]. Many clinical trials using antibodies targeting PD-1 or PD-L1 have shown promising results in improving the survival rate in human prostate cancer, melanoma, and breast cancer [198, 199]. Expression of programmed death ligand 1 (PD-L1) has been found in different types of cancer, such as non-small cell lung carcinoma [200], esophageal cancer [201], pancreatic cancer [202] and breast cancer [203]. Ghebeh and colleagues reported that PD-L1 is expressed in 34% of primary breast cancer samples but not in healthy controls. The expression level of PD-L1 is significantly correlated with grade 3 tumors [203, 204].
With a combination of different therapy methods and the discovery of novel therapies, the mortality rate of many types of cancer has decreased during the past decades. However, due to the complexity of metastasis and immune suppression in late stages, a diagnostic tool that can detect tumors early will provide more therapy options and eventually improve patients’ outcomes.
CHAPTER 2

Early Detection Of Breast Cancer

2.1 Introduction

In the United States in 2014, 235,030 new breast cancer cases were diagnosed, and there were 40,430 breast cancer deaths, making breast cancer the second leading cause of death among females [205]. Currently, the standard diagnostic tool for breast cancer is mammography. However, with its low sensitivity and specificity and its limitation to a specific population, mammography detects only 70% of breast cancers [33, 36, 37, 53]. The five-year relative survival rate decreases from 100% to 22% when breast cancers are detected at stage 4 instead of stage 1 [206]. Therefore, there is an urgent need for early diagnosis of breast cancer, as it could improve disease outcome and be the key to decreasing the death rate from breast cancer. This study proposes a serum test called immunosignature to identify specific antibody signatures to diagnose breast cancer early.

Serum proteins, circulating miRNAs, circulating tumor cells (CTCs) and glycan biomarkers have been investigated to discover breast cancer biomarkers to diagnose the disease [72, 207-210]. Current plasma protein biomarkers in the human proteome organization (HUPO) project are in the range of ug/ml to mg/ml and the concentration of target molecules shed from a small number of tumor cells into the blood is 104 fold too low to be detected by current technology [100-102]. This situation is especially true in early-stage breast cancer, when only a small number of tumor cells initiate and secrete molecules. The challenge for the
foregoing approaches is the detection of low-abundance valuable biomarkers of early stage breast cancer.

Circulating serum antibodies could provide a comprehensive reflection of patients’ health status. Immunosurveillance of B cells and T cells begins in the early stage. Tumor-specific B cells can be activated and amplify the antibody level, making the detection of a minimal exposure of antigen at the early stage feasible. Pre-diagnostic antibodies have shown value in the early detection of human breast cancer [19, 211, 212]. Chapman and colleagues found an antibody response to six antigens (c-myc, p53, HER2, NY-ESO-1, BRCA2 and MUC1) in early-stage primary breast cancer patients, although with low sensitivity [213]. Different panels of biomarkers all proved that a tumor-specific antibody response exists in early-stage breast cancer, such as the carcinoma in situ stage [117, 118]. Multiple antibody-based approaches have been explored to leverage the specific amplification effect of the immune system in the early-stage breast cancer.

Serological screening of the recombinant cDNA library (SEREX) uses phages to display proteins from a recombinant cDNA library of interests and probed with tumor patients’ sera of interest to select antigens with high binding interactions with tumor related antibodies in the sera [122, 125, 126]. Serological proteomic analysis (SERPA) separates proteins from cell lines or tissue lysates into spots by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) [143]. Selected spots will be analyzed by mass-spectrometry to find protein biomarkers for breast cancer. Protein arrays printed with pre-synthesized recombinant proteins or directly synthesizing proteins by an in-vitro expression system can also be used to
screen autoantibodies biomarkers in a high-throughput way. However, SEREX and SERPA require large amounts of serum for profiling, and the initial construction of the cDNA library in SEREX can easily cause biased result. Proteins may not be folded and modified correctly in all three methods. More importantly, these approaches can detect only antibody responses that target a limited number of known proteins in each system. This cannot capture neo-antigens by mutations or post-translational modification.

Here, we propose an unbiased antibody profiling technology using high-density peptide microarrays for the early detection of breast cancer. The technology, called immunosignature (ImS), uses a peptide array with 10,000 spotted peptides selected without bias from a non-biological peptide sequence space. The assay consists of applying diluted sera to the array and then detecting the binding pattern of the primary antibody with a labeled secondary antibody. In previous work, we successfully used immunosignature technology to find antibody signatures in early-stage pancreatic cancer and Alzheimer’s disease in human patients [168, 175, 214]

This murine model develops invasive mammary carcinomas and can recapitulate the morphologic, pathologic, and molecular features of human luminal breast cancer at many stages [9]. We hypothesize that the FVB/N neuN mouse model can recapitulate human humoral responses to breast cancer and provide insights into the dynamic antibody profile from the early to late stages of breast tumor development. It can test whether, in principle, immunosignature can detect early stage of cancer or not.
We have found immunosignatures of early stage tumors as early as 12 weeks before the tumor is first palpable. The existence of different immunosignatures at different stages reflects the dynamic changes in the antibody profiles during tumor development. By using the immunosignature at each stage, we achieved high sensitivity and specificity by cross-validation in multiple classification methods. Applying the late-stage immunosignature to early-stage samples resulted in low-prediction performance. This implies that the late-stage tumor immunosignature cannot be used to develop the early-stage signature. This study could determine the potential application of immunosignature technology in preclinical diagnosis of early-stage breast cancer.

2.2 Results

2.2.1 HER2 Transgenic Mouse Model

This study uses FVB/N neuN transgenic mice, a well-characterized murine model for HER2 breast cancer, to compare antibody profiles at different stages of breast cancer. Mice are engineered with neuN with an MMTV promoter that restricts the neuN gene from being expressed in mammary glands. We collected blood samples from 23 transgenic mice and ten wild-type mice aged 12 weeks to terminal age. The mice were examined every week for any tumor formation, and any tumors found were measured once a week. Among the 23 transgenic mice, the first tumor was detected, on average, at 33 weeks of age (Fig. 2.1a) with an average tumor size of 96.3 mm³. We defined the point in time when the first palpable tumor was found in a mouse to be window 0 week. A window is used to calculate intervals between a certain age and the age at which the mouse had the
first tumor detected. The phenotypes of tumors at different stages are well
characterized into hyperplasia, dysplasia and carcinoma in situ, multiple foci of
carcinoma in situ or small lobular carcinoma, and carcinoma. Studies from
Boggio and Abe have shown a transition from normal mammary glands to
hyperplasia around 8.6 to 15 weeks [23, 215]. Figure 1b shows the tumor growth
curve of this mouse model. The average tumor size and its standard error are
plotted against days after the first palpable tumor was detected.

Figure 2. 1: Tumor Free and Tumor Growth Curve of FVB/N NeuN
Transgenic Mouse Model
23 FVB/N NeuN transgenic mice and 10 FVB/N wild type controls were
monitored every week for any palpable tumor in mammary glands. The median
tumor free time is at 33.1 weeks with an average tumor size of 96.3 mm³. 94.3%
of FVB/N NeuN transgenic mice will develop palpable tumors in 47.6 weeks
monitoring.
2.2.2 Scheme of Experiment Design

Figure 2.2: Scheme of Experiment Design

The scheme of the early diagnosis of breast cancer was layout in Figure 2.2. Serum samples from FVB/N transgenic mice and wild type controls were collected every two weeks during tumor development. Mice were palpated every week to find any tumor occurrence. Serum samples were categorized into four stages: ‘Stage-1’ for samples at 12 weeks before 1st tumor, ‘Stage-2’ for samples at 8 weeks before 1st tumor, ‘Stage-3’ for samples at 4 weeks before 1st tumor, ‘Stage-4’ for samples at 1st tumor. For each time point, transgenic mice were compared with age-matched wild type controls to find significant peptides that passed statistic test.
2.2.3 Influence of Genetic Background on Tumor-specific Immunosignature

In order determine the extent to which transgenes might affect the ImS, we compared a sample set of seven transgenic mice at 12 weeks and seven age-matched wild-type controls. Principle component analysis of these samples did not reveal clustering by groups (Fig. 2.3a). Fig. 2b shown a Heatmap analysis of four transgenic mice and six wild-type controls at 12 weeks. Although transgenic and wild type mice were not mixed after hieratical clustering, the intensity of 833 selected peptide did not show distinguishable pattern between two groups (Fig. 2.3b). No peptides could differentiate the two groups by hypothesis test in this ten-sample set with a two-tail t-test with p value cut off at 0.05 (Fig. 2.3c). An independent sample set of seven transgenic mice and nine wild-type controls resulted in 37.5% accuracy using the top 833 peptides with the highest p value in the previous set of seven transgenic and seven control mice (Fig. 2.3d). We concluded that the ImS was not influenced by the transgenes before the tumor activated.
Figure 2.3: Influence of Genetic Background on Immunosignature
A) Principle component analysis of 7 transgenic and 7 wild type mice at 12 week age by all peptides.
B) Heatmap of 4 Tg and 6 WT at 12 weeks age by 833 peptides with top p value by t-test without MCC. No distinguishable patterns between Tg and WT were observed.
C) Hypothesis test of 4 Tg and 6 WT at 12 week age. No peptides pass T-test. 1% peptides have P<0.01. 0.5% have <1.3 fold change.
D) Classification result by 7 Tg and 9 WT at 12 weeks age by 833 peptides with top p value by t-test without MCC gives 37.5% accuracy.

2.2.4 Existence of Pre-tumor Immunosignature Prior to Tumor Palpation

To test whether immunosignature technology can detect changes at the pre-palpable tumor stage, we used eight transgenic mice (two weeks before the first tumor) and eight wild-type mice samples. We were able to detect a significant difference in antibody binding of 51 peptides (p≤0.05 with Benjamin and Hochberg FDR) between the transgenic and wild-type samples (Fig. 2.4a). The 51 peptides signature was further validated using an independent set of samples, including eight transgenic and eight wild-type mice. By using K-means to classify, the 51 peptides gave 100% sensitivity and 66.7% specificity (Fig. 2.4b). Two experiments with different peptide libraries and samples all showed
that significant peptides could be found between pre-tumor transgenic samples and wild-type controls. The average age of these pre-tumor samples were greater than 17 weeks, at which time the mammary glands had already progressed into carcinoma in situ. We concluded we can use immunosignatures to segregate transgenic and wild-type mice before the first palpable tumor was detected.

**Figure 2. 4: Immunosignature Can Show a Separation between Transgenic Mice and Age Matched Controls Two Weeks before First Palpable Tumor**

A) Independent sample sets with 8 transgenic mice (2 weeks before 1st tumor) and 8 wild type mice samples using CIM 10K Version 3 peptide library. 51 peptides (p≤0.05 with Benjamin and Hochberg FDR) were significant between transgenic and wild type samples. Principle component analysis shows separation.

B) Cross validation result of training sample and prediction performance in an independent samples gives a sensitivity of 87.5% (7 right in 8 cases) and specificity of 87.5% (7 right in 8 cases).

**2.2.5 Earliest Time Point of the Detectable Immunosignature**

After showing the existence of immunosignature at two weeks before first palpable tumor, we were interested in how early immunosignature could detect tumor. In order to test this, we used mice at 12 weeks, and 16 weeks to estimate the earliest time point in CIM-Version 2 arrays. Four transgenic and seven wild type mice at age of 12 weeks were ran in the same batch of slides (Fig. 2.5). No
peptide passed T-test with p value cutoff at 0.05. At age of 16 weeks, fifty three peptides were significant between five transgenic and seven wild type mice. The lack of difference between transgenic mice and wild type control at age of 12 weeks and existence of significant difference at age of 16 weeks implied the earliest detectable time point for breast tumor is between 12 to 16 weeks by immunosignature in CIM-Version 2 arrays.

**Figure 2. 5: Heatmap of Immunosignatures of Different Time Points in Tumor Development**

Four NeuT transgenic FVB/N (Tg) mice and seven wild type FVB/N (WT) mice are assayed at age of 12 weeks and 16 weeks. Each colored square represents relative median intensity. Blue: low expression; Red: high expression; Yellow: average expression.

2.2.6 Time Course of Immunosignatures During Tumor Progression in FVB/N neuN Mice

Previous studies have shown that tumors present different tumor antigens at different stages [86, 87]. Also, biomarkers that were selected using late-stage patients gave poor prognostic performance in an early-stage setting. These facts imply that the immune system develops different antibody profiles along with the
development of a tumor. The immunosignature is a representation of the antibody profile at a given point in time. Therefore, we were interested in comparing immunosignatures from different stages to quantitatively investigate differences in antibody profiles. In this experiment, we selected 23 transgenic mice with ten wild-type controls from previous populations and collected serum at multiple points in time. Both the transgenic and control samples were grouped into four stages by the numbers of weeks before the first palpable tumor. We defined ‘IS-1’ as 12 weeks before the first tumor; ‘IS-2’ as eight weeks before the first tumor; ‘IS-3’ as four weeks before the first tumor; and ‘IS-4’ as the time of the first tumor. Wild-type mice (aged 15.3 to 45.4 weeks) were used to make between-stage comparisons comparable and to eliminate age influence.

By using 23 transgenic mice and ten wild-type controls, we selected the top 200 significant peptides using an FDR-corrected T-test with a larger than 1.3-fold change between transgenic and wild-type mice for each point in time (Fig 2.6). To visualize the differences between transgenic mice and the wild-type controls, we plotted heatmaps containing these top 200 peptides between the transgenic and wild-type groups.

We observed a change in immunosignatures from a greater to a less signature difference between the wild-type and transgenic mice from early to late stage tumors (Fig 2.6). This demonstrates that immunosignature can capture tumor development. The IS-1 heatmap showed two clear clusters of transgenic and wild-type mice. The majority of the peptides were highly reactive in the wild-type controls and low reactive in the transgenic mice. This pattern reversed in IS-
2, when the majority of peptide intensities were higher in transgenic mice. In IS-3 and IS-4, the wild-type mice began to be clustered into the transgenic group. Distinguishing patterns of peptides intensities in transgenic and wild-type mice became less obvious in later stages of the tumor, while small clusters in these two groups appeared instead.

**Figure 2.6: Time Course of Immunosignatures during Tumor Progression in FVB/N NeuN Mice**

23 NeuT transgenic FVB/N (Tg) mice are assayed at four time points - 12, 8, 4 weeks before the first tumor and at the first tumor. 10 wild type FVB/N (WT) mice are assayed. We select the top 200 most significant peptides using FDR-corrected T-test for each time point. Heatmap and PCA using each stage signatures are used to show the difference between transgenic mice and wild-type controls. Significant peptides are judged by p<=0.05 with fold change >=1.3. Each colored square represents relative median intensity. Blue: low expression; Red: high expression.

The Venn diagrams show the overlap of immunosignatures at the four stages (Fig. 2.7). We found that different tumor stages have different and partially overlapping signatures. The adjacent stages share more peptide overlap (Fig. 2.8).
The immunosignatures of IS-1 and IS-2 shared 25 peptides. The immunosignatures of IS-2 and IS-3 shared 35 peptides, while the immunosignatures of IS-3 and IS-4 shared 90. Sixteen peptides were shared by all four stages. These numbers are larger than the overlaps expected by peptides chosen at random. The number of overlapping peptides between the earliest stages (IS-1) and all the following stages remained consistent, but the number increased when comparing the latest stage (IS-4) with all the previous stages (Fig. 2.9). Taken together, this also shows that immunosignature can detect differences between transgenic mice and healthy control mice caused by tumor development at multiple stages.
Different stages in tumor development have different immunosignatures, a potential reflection of antigens profile changes by immunoediting in tumor development. 200 significant peptides with highest fold change between transgenic and control group are selected at four stages, corresponding to 12 weeks before, 8 weeks before, 4 weeks before and at 1st tumor. Peptides shared among 4 stages take <10% in any stage.

Figure 2. 7: Venn Diagram of Different Immunosignatures Along Tumor Development

Figure 2. 8: Numbers of Selected Peptides Overlap between Two Adjacent Stages during Tumor Development
Recent biomarker failures implied that using late-stage biomarkers for early prediction leads to low performance. Here, we tried to quantify cross-stage performance by using an immunosignature from one age to predict another. We use SVD in a PAMR package to predict twelve combinations of any two of the
four ages (Table 2.1). Using 200 peptides from IS-1 to predict IS-2, IS-3, and IS-4 samples gave 0% accuracy. Using IS-2 to predict IS-1, IS-3, and IS-4 samples gave 0%, 4.2% and 8.3% accuracy, respectively. Using IS-3 to predict IS-1, IS-2, and IS-4 samples gave 28.6%, 58.3% and 87.5% accuracy, respectively. Using IS-4 to predict IS-1, IS-2, and IS-3 samples gave 9.5%, 41.7% and 75.0% accuracy, respectively. This implies that late-stage tumor immunosignature cannot be used to develop an early-stage signature.

### Table 2.1: Cross-Stage Prediction Performance in the Training Set

Using peptides selected from 10 week before, 4 weeks before or 6 weeks after first tumor to predict samples from another time points. Accuracy is calculated by the percent of correct prediction of both transgenic and wild type mice in both groups.

<table>
<thead>
<tr>
<th></th>
<th>-12 wks</th>
<th>-8 wks</th>
<th>-4 wks</th>
<th>1st tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>-12 wks to predict</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>-8 wks to predict</td>
<td></td>
<td>4.2%</td>
<td>8.3%</td>
<td>87.5%</td>
</tr>
<tr>
<td>-4 wks to predict</td>
<td>28.6%</td>
<td>58.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st tumor to predict</td>
<td>9.5%</td>
<td>41.7%</td>
<td>75.0%</td>
<td></td>
</tr>
</tbody>
</table>

### 2.3 Methods

#### 2.3.1 Mice and Sample Collection

Female FVB/NJ mice (n=10, age 6 weeks) were purchased from Jackson Laboratories (Bar Harbor, ME). Female FVB/N neuN transgenic mice were kindly gifted from Dr. Chella David at Mayo, Rochester. Transgenic FVB/N NeuN mice and wild-type FVB/N mice were maintained and weaned in the same barrier isolation housing with standard rodent feed and water. Mice were palpated every week to detect mammary tumor growth. Blood samples were collected by submandibular venipuncture using a 5.0 mm lancet (MEDIpoint, Inc. NY). Serum
was separated by serum separating Microtainer (BD Biosciences, San Jose, CA) and stored at -20 C with aliquots. Mice were euthanized by 2,2,2-tribromoethanol at a dose of 125mg/kg through intraperitoneal injection when tumor size exceeded 10% of the body weight. All murine experiments were conducted following animal protocol reviewed and approved by Arizona State University Institutional Animal Care and Use Committee.

2.3.2 Preparation of Microarray

CIM10K peptide arrays were printed as earlier described [110, 160]. Two libraries of peptides were used separately: CIM10K Version 2 and Version 3. Both libraries contained 10,000 random 20-residue peptide of 17 random sequence amino acids. Peptides were designed with random sequences with nineteen amino acids (cysteine excluded), except for three residues at N-terminal as a linker to attach peptides covalently on Nexterion A+ aminosilane slides (Schott). The CIM10K Version 2 library contained L amino acid with Gly-Ser-Cys-NH2 linker. Version 3 contained both L and D amino acids with Cys-Ser-Gly-NH2 linker. Peptides were synthesized by Sigma Genosys (St. Louis, MO). Slides were printed using the two-up format by piezoelectric non-contact printer at Applied Microarrays (Tempe, AZ). Quality-control tests were performed using subsamples of each batch with pooled human naïve serum to select high array-to-array correlation and dynamic range. After being produced, arrays were stored in containers, protected from light.
2.3.3 Probing Serum Antibodies on Peptide Microarrays

Slides were first washed by 33% isopropanol, 7.5% acetonitrile, and 0.5% trifluoroacetic acid for five minutes, then dried by centrifuge at 800 rpm for five minutes. After pre-wash, slide processing was automated by a TECAN HS4800-Pro automated incubator, following a protocol optimized for antibody binding. Briefly, arrays were blocking by 0.014% mercaptohexanol, 3% BSA, and 0.05% Tween 20 in 1×PBS solution for 1 h at 23° C. Serum samples were then diluted at 1:500 in incubation buffer (3% BSA 0.05% Tween 20 in 1×PBS) and probed on arrays for one hour at 37° C. Biotinylated anti-mouse IgG(H+L) antibodies (Bethyl Laboratories, Montgomery, TX) were diluted in the incubation buffer at 5nM for one hour at 37° C. Bound secondary antibodies were visualized by Alexa Fluor-647 labeled streptavidin (Invitrogen, CA) at 5nM for one hour at 37° C. The injection volume of each step was 110 μl per array, and TBST solution was used to wash the array between steps. The final wash used TBST and distilled water to remove residual salt on slides. Correlation for technical replicates less than 0.85 were re-probed [175].

2.3.4 Microarray Data Processing

After the final wash, slides were scanned by an Agilent C scanner at 633nm emission (Agilent Technologies, CA) with 100% laser power and 100% PMT. Images were extracted by GenePix Pro 6.0 (Molecular Devices, Sunnyvale, CA) and saved as gpr files. Poor-quality peptide spots were excluded by flagging them as “absent” by visual inspection. Samples with Pearson correlation coefficient >0.85 among technical replicates were used. Raw data from gpr files
were collected and imported into GeneSpring GX 7.3 or R for later processing (Agilent Technologies, Palo Alto, CA). Data were normalized to the 50th percentile to minimize array- to-array variation. Signal intensities below 0.01 were set to 0.01. Batch effects were adjusted by SVA algorithm with irw as adjustment method with 10 iterations. The biological group was used as a covariate in SVA.

2.3.5 Statistical Analysis

Statistical analyses were done using Genespring GX 7.3 (Agilent Technologies, Palo Alto, CA), R version 3.1.1, and JMP version 11. P-values were corrected for multiple testing by the Benjamin-Hochberg false discovery rate method. The Welch t test was used in R to select significant peptides with greater than 1.3-fold changes and p<0.05 between two groups. We used a support vector machine with settings including polynomial dot order at 1 and scaling order at 0 for cross-validation and prediction. Packages used in R were sva, caret, e1071 [216, 217].

2.4 Discussion

In this study, we questioned the feasibility of breast cancer early detection by immunosignature technology. In comparing transgenic mice and their non-transgenic littermates, we found the tumor-specific immunosignature of significantly different peptides as early as 12 weeks before a tumor became palpable. We also tested four more time points at early stages between transgenic and wild-type groups and found the immunosignature of significant peptides for each stage during tumor development. Fewer than 10% of the peptides were
shared among four time points. Using early-stage immunosignature to predict later-stage sample results in low classification performance.

Early detection of breast cancer has the advantage of giving patients early treatment intervention and improving their outcomes. In our previous work, we successfully used immunosignature technology to find early antibody signatures in Alzheimer’s disease in a mouse model [168]. Here, we wanted to test whether we could detect the immunosignature of the early stages of breast cancer. However, serum samples from the same cancer patients before and after clinical diagnosis are hard to get. It is difficult to study biomarkers of early-stage breast cancer without proper early-stage samples. The FVB/N neuN transgenic mouse model is a well-established model to study breast cancer. In order to cover samples at the early stages of tumors, we collected serum samples for each transgenic mouse at time points from zero to twelve weeks before the first palpable tumor—aged 12.36 to 45.4 weeks. In this mouse model, hyperplasia occurred at around 8.6 to 15 weeks [23, 215]. So, selected samples could have good coverage of early stage breast cancer. Samples at multiple time points were tested to see whether we could find the immunosignature that distinguishes transgenic mice from wild-type controls.

In human clinical trials, Lu and colleagues showed that autoantibodies against HER2 and p53 can be detected by protein microarrays 150 days before a breast cancer diagnosis. Opstal-van Winden and colleagues also showed that breast cancer autoantibodies can be found in patients’ samples a median of 21.3 months before diagnosis. These findings proved the feasibility of early diagnosis
of breast cancer. However, a few protein biomarkers that were discovered in late-
stage cancer were evaluated each time, and their samples did not cover a wide
range of early-stage breast cancers in humans. Using the same mouse model,
Jianning Mao and colleagues used serological screening of cDNA expression
libraries to identify tumor-associated autoantibodies in serum one month prior to
the development of a palpable tumor. Six proteins were selected [19]. They found
that autoantibody responses of six proteins appeared in mice as early as 20-30
weeks of age. In our study, the earliest time point at which we found a difference
by immunosignature was between ages 12 and 16 weeks, which is still during the
time of hyperplasia development in this mouse model. This found
immunosignature is more sensitive to the detection of a minimal antibody
response in early-stage breast cancer. Our study also used more time points to
reflect antibody responses in early-stage breast cancer more comprehensively.

When we compared immunosignatures at different stages, the early
signature was quite different from the late one. Antibodies that specifically target
antigens that emerge in the late stages do not exist at an earlier stage. Fewer than
10% of the peptides overlapped the four stages. 34 significant peptides
overlapped Stages I and IV, while 90 peptides overlapped Stages III and IV. This
showed that the binding pattern of the extra 56 peptides had been generated since
Stage III. In our previous study of Alzheimer’s disease, we also found that the
early-stage was different from the late-stage immunosignature. These findings
support the immunoediting theory of new antigens generated by tumors during an
active immunoediting process.
Several studies have shown that biomarkers discovered in established tumor patients did not predict well in the early detection of breast cancer. Lu and colleagues used eight antigens to detect breast cancer samples 150 days before diagnosis. HER2 and p53 among the eight antigens gave a ROC score of 0.6 and 0.63, respectively. Opstal-van Winden and colleagues also showed that markers discovered in diagnostic samples might not be valid in pre-diagnostic samples. Here, we use late-stage immunosignature to predict early-stage samples, and the performance of cross-stage prediction decreased significantly compared with cross-validation within each stage. Our findings suggest that using biomarkers discovered in diagnostic samples may be not valid in pre-diagnostic samples. They further suggest the importance of using proper pre-diagnostic samples to discover early-detection biomarkers.

In terms of intensity level, the majority of peptides in early breast cancer sera are less reactive than non-cancer controls. We hypothesize that the down regulation that of antibodies may be caused by immunosuppression. Edward Y. Woo and colleagues reported CD4+ CD25+ T cells secrete immunosuppressive cytokine transforming growth factor-\(\beta\) in early-stage non-small cell lung cancer [218]. In breast cancer, Isabel Poschke and colleagues observed immunosuppression in early-stage breast cancer patients in terms of the exhaustion of tumor-associated T cells [219]. A report also showed that Atypical T cell-suppressive neutrophils occurred during early tumor progression [220]. In terms of antibody response, data from Jianning Mao and colleagues showed a low reactivity of autoantibodies to four tumor-associated proteins in early-stage breast
cancer [19]. Our findings further showed that immunosuppression already occurs at early stages of tumor development in humoral response.

We also observed that the immunosignature of transgenic mice is homogeneous among different mice at Stage I but become sparsely distributed at later stages. Because the tumorigenesis of this mouse model is initiated by a single oncogene (NeuN) expression, the immune system in different mice may respond in a similar way. However, when more mutations occur during tumor development, the immune system in each mouse will respond differently to new tumor antigens. At the scope of 12 weeks, the dramatic changes in the immunosignature imply that the immune system has gone through significant changes along with tumor development.

We recognize that the limited sample size and the use of a random sequence array may not be the best conditions for finding gene and protein targets. Because serum from each mouse needed to be collected as the tumors developed, the tumor’s histology information could not be known. However, this study established the methodology and demonstrated the feasibility of early detection of breast cancer by immunosignature technology in a mouse model.

Early detection of breast cancer could lead to early treatment intervention and improved patient outcomes. Future studies can use retrospective sample collections in clinical trials to develop a specific immunosignature for early-stage breast cancer in humans. If these immunosignatures are found, we could use multiple immunosignatures at different stages in early breast cancer to predict breast cancer early and reduce the false-positive rate in cancer prediction.
2.5 Supplements

2.5.1 Selection of multiple batches of array

Due to the large number of slides that were used in this study, slides from multiple batches of printing were selected. In order to reduce the variance caused by different printing of batches, we tested the correlation of the same quality control sample in eight different batches. Statistic characterization were also calculated to evaluate the overall performance for different batches, including intensity of empty spots on a slides, ratio of median intensity of all peptides to empty spots, coefficient of variance of all peptides, correlation between two technical replicates using same quality control samples (Table 2.2, table 2.3).

From the analysis, batches from 474, 493, 494, 524, 530, 538 and 540 were selected.

Table 2. 2: Correlation Analysis for Different Printing Batches of Slides Using Quality Control Sample
First column and last row are the batch number. Square of Pearson Correlation is calculated for each comparison between eight different batches.
Table 2.3: Statistic Characterization of Different Batches of Slides
Correlation analysis for different printing batches of slides using quality control sample. Empty: intensity of empty spots on a slides; f647 M/E: ratio of median intensity of all peptides to empty spots; f647 CV: coefficient of variance of all peptides; Correlation: correlation between two technical replicates using same quality control samples.

2.5.2 Removal of Batch Effect

One important issue in microarray experiment is the variance caused by batch effect. In this experiment, seven different batches of slides were used. Principle Component Analysis (PCA) of 336 samples using all peptides with median normalization was plotted to illustrate batch effect (Figure 2.10). From the figure, we can see batch 474 and some samples in batch 493 were separated from other batches. Batch removal methods were needed to remove these difference from non-biological difference.
Our lab had employed ComBat to remove batch effects [221, 222] in previous studies. Surrogate variable analysis (SVA) is also a widely used batch effect removal method [217]. Batch removal methods need to be selected based for the specificity of different studies [223, 224]. Here, we evaluated SVA and ComBat methods in a data set from array result of 168 different mice sera samples each with two replicates (Fig 2.11, Fig 2.12, Fig 2.13). Packages from SVA and ComBat were run in R. Correlations between median normalized data and SVA or ComBat adjusted data for each sample were plotted. PCA was also plotted to show the variance between different batches after adjustment of batch effects. In

---

**Figure 2.10: Principle Component Analysis of Different Batches**

Our lab had employed ComBat to remove batch effects [221, 222] in previous studies. Surrogate variable analysis (SVA) is also a widely used batch effect removal method [217]. Batch removal methods need to be selected based for the specificity of different studies [223, 224]. Here, we evaluated SVA and ComBat methods in a data set from array result of 168 different mice sera samples each with two replicates (Fig 2.11, Fig 2.12, Fig 2.13). Packages from SVA and ComBat were run in R. Correlations between median normalized data and SVA or ComBat adjusted data for each sample were plotted. PCA was also plotted to show the variance between different batches after adjustment of batch effects. In
SVA, we can choose irw and twostep as different methods. I tested the irw method with 10 iterations and twostep method in SVA (Fig 2.11).

Figure 2. 11: Correlation of Technical Replicates Before and After SVA Adjusted and PCA Analysis of Batch Effects
Figure 2. Correlation of Technical Replicates Before and After Combat Adjusted and PCA Analysis of Batch Effects

After both irw and two-step method in SVA, batch effects were largely reduced by the visualization of PCA. Correlation between technical replicates for the same sample increased. By comparing statistics of technical correlation after different SVA and ComBat, we found irw with 10 iterations in SVA package gave the best performance in both reducing batch effects and increasing correlation of technical replicates for the same samples (Figure 2.).
Figure 2. 13: Comparison of Different Batch Effect Removal Methods
Correlation of technical replicates for 168 samples after different batch effect removal methods are summarized.
CHAPTER 3

A MODEL FOR THE EARLY DETECTION AND TREATMENT OF CANCER

3.1 Abstract

Cancer kills ~8M people/year and the WHO predicts an epidemic of cancer in the developing world over the next 20 years due to increased longevity. Development of a simple method to detect and treat cancer early could be a solution to this challenge. Here we test this approach in principle in a mouse mammary tumor model. We had previously demonstrated that tumors could be detected at an early stage in the mouse FVB-NeuN (MMTVneu/202Mul/J) mammary cancer model by immunosignature diagnostics. Immunosignatures are profiles of antibodies in the blood displayed on peptide arrays. We determined if early treatment at the time of diagnosis with the checkpoint blockade, anti-PD-L1, would inhibit the tumor growth. A course of 3 injections at the time of diagnosis significantly retarded tumor growth relative to untreated controls. The early treatment elicited long protection which could restrict the tumor growth even four months after the treatment. The treated mice share a specific immunosignature 28 days post treatment and 28 days post the first palpable tumor, compared to the untreated mice. This indicates that the early treatment elicited an anti-tumor immune response that could inhibit tumor growth. This preliminary experiment suggests that early diagnosis combined with systemic treatment might be a method to control or eliminate cancer.
3.2 Introduction

Since most of the deaths from cancer occur in the developing world, any solution to the cancer problem must be applicable to the whole world population. A possible solution is to develop a simple, inexpensive technique to detect cancer early and couple it to a systemic treatment that would eradicate the tumor. Early detection and early treatment has the advantages that the tumor cell number is small, the chance for evolving evasive mutations is less and the probability of metastasis is greatly decreased. It also is likely a lower dose and shorter treatment would be efficient, so the cost and side effect will be reduced.

We developed the immunosignature diagnostic technology for the early detection of changes in health status, including early cancers [168, 175]. The technology is based on the peptide array with 10,000 spotted peptides. These peptides are disease agnostic as they are chosen from non-biological peptide sequence space [225]. The assay consists of applying diluted sera to the array and then detecting the binding pattern of primary antibody with a labeled secondary antibody. The technology relies on the activation of humoral immune response to diseased cells, such as nascent transformed cells. The humoral immune response changes antibody profile and can be detected by the immunosignature. It has been used to detect multiple diseases. We have shown that this technique can detect Alzheimer’s disease at an early stage in a mouse model [168] and recently showed that mammary tumors could be detected by this method at a very early stage in FVB-NeuN mice, a mouse mammary tumor model.
Immune checkpoint blockades are showing promise as systemic cancer therapeutics. A portion, but not all, patients treated have demonstrated definitive responses [195]. The underlying concept is that an antitumor immune response is naturally mounted against any tumor but it is suppressed in multiple ways by the tumors cells directly (PD1, PD-L1) or through T-regulatory cells (CTLA4) [226]. Checkpoint blockades release this suppression allowing the native antitumor response to limit or even kill the tumor cells. For example, anti-PD-L1 is an antibody therapeutic that has been reported to have efficacy in treatment of late stage melanoma and lung cancer. To date, three such drugs have been approved by FDA. Many more human trials are in the progress. Patients in all clinical trials are at late stage disease [198, 199].

We proposed that combining early detection and a checkpoint inhibitor treatment (eg. anti-PD-L1) would be an effective way to treat tumors, before they were imageable or symptomatic. The question is whether the immune response to a tumor at an early stage, even if unsuppressed, would be sufficient to inhibit the tumor. We tested this basic concept using a transgenic mouse model of breast cancer. In the FVB-NeuN model the MMTV promoter activates the wild type rat ERBB2 gene in mammary cells, and every mammary cell potentially turns to be cancer cell. The mice develop 1-3 (2.7 ± 0.33) tumors at 19-45 (32.6 ± 1.2) weeks of age in our hand. Studies from Boggio and Abe using same mouse model have shown a transition from normal mammary glands to hyperplasia around 8.6 to 15 weeks [23, 215]. We had previously demonstrated that an immunosignature of the initiation of the tumor could be detected at least 12 weeks before a palpable tumor.
We defined this signature as the early stage of the cancer, namely stage I signature. Here we ask whether checkpoint treatment by PD-L1 blockade can be effective at this early stage.

3.3 Results:

3.3.1 Early Detection of Breast Cancer and PD-L1 Treatment

The basic outline of the experimental design is cartooned in Figure 1. Mice were monitored every four weeks for their immunosignature. The treatment was immediately initiated when a mouse was scored positive for the stage I cancer signature in two contiguous monitors. The summary heatmap of the confirmed stage I signature comparing to the historical data is given in Figure 3.1. The treatment consisted of 3 intraperitoneal injections of 200ug anti-PD-L1 (10F.9G2) every three days. Serum was collected in four week intervals after treatment. The timeframe for occurrence of signatures, palpable tumor and treatments is given in Table 3.1. A group of 3 mice were not treated after the signature was identified. The anti-PD-L1 antibody injection step was done by Luhui Shen.
Mice were monitored every four weeks for their immunosignature. The treatment was immediately initiated when a mouse was scored positive for the stage I cancer signature in two continue monitors. Treatment consisted of 3 injections of anti-PD-L1 (10F.9G2) over 9 days. Serum was collected in four week intervals after treatment.

**Figure 3.1: Cartoon of Experimental Design Outline**

**Table 3.1: The Timeframe for Occurrence of Signatures, Palpable Tumor and Treatments**

Sample from anti-PD-L1 treatment group and non-treated control are labeled in the rows. Columns are ages of mice at seven time points. “+” means samples at that time point scores positive for the stage I cancer signature. Green shows the
time point of giving the treatment. Red “Tumor” shows the time point of having first palpable tumor.

3.3.2 Inhibition of Tumor Growth Rate by Early Treatment

The tumor volumes were measured every week and plotted in Figure 3.2. As evident the PDL1 treatment had several effects. Four of ten treated mice did not developed a palpable tumor at 40 weeks old. Tumor growth rate was significantly retarded in the treated mice that eventually developed tumors. This tumor growth inhibition was a long term protection. It was still effective to hinder the growth of tumors in the mice that started to develop a tumor 4 months after of the treatment. Tumors from the untreated controls and the treated mice were examined at the same time point after detection.
3.3.3 RNA Expression of PD-L1 in Mammary Glands

PD- An implication of the effect of anti-PD-L1 treatment is that the gene is expressed in the early stage of tumor development. To test this we performed an RNA analysis on mammary glands with normal morphology and established tumors with different size. (Figure 3.3). The PD-L1 expression level was significantly higher in normal mammary glands than established tumors, especially in mammary gland with increasing expression of MMTV, which indicates the initiation of tumor development by increasing the expression of rat
ERBB2. The PD-L1 expression was decreased after the tumor was established. All of the tumors had similar expression level of PD-L1 regardless the size and expression level of rat ERBB2 (Figure 4). We concluded the L1 gene and presumably protein were expressed on mouse mammary cells and the expression was lower in established tumor.

The RT-PCR step was done by Luhui Shen.

Figure 3. 3: RT-PCR Analysis of MMTV Promotor and PDL1 Expression in Established Tumors and Adjacent Normal Mammary Glands in No-Treated FVB-Neun Mice
RNA analysis on different stages of established tumors and mammary glands. MMTV is the promotor of Rat Her2 gene in FVB-NeuN mouse. Both PDL1 and rat Her2 were at similar expression levels in different stages of the established tumors. PDL1 is express high in normal mg and decrease in tumor.
3.3.4 Changes of Immunosignatures After PD-L1 Treatment

The long term protection of the short term early treatment suggested that the treatment elicited an efficient anti-tumor immune response. The immune checkpoint blockades can induce unique immunologic changes in cancer patients, which may be related to the anti-tumor immune response [227]. It is possible these changes are reflected in the antibody profiles. We found a specific immunosignature of 321 peptides that can distinguish the treated mice 28 days after the treatments with both the age matched untreated mice and their own signature before the treatment. The 321 peptides gave a separated clustering between the treated mice 28 days after the treatments PD-L1 and all samples without PD-L1 treatment at three time points (Figure 3.4a). The same immunosignature was also detected in the treated mice 28 days after treatment and 35 days after their first palpable tumors, compared with non-treated mice at age of 16 weeks (Figure 3.4b). In 46.4% selected peptides, their antibody activities were significantly increased after the tumor established. Taken together, this immunosignature suggests that the anti-tumor immune response was elicited in the treated mice for long term protection.
Figure 3.4: Heatmap Showing PD-L1 Generated Immunosignature

Sera were collected from 8 FVB neuN mice in the PD-L1 treatment group at 28 days after the treatment (AT_T28), and at 35 days after 1st tumor (AT_F35). 8 age-match transgenic controls without any treatment were also collected for sera at corresponding time points (NT_T28, NT_F35). Sera were collected from all mice at age of 16 weeks (NT_Wk16) before any PD-L1 treatment. The immunosignature of 321 peptides selected distinguished samples of AT_T28 and NT_T28.

A) Heatmap of selected peptides depicting simultaneous analysis of samples at 28 days after the treatment (AT_T28) compared with samples without PD-L1 treatment at 3 time points (NT_Wk16, NT_T28, NT_F35).

B) Heatmap of selected peptides showing the immunosignatures at one time point without treatment (NT_Wk16) and two following time points with treatment (AT_T28, AT_F35)
3.4 Methods

3.4.1 Mice and Sample Collection

Female FVB/NJ mice (n=10, age 6 weeks) were purchased from Jackson Laboratories (Barharbor, ME). Female FVB/N NeuN transgenic mice were kindly gifted from Dr. Chella David at Mayo, Rochester. Transgenic FVB/N NeuN mice and wild-type FVB/NJ mice were maintained and weaned in the same barrier isolation housing with standard rodent feed and water. Mice were palpated every week to detect mammary tumor growth. Blood samples were collected by submandibular venipuncture using a 5.0 mm lancet (MEDIpoint, Inc. NY). Serum was separated by serum separating Microtainer (BD Biosciences, San Jose, CA) and stored at -20 C with aliquots. Mice were euthanized by 2,2,2-tribromoethanol at a dose of 125mg/kg through intraperitoneal injection when tumor size exceeded 10% of the body weight. All murine experiments were conducted following animal protocol reviewed and approved by Arizona State University Institutional Animal Care and Use Committee.

3.4.2 Binding of Sera to the Peptide Microarrays

CIM10K Version 3 peptide arrays were printed as earlier described [110, 160]. CIM10K Version 3 library contained 10,000 random 20-residue peptide of 17 random sequence amino acids. Peptides were designed with random sequences with nineteen amino acids (cysteine excluded), except for three residues at N-terminal as a linker to attach peptides covalently on Nexterion A+ aminosilane slides (Schott). Version 3 contained both L and D amino acids with Cys-Ser-Gly-NH2 linker. Peptides were synthesized by Sigma Genosys (St. Louis, MO). Slides
were printed using the two-up format by piezoelectric non-contact printer at Applied Microarrays (Tempe, AZ). Quality-control tests were performed using subsamples of each batch with pooled human naïve serum to select high array-to-array correlation and dynamic range. After being produced, arrays were stored in containers, protected from light.

Slides were first washed by 33% isopropanol, 7.5% acetonitrile, and 0.5% trifluoroacetic acid for five minutes, then dried by centrifuge at 800 rpm for five minutes. After pre-wash, slide processing was automated by a TECAN HS4800-Pro automated incubator, following a protocol optimized for antibody binding. Briefly, arrays were blocking by 0.014% mercaptohexanol, 3% BSA, and 0.05% Tween 20 in 1×PBS solution for 1 h at 23° C. Serum samples were then diluted at 1:500 in incubation buffer (3% BSA 0.05% Tween 20 in 1×PBS) and probed on arrays for one hour at 37° C. Biotinylated anti-mouse IgG(H+L) antibodies (Bethyl Laboratories, Montgomery, TX) were diluted in the incubation buffer at 5nM for one hour at 37° C. Bound secondary antibodies were visualized by Alexa Fluor-647 labeled streptavidin (Invitrogen, CA) at 5nM for one hour at 37° C. The injection volume of each step was 110 μl per array, and TBST solution was used to wash the array between steps. The final wash used TBST and distilled water to remove residual salt on slides. Correlation for technical replicates less than 0.85 were re-probed [175].

3.4.3 PD-L1 Treatment Regime

Stage I cancer signature of FVB/N-NeuN model was established in Chapter 2. The age when the mice showed the stage I signature was 16 weeks. 10
mice were start immunosignature analysis every four weeks since 16 weeks old. The mouse was confirmed with stage I cancer immunosignature in two continual analysis was treated 200ug anti-mouse PD-L1 antibody in PBS every three days through i.p. for total three treatments. Monoclonal anti mouse PD-L1 antibody (10F.9G2) were purchased from BioXcell (Cat#: BE0101).

### 3.4.4 RNA Expression in Mammary Gland

All tissues and tumor were collected and storage in RNA-later (Ambion, Cat#AM7020). Total RNA were purified with TRIzol (Life Tech. Cat#15596-018) and PureLink RNA purification system (Life Tech. Cat#12183-018). Genomic DNA were removed with PapidOut DNA Removal kit (Thermo Sci. Cat#K2981). cDNAs were synthesized from 2ug of total RNA from each sample with cDNA synthesis kit (Thermo Sci. Cat#K1641). PCR reactions were set up with PCR SuperMix (Life Tech. Cat#10572-014) in 20ul reaction volume. The expression of β-actin, PD-L1 and MMTV promotor (the promotor for Rat Her2 gene) were analyzed.

#### PCR primers:

**β-actin:** 5’:CCTGTATGCCTCTGGTCG; 3’: TGTACGCACGATTCC.

**PDL1:** 5’:TTCACAGCCTGCTGACTTT; 3’:GGGATTGACTTTCAGCGTG.

**MMTV:** 5’: ATCGGATGTCGGCGAT; 3’:GTAACACAGGCGATGAGGA.

**PCR condition:** β-actin: 95°C 2’, (95°C 30”, 60°C 30”, 72°C 30”) X 25 cycles, 72°C 5’. PDL1: 95°C 2’, (95°C 30”, 65°C 30”, 72°C 30”) X 27 cycles, 72°C 5’. CMV: 95°C 2’, (95°C 30”, 60°C 30”, 72°C 30”) X 30 cycles, 72°C 5’.
3.4.5 Statistical Analysis

After the final wash, slides were scanned by an Agilent C scanner at 633nm emission (Agilent Technologies, CA) with 100% laser power and 100% PMT. Images were extracted by GenePix Pro 6.0 (Molecular Devices, Sunnyvale, CA) and saved as gpr files. Poor-quality peptide spots were excluded by flagging them as “absent” by visual inspection. Samples with Pearson correlation coefficient >0.85 among technical replicates were used. Raw data from gpr files were collected and imported into R for later processing (Agilent Technologies, Palo Alto, CA). Data were normalized to the 50th percentile to minimize array-to-array variation. Batch effects were adjusted by SVA algorithm with irw as adjustment method with 10 iterations. The biological group was used as a covariate in SVA. Statistical analyses were done using R version 3.1.1, and JMP version 11. P-values were corrected for multiple testing by the Benjamin-Hochberg false discovery rate method. The Welch t test was used in R to select significant peptides with greater than 1.3-fold changes and p<0.05 between two groups. We used a support vector machine with settings including polynomial dot order at 1 and scaling order at 0 for cross-validation and prediction. Packages used in R were sva, caret, e1071 [216, 217].

3.5 Discussion

We had shown that transgenic mice display an immunosignature of having a tumor at least 12 weeks before the tumor is palpable. Short treatment at this early stage with i.p. injection of anti-PD-L1 retarded the tumor growth compared to the untreated mice. This early treatment has long term protection, which can
slow the tumor growth four months after the treatment with similar efficiency in the mice that develop tumor late. High PD-L1 gene expression was evident in the transforming initiated mammary glands, consistent with the effects of anti-PD-L1 treatment at early stages. The early treatment elicited a specific antibody profile which can be detected one month after the treatment and the first palpable tumor established in the treated mice.

It was not obvious that early treatment of a tumor would be effective. It would depend first on the tumor actually producing PD-L1 at this early stage and second, that there is a potential anti-tumor immune response at the same time that was being inhibited by the PD1/PD-L1 immune checkpoint pathway. In other studies, higher mRNA expression of PD-L1 in the diagnosed tumor were observed in the patient’s tissue and also in multiple breast cancer cell lines [204, 228-232]. Triple-negative breast cancer (TNBC), basal type breast cancer cells expressed higher levels of PD-L1 constitutively [228, 232]. Although subtypes of breast cancer can influence clinical treatment strategy, timing is also an important factor when giving treatments. However, few study investigate PD-L1 mRNA expression level at prediagnosed stage and its relationship between anti-PD-L1 treatment effects. Here we demonstrated that PD-L1 was highly expressed in the early stage of the tumor when transformation was just initiated.

The immune checkpoint blockade treatment in cancer patients could induce unique systemic immunologic changes [227], which are possible to elicit a long term anti-tumor immune response against tumor developing in different location. In the late stage cancer, increasing number of tumor-infiltrating
lymphocytes (TIL) in tumors were reported to be significantly correlated with anti-PD-L1 treatment [198, 199, 229]. Therefore, we hypothesize that for protection, one underline mechanism could be anti-PD-L1 treatment can also boost the immune system even at early stage by increasing tumor infiltration.

The long term protection of the early treatment in this experiment indicates the anti-tumor immune responses were elicited and boosted by the treatment. This is supported by a specific and consistent immunosignature of the treated mice in both one month after the treatment and the first palpable tumor was established, comparing to the non-treated mice with matched age and tumor status. Also in multiple advanced human cancers, BMS-936559, an anti-PD-L1 antibody drug, showed long term protection against cancer with a response rate of 6-17% [198]. By using breast cancer [233], melanoma, lung and colon mouse models [234] injected with cancer cells, it was shown that anti-PD-L1 treatment can provide a protection against tumors after the first injection and rechallenges of cancer cells.

Taken together, this suggests that there are cancer antigens at an early stage, which can induce the specific anti-tumor immune response and anti-PDL1 treatment is more beneficial at the early stage of tumor development.

In humans, generally a single primary tumor develops that can metastasize. This contrasts to the mouse mammary tumor model in this study, where multiple tumors can arise as all the cells in the mammary gland can overexpress the oncogenic rat Her2. Also in contrast, therapeutic treatment in humans with checkpoint inhibitors to date, has involved treatment up to 111 weeks [235]. We
choose to only treat 3 times after early diagnosis to determine if even a brief intervention was beneficial. As shown in Figures 3 the early treatment had a significant inhibition on tumor growth late. We suspect that in the case of natural cases of tumors, the early treatment would be more efficacious. It still may require a longer treatment regime than we tested to get better tumor inhibition. We note that the treatment was not started until the diagnosis was confirmed 4 weeks later. While a second monitoring is likely for human early detection and treatment scenarios, in the mouse model this would allow for significant tumor development. It should improve the protection if we treat the mice earlier.

One thing worth noticed is that the average time of first palpable tumor in anti-PD-L1 treatment group was at 30.8 weeks, 35.8 weeks for three mice in non-treatment group, and 33.8 weeks for all non-treatment mice we recorded. There was no significant difference between those three groups. This difference of average time of first palatable tumor by anti-PD-L1 treatment may be due to high variance in small sample size or may be caused by increasing TIL in tumor tissue resulting in increasing tumor size. Topalian et al. shown that in cancer patients after anti-PD-1 treatment the observed tumor size first increased after 2 months treatment then decreased at 4 months [198, 199].

A combination of immune checkpoint blockades therapeutics has been tested in different mouse tumor models and shown better responses than the mono-blockade treatment [236, 237]. The clinical trial of combined nivolumab and ipilimumab treatment on human also showed better response than the mono-treatment with a manageable safety profile [238]. The combined treatment
induced distinct immunologic changes in the patients [227]. Combination therapy may also improve the outcome for the cancer early treatment. Future studies will explore the effect of earlier, longer and combined treatment regimes.

We propose that the work described here could be applied as a general plan for effectively eradicating cancer. People would regularly send in a blood sample to be screened for the early signature of cancer. If a positive was detected the person would be rescreened soon after to confirm the diagnosis and assess the vector of the signature. A second positive diagnosis may be sufficient to initiate treatment. This decision could depend on factors discussed below. The response to the treatment could also be monitored by immunosignature. We believe the principle of all the elements of this scenario have been demonstrated in this work.

One major concern relative to this scenario is the false positive rate. If a tumor was misidentified, or more likely, was a tumor that would self-resolve or remain indolent, the concern is that false positive treatments would lead to unnecessary costs and increasing risk of side effects. We believe that if the immunosignature protocol was both inexpensive and simple, it could substantially relieve the cost of false positives through retesting and assessment of the vector of change in tumor signatures. Studies to date indicate the immunosignature diagnostic can be both simple and inexpensive [225].

The other consideration is the cost and side-effects of the treatments. If the systemic treatment with immune checkpoint inhibitors, or other systemic treatments, had little side effects, then it would not be a risk to treat positives, even though the eradicated tumor was benign or otherwise not life-threatening.
One, as yet to be tested possibility, is that treatment at an early stage would decrease the time of treatment and amount of therapeutic required, therefore reducing the risk and cost of treatment. Checkpoint inhibitor trials to date have been in cases of advanced tumors and have had significant occurrences of adverse effects [226].

Cost of treatment is a significant concern for the model depicted in Figure 6. Currently, the cost of a course of treatment with anti-PD-1 is reported to be over $12,500 a month [239]. If such a treatment were to be administered to kill any tumor arising the cost issue could be of major concern. There are ~1.7M new cancer cases/year in the US. At current checkpoint therapeutic treatment costs, it would require ~$200B/yr to treat all cancers. This is still less than the estimated current cost of treating cancer. Worldwide there are ~14M new cases/yr. ~70% of the new cases are in the developing world. Treatment of these cases at current costs would incur ~$1T/year which is approximately the current cost per year for cancer effects estimated in this sector (ACS/Livestrong report). Even though current costs of treatment might be justified, there is a clear incentive to develop less costly equivalent therapeutics. We also point out that any drug, not only checkpoint inhibitors, that could be administered systemically and be effective could be used in this scenario.
REFERENCES


239. Carroll, J., Bristol-Myers' pioneering PD-1 drug Opdivo OK'd by FDA for melanoma. 2014.


APPENDIX A

DETECTION OF OSTEOSARCOMA RECURRENCE IN CLINICAL CANINE SAMPLES BY IMMUNOSIGNATURE
AP1.1 Introduction

AP1.1.1 Background of Osteosarcoma

Osteosarcoma (OSA) is a malignant tumor of the bone [240]. It often starts in the shoulder, wrist, knee, cranium, spinal column, or ribs. Appendicular osteosarcoma, which is OSA of the limbs, accounts for 75-85% of all bone cancer in dogs. Bones with OSA are destroyed from the inside out. Intensive pain, obvious swelling-like lumps and lameness are evidence of tumor growth in OSA. Tumorous bones are easy to break and difficult to heal. OSA is also highly aggressive. Over 90% of clinical cases have micro-metastasized to the lungs and other organs by the time of diagnosis[241]. OSA is the most common canine bone tumor[242] and accounts for 85% of all primary canine bone tumors[243]. Bone tumors account for approximately 5% of all canine cancer. OSA has a high prevalence in middle-aged to elderly dogs (8 to 10 years) [244] with around 8,000 cases diagnosed each year [240]. Large and giant breed dogs, including Rottweilers, Great Danes, Greyhounds, Saint Bernards and Doberman Pinschers [245-247], tend to develop OSA at younger ages [242]. Male dogs have a higher rate of OSA than female dogs [243].
AP1.1.2 Current Diagnosis and Treatment of OSA:

Currently, radiographic imaging, age, gender and breed are the main diagnostic tools. Tissue biopsy is the gold standard to determine OSA. After the diagnosis of OSA, dogs can be treated with radiation therapy and analgesic medication (carprofen, aspirin, butorphanol and others). Surgery can also be performed. Limb-sparing techniques are adapted from human medicine and involve replacing the tumorous part of the bone with a bone graft.

AP1.1.3 Recurrence of OSA

Because of the high metastasis rate, many the dogs with OSA will already have micro- metastasis or recurrence after treatment of the primary tumor. Metastasis is the key factor that influence patient’s survival rate both in dog and human. In human cancer treatment, patients with localized cancer have better survival rates than those with metastatic disease [55]. According to American Cancer Society projected data for 2015, 61% of breast cancer cases are detected at a localized stage with a five-year relative survival rate of 99% [1]. However, when tumors spread to nearby lymph nodes or other tissues, the survival rate decreases to 85% [1]. If the breast cancer is in a late stage, in which tumors are found in lymph nodes around the collarbone or more distant, the survival rate falls to 25% [1]. The early detection of primary osteosarcoma and its recurrence will improve prognostic capability and treatment modality, relieve pain in dogs, and potentially extend their lives.
AP1.2 Result and Discussion

AP1.2.1 Canine Osteosarcoma Clinical Sample Summary

Dog samples were generously provided by Dr. Douglas H. Thamm from Colorado State University. Five dogs were serially collected for four time points. The four time points were labeled as Initial, A5, PTOR and OR, respectively. Sample IDs and dates for each time point are described below (Table AP1.1). The disease course interval is the difference of days between two time points. Initial was the time point of diagnosis before surgery. A5 was the time point of the fifth chemotherapy treatment, approximately ten weeks after diagnosis. PTOR was the time point of the last visit prior to relapse. OR was the time point of the visit at which the relapse was detected. Age-matched normal serum samples were selected from clinically normal client-owned dogs.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Sample ID</th>
<th>Sample ID</th>
<th>Sample ID</th>
<th>Sample ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTOR</td>
<td>1998/12/11</td>
<td>1999/9/20</td>
<td>2000/2/15</td>
<td>1999/3/19</td>
</tr>
<tr>
<td>OR</td>
<td>1999/3/12</td>
<td>1999/9/27</td>
<td>2000/5/18</td>
<td>1999/6/10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disease course interval time (days)</th>
<th>94</th>
<th>84</th>
<th>76</th>
<th>69</th>
<th>336</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial to A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5 to PTOR</td>
<td>84</td>
<td>32</td>
<td>176</td>
<td>72</td>
<td>168</td>
</tr>
<tr>
<td>PTOR to OR</td>
<td>91</td>
<td>7</td>
<td>93</td>
<td>83</td>
<td>84</td>
</tr>
</tbody>
</table>

Table AP1.1: Canine Osteosarcoma Clinical Sample Summary
Blood samples from five dogs were serially collected for four time points. The four time points are labeled as Initial, A5, PTOR and OR, respectively. Sample IDs and dates for each time point have been described. The disease course interval is the difference of days between two time points.

AP1.2.2 Correlation Analysis Of Overall Antibody Profile During OSA Progression

For this time series study, it was interesting to know how the overall antibody profile changes at different time points by immunosignature technology. At each time point, the intensity of each peptide for the five samples was averaged. Pearson correlation, using all peptides between two time points, was calculated (Table AP1.2).

The highest correlation was 0.948 between OR (the time point of the visit at which the relapse was detected) and PTOR (the time point of the last visit prior to relapse). This means the antibody response at PTOR was similar to that at OR and implies that the samples at PTOR might already have been at the recurrence stage and have a signature of recurrence. The lowest correlation was 0.739, between PTOR and Normal. This implies that compared with the primary tumor, the recurrence was more distinguishable from normal.
Table AP1. 2: Correlation Analysis of Overall Antibody Profile

The experiment was done by using the CIM 10K version-2 peptide arrays from Batch 324. Each sample was run with two to three replicates. Informative peptides were selected by comparing different stages with p value and fold change cutoff. P values were calculated by GeneSpring software using a two-tail T-test with FDR as a multiple comparison correction. The fold change of a peptide was calculated by using the average intensity of samples from one condition divided by the average intensity of samples from the other condition.

<table>
<thead>
<tr>
<th></th>
<th>All sample</th>
<th>Initial</th>
<th>A5</th>
<th>OR</th>
<th>PTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td></td>
<td></td>
<td>0.943</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>0.95</td>
<td></td>
<td>0.942</td>
<td>0.948</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>0.889</td>
<td>0.946</td>
<td></td>
<td>0.739</td>
<td></td>
</tr>
</tbody>
</table>
| PTOR      | Normal     | 0.912   | 0.83 | 0.854 |}

Table AP1. 3: Distribution of Fold Change in Selected Peptides

Fold is calculated by using the average intensity of each peptide at one time point divided by the other. Samples of recurrent disease include samples at both OR and PTOR.

<table>
<thead>
<tr>
<th></th>
<th>from normal to initial</th>
<th>from normal to recurrence</th>
<th>from normal to OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fold</td>
<td># increase</td>
<td>decrease</td>
<td>Fold</td>
</tr>
<tr>
<td>1.5</td>
<td>115</td>
<td>99</td>
<td>67</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>2.5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

#: Numbers of peptides have different fold change between two conditions. An increase means that peptides have a certain fold change from one condition to the other, and vice versa for a decrease.

Table AP1.3 lists the distribution of the numbers of selected peptides by fold change in three comparisons. I compared dog samples without canine cancers
(normal) with samples at initial diagnosis to find the immunosignature for primary OSA. As mentioned above in the introduction, OSA is highly aggressive, and micro-metastases may already be present when primary OSA is detected. The time difference between PTOR and OR was 71.6 days on average (ranging from 7-93 days) for the five dogs. At the time of PTOR, the dogs could already have had micro-recurrent tumors. By combining samples at PTOR and OR together, with label “recurrent,” this could cover all stages of recurrent OSA comprehensively. By comparing normal with recurrent samples, I could investigate the immunosignature of recurrent OSA. OR was the time point when a relapse was determined clinically. By comparing OR with normal samples, I could investigate the immunosignature of recurrent OSA for that specific time point.

The result shows that most selected peptides had 1.5- to two-fold change. A few had a greater than four-fold change. Most selected peptides were increased, with fewer than 20% decreased, on average. The number of decreased peptides dropped faster than the increased ones with the increase of fold change. These data show that we could find immunosignatures to distinguish primary OSA and recurrent OSA from healthy controls. They also show that most of the selected peptides had higher intensity levels in dogs with OSA compared with health controls.

AP1.2.3 Comparison of Immunosignature at Multiple Time Points

After finding informative peptides for each time point, I wanted to ask how the intensities for those peptides changed along tumor development. This
could reveal the role of tumor-related antibodies during the disease progression of OSA.

I first selected two peptide sets: 1) the 94 overlapped peptides between 115 peptides of primary OSA and 278 for recurrent OSA at both OR and PTOR; and 2) the 86 overlapped peptides between 115 peptides of primary OSA and 197 for recurrent OSA at OR. 85 peptides of the 86-peptides set were shared in the 94-peptide set.

Figure AP1. 1: Proportion of Overlapped Peptides Shared Among Different Time Points
Intensity level of the overlapped peptides are plotted against each time point.
Table AP1.4: Proportion of Overlapped Peptides Shared Among Different Time Points

From the Figure AP1.1, the intensity of the majority of the overlapped peptides intensity increased from Initial to A5 and to PTOR, but decreased in OR. From the Table AP1.4, the number of significant peptides overlapped between primary OSA and recurrent OSA was 74.8%–81.7% in the 115-peptide set and decreased to 43.7%–47.7%.

These data mean that different antibodies were dominant at different time points. This might be due to the immunoediting process between OSA and the immune system in dogs. Antigens that presented in the primary OSA disappeared due to high immunogenicity. After the tumor developed, more neo-antigen may have appeared and triggered different antibody responses at the recurrent stage.

Figure AP1.2: Intensity Plot of Significant Peptides between Primary and Recurrent OSA
In order to further test whether there was a difference between primary tumor and recurrent tumor, I asked how many peptides were statistically significant between primary and recurrent by comparing these two time points. 35 peptides were selected. As Figure AP1.2 shows, most of them decrease from primary to recurrent. The PCA result on the 35 peptides shows cluster of primary samples and recurrent samples by each time point. These show that primary OSA differed from recurrent OSA in terms of their immune response to tumor antigens.

**Conclusion and Future Direction**

I recognize the limited sample size for this experiment, and I don’t have further validation of the discovered immunosignature. Some of the selected peptides may have been due to randomness. The signature discovered may not well represent different breeds of dogs with osteosarcoma. However, in this study, I have used multiple comparison methods to reduce the possibility of randomness and carefully test each sample. This study could still be beneficial to studying osteosarcoma.

With the development of OSA, the overall antibody profiles changes substantially. Different immunosignatures were found for primary OSA and recurrent OSA. The significant peptides are increased from normal controls to tumor samples, showing increasing anti-tumor immune response. The difference between primary OSA and recurrent OSA exists. Most significant peptides decrease from primary OSA to its recurrence. Based on the data, the selected peptides can be grouped into the following categories: Intensities were high in
normal and low in initial. This may be due to the immunosuppression or Treg by
the primary tumor. Intensities increased after treatment: suppression was relieved
and IS increased in the recurrence. Intensities decreased at late recurrence: the
recurrent tumor seemed to suppress these peptides but not significantly. It
behaved differently from the primary tumor. This may imply that the immune
suppression from primary and recurrent tumors was different.

In order to get a more solid immunosignature of primary and recurrent
osteosarcoma, more dog samples with initial, A5, PTOR, OR time points are
needed. Current PTOR is the time point before relapse is detected, with an
average of 71.6 days before recurrence. In order to address the early detection of
OSA recurrence, it will be better to get a time point that precedes recurrence
detection by a longer period. It will be better to have samples for which the
interval between “initial to A5” and “PTOR to OR” is long enough to have more
distinguished stages.

AP1.3 Methods

AP1.3.1 Sample Summary

Dr. Douglas H. Thamm from Colorado State University generously
provided five canine osteosarcoma clinical samples. The breeds of the five canine
patients are Borzoi, Labrador retriever, golden retriever, and mix breeds. Each
dog had been confirmed with osteosarcoma histologically or cytologically. Sera
from each dog were serially collected for four time points. The four time points
were labeled as Initial, A5, PTOR and OR respectively. Initial is the time point of
diagnosis before surgery. A5 is the time point of firth chemotherapy treatment,
approximately ten weeks after diagnosis. PTOR is the time point of the last visit prior to relapse. OR is the time point of the visit at which relapse was detected. Sera were collected and stored at -80°C through all collections.

Chemotherapy-naïve primary tumor samples were selected from the Colorado State University Animal Cancer Center's tissue archive based on the criteria that the patient had undergone surgical amputation followed by chemotherapy with doxorubicin and/or a platinum-based drug. Doxorubicin (Adriamycin) was given every two weeks for five treatments.

AP1.3.2 Binding of Sera to the Peptide Microarrays and Microarray Data Processing

CIM standard clinical protocol with four steps was used. Detailed methods can be found in Chapter 2 and 3.

AP1.3.3 Statistical Analysis

Statistical analysis of microarray data was done with Gene-Spring 7.3.1 (Agilent, Inc., Palo Alto, CA) by importing image-processed data from GenePix Pro 6.0 (Axon Instruments, Union City, CA). Calculations based on the GenePix-prepared gpr text files were done on the median signal intensity per spot. Poor quality spots were excluded from analysis by flagging as “absent” upon visual inspection. Prior to analysis, each array was normalized to the 50th percentile to eliminate array-to-array variation, and signal intensities of less than 0.01 were set to 0.01. Values from replicate arrays were averaged and used in the analysis. P-values were corrected for multiple testing by the Benjamin-Hochberg false discovery rate method in Genespring GX 7.3 software (Agilent Technologies,
Palo Alto, CA). The Welch t test was used in GeneSpring to select significant peptides with p<0.05 between two groups.
APPENDIX B

CHARACTERIZATION OF ANTIBODY PROFILE DURING CANCER DEVELOPMENT
AP 2.1 Quantitative Characterization Of Antibody Profiles During Tumor Development

AP2.1.1 Introduction

Immunosignatures allow us to study the overall antibody profile of a sample. We can quantitatively evaluate the distribution of overall antibodies at a certain time point in a disease to potentially reveal its internal patterns. A quantitatively method to measure statistical dispersion of antibodies can indirectly show antigen heterogeneity. This can help researchers to understand the interaction between the immune system and cancer.

Different statistical parameters are available to characterize the distribution of information with a single quantitative value. With the increasing popularity of mobile and wearable devices, customers can collect data about their health easily. A simple visualization of major biological information in our bodies will make it possible for everyone monitor their health. In Dr. Kurt Whittemore’s dissertation research, he used AbStat as a tool to characterize the antibody profile of a patient through immunosignature technology. The current work is a collaboration with Dr. Whittemore and uses his AbStat algorithm to calculate entropy, coefficient of variation, kurtosis, skew, and dynamic range [248-251].

Entropy is a measurement of the homogeneity of a frequency distribution. If all information in the distribution is ordered (with homogeneous value), its entropy decreases. The minimum value is zero when all values in that distribution are identical. The value increases when the distribution is more heterogeneous. The maximum entropy is achieved when values of every element are different
from each other. Kurtosis measures the peakedness of a frequency distribution, with higher values indicating a sharper peak. Skewness measures symmetry: a larger absolute amount of skewness indicates a more asymmetric distribution, which means that values in that distribution lean to one side of the mean. Dynamic range is the ratio of the value of 95% intensity to the value of 5% intensity. A high dynamic range indicates a wider range of distribution. The coefficient of variation is the ratio of the standard deviation to the mean. A high coefficient of variation indicates a high degree of dispersion.

We performed a time series study to characterize the overall antibody profile during tumor development, using two transgenic mice and two wild type controls (Figure AP2.1, Figure AP2.2). By doing this, we hope to find some interesting general pattern between the immune system and cancer.

AP2.1.2: Result and Discussion

Sera were collected every two weeks for two transgenic mice and two age-matched wild type mice (Table AP2.1, Table AP2.2). Sera from these samples were run on the CIM Version-2 arrays using a standard CIM clinical protocol with one replicate at each time point. We used one replicate because the abundance of continuous time points allowed us to get a good representation of the antibody profile during tumor development with one replicate. Slides were randomly chosen from all available slides in the same batch of slides by “sample” function in R. Arrays, and samples were also randomly assigned to different positions on the Tecan.
Figure AP2. 1: Summary of Mice Samples
Sixteen time points of Tg8-1-1, twelve time points of Tg8-3-9 and WT1-3, and eight time points for wild type mice WT2-3 were used.

Figure AP2. 2: Tumor growth Curve for Tg8_1-1 and Tg8_3-9 Transgenic Mice

Entropy, kurtosis, skewness, dynamic range, and CV were calculated for two transgenic mice and two wild type controls at multiple time points during tumor development (Figure AP2.3, AP2.4). The results show that the value of kurtosis, skewness, dynamic range, and CV were similar and correlated between
transgenic mice and control mice. The chaotic pattern of these parameters showed
dynamic changes occurring in the antibody repertoire during mouse development
(Table AP2.1). These changes might be the result of the aging process instead of
tumor development, as there was no significant difference between the transgenic
mice and the wild type controls for these parameters.

Figure AP2. 3: AbStat Time Course for Transgenic Tg8-1-1 and Wild Type WT2-3 Mice
Only entropy showed some level of difference between the transgenic samples that developed tumor later and the wild type controls (Figure AP2.5). In WT 2-3, the entropy for the wild type controls was linear increasing with less variance, compared with Tg8-1-1. The entropy for Tg8-1-1 fluctuated along tumor development with significant variance. However, this pattern was not
evident in the wild type mouse. Dr. Kurt Whittemore’s study showed a positive correlation between entropy and age. The constant increase in wild-type mice may be the result of the aging process, when more antibodies were produced during their lifetimes. However, in the transgenic mice, entropy of the antibody repertoire was chaotic instead of showing a constant increase. This might have been the result of an immunoediting process between tumor and immune system. Different antibodies were produced to target tumors at different development stages. The immune system might have produced a large amount of antibodies when antigens were highly immunogenic in tumor cells and then decreased antibody production because of mutations in those antigens that made them not immunogenic. This distinct difference of entropy in transgenic mice and wild-

differentiation graphs.jpg

Figure AP2. 5: Entropy Time Course for Transgenic and Wild Type Mice

In conclusion, this study provided an interesting approach to evaluating
the antibody repertoire during tumor development. A time series study of cancer was first conducted in the immunosignature platform to discover hidden patterns over time. The results implied that entropy could be a useful parameter to evaluate tumor development and to measure statistical dispersion of antibodies and indirectly show antigen heterogeneity.

AP2.2 A Time Series Study of Antibody Profile Along Tumor Development

AP2.2.1 Introduction

Although immunoediting theory suggests that tumor cells present different antigens along tumor development, few studies have investigated the dynamic process of the immune system during tumor development. Immunosignature can reflect different antibody profiles at different stages. It provides us an opportunity to evaluate the antigen heterogeneity along tumor development by analyzing immunosignature in a time series study.

Compared with static comparisons, a time series study of biomarkers has the advantage of improving the performance of diagnostic methods. The pitfall of current biomarkers in bigger populations is that there is substantial heterogeneity even in populations with similar genetic backgrounds. Although common profiles can be found and an absolute cutoff can be drawn, the results may not be relevant to each individual. Therefore, dynamic measurements that allow the assessment of within-patient expression changes will be more useful [56]. The recent study on continuous monitoring of multiple omics by Rui Chen and colleagues showed the promise of self-monitoring in personalized medicine and diagnosis [252].
The method for analyzing multiple time points of a feature falls under time series analysis. Time series studies have a number of advantages: they can use the time series structure in the data; can increase the accuracy of prediction; and can capture information about features with transient expression changes. These are not possible in static studies that compare only two populations at a single time point.

In this study, an individual dynamic monitoring of the antibodyome was conducted in a murine breast cancer model.

AP2.2.2 Result and Discussion

In order to study the time serial pattern of the antibodyome, we used serum samples from the same individual at multiple time points. In fields in which time series analyses is widely used, such as economics, astronomy, physics and industrial engineering, time series data normally include more than 30 time points. However, in biology, such an extended sequence of samples is rarely accessible. For this study, I had a time series serum collection for FVB/N neuN transgenic mice and FVB/N wild-type mice with more than ten time points for each mouse. Serum for each mouse was collected every two weeks, from mice aged 5-10 weeks to 40-45 weeks, which covered the complete tumor development stages.

In this study, I chose two transgenic mice (Tg8-1-1, Tg8-3-9) and two wild-type mice (WT1-3, WT2-3) (Table AP2.2, Table AP2.3). Sixteen time points of Tg8-1-1, twelve time points of Tg8-3-9 and WT1-3, and eight time points for wild type mice WT2-3 were used. The samples used in this study were the same samples run in the AP2.1 section. The interval was the weeks between two time
points. It should be noted that in order to have a better result in time series analysis, a constant interval is important. In Tg8-1-1, the average interval is 2.02±0.37 weeks. In Tg8-3-9, the average interval is 1.98±0.4 weeks.

<table>
<thead>
<tr>
<th>Tumor size (mm³)</th>
<th>Tg8-1-1</th>
<th>WT2-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age (wks)</td>
<td>Window (wks)</td>
</tr>
<tr>
<td>15.1</td>
<td>-18</td>
<td>2.3</td>
</tr>
<tr>
<td>17.4</td>
<td>-15.7</td>
<td>1.9</td>
</tr>
<tr>
<td>19.3</td>
<td>-13.8</td>
<td>2.1</td>
</tr>
<tr>
<td>21.4</td>
<td>-11.7</td>
<td>2.7</td>
</tr>
<tr>
<td>24.1</td>
<td>-9</td>
<td>1.2</td>
</tr>
<tr>
<td>25.3</td>
<td>-7.8</td>
<td>2.1</td>
</tr>
<tr>
<td>27.4</td>
<td>-5.7</td>
<td>2</td>
</tr>
<tr>
<td>29.4</td>
<td>-3.7</td>
<td>2.2</td>
</tr>
<tr>
<td>31.6</td>
<td>-1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>0.5(1)</td>
<td>33.1</td>
<td>0</td>
</tr>
<tr>
<td>89.43(2)</td>
<td>35.6</td>
<td>2.5</td>
</tr>
<tr>
<td>261.71(2)</td>
<td>37.4</td>
<td>4.3</td>
</tr>
<tr>
<td>423.64(2)</td>
<td>39.1</td>
<td>6</td>
</tr>
<tr>
<td>711.49(3)</td>
<td>41.1</td>
<td>8</td>
</tr>
<tr>
<td>1626.01(4)</td>
<td>43.3</td>
<td>10.2</td>
</tr>
<tr>
<td>3326.36(4)</td>
<td>45.4</td>
<td>12.3</td>
</tr>
</tbody>
</table>

Table AP2. 2: Sample Information for Tg8-1-1 and WT2-3
Table AP2. 3: Sample Information for Tg8-3-9 and WT1-3

<table>
<thead>
<tr>
<th>Tumor size (mm3)</th>
<th>Age (wks)</th>
<th>Window (wks)</th>
<th>Interval (wks)</th>
<th>Age (wks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19.3</td>
<td>-12.3</td>
<td>2.1</td>
<td>17.3</td>
</tr>
<tr>
<td>0</td>
<td>21.4</td>
<td>-10.2</td>
<td>2.7</td>
<td>19.1</td>
</tr>
<tr>
<td>0</td>
<td>24.1</td>
<td>-7.5</td>
<td>1.2</td>
<td>21.1</td>
</tr>
<tr>
<td>0</td>
<td>25.3</td>
<td>-6.3</td>
<td>2.1</td>
<td>24.1</td>
</tr>
<tr>
<td>0</td>
<td>27.4</td>
<td>-4.2</td>
<td>2</td>
<td>27.3</td>
</tr>
<tr>
<td>0</td>
<td>29.4</td>
<td>-2.2</td>
<td>2.2</td>
<td>29.1</td>
</tr>
<tr>
<td>35.93(1)</td>
<td>31.6</td>
<td>0</td>
<td>1.5</td>
<td>31.3</td>
</tr>
<tr>
<td>165.38(1)</td>
<td>33.1</td>
<td>1.5</td>
<td>2.5</td>
<td>33.1</td>
</tr>
<tr>
<td>493.32(1)</td>
<td>35.6</td>
<td>4</td>
<td>1.8</td>
<td>35.1</td>
</tr>
<tr>
<td>1105.01(2)</td>
<td>37.4</td>
<td>5.8</td>
<td>1.7</td>
<td>37.4</td>
</tr>
<tr>
<td>2215.85(2)</td>
<td>39.1</td>
<td>7.5</td>
<td>2</td>
<td>39.1</td>
</tr>
<tr>
<td>2932.92(2)</td>
<td>41.1</td>
<td>9.5</td>
<td>4.1</td>
<td>41.3</td>
</tr>
</tbody>
</table>

Autocorrelation function (ACF) is a standard method to analyze time series data. A feature with time series patterns is autocorrelated, and the Durbin-Watson test can calculate and test its significance (Figure AP2.6). The Durbin-Watson Test in the R package was used to calculate the significance of ACF with a cutoff of p value at 0.05. Peptides passing the test should have autocorrelation structure instead of randomness.
249 peptides in Tg8-1-1 and 849 peptides in Tg8-3-9 passed Durbin-Watson test with p value less than 0.05. These peptides showed an autocorrelated structure. Self-Organizing Map was used to cluster the significant peptides into Tg8-1-1 and Tg8-3-9 and to reveal potential patterns (Figure AP2.7). Several clusters showed a correlated trend between signal intensity and tumor development, although patterns were different for each individual. Tg8-1-1 had increased peptides around a tumor-initiated point, while Tg-3-9 only had decreased ones. This pattern could be a representation of the immunoediting process in the mouse immune system.
Figure AP2. 7: Self-Organizing Map to cluster 249 autocorrelated peptides in Tg8-3-6.
Setting for SOM in GeneSpring are Rows 3, Columns 3, Iterations 10000, Radius 4.0.

Tg8-1_ACF_249: Rows 3, Columns 3, Iterations 10000, Radius 4.0.
One cluster of the 249 peptides by K-means included a cluster of 24 peptides (Figure AP2.9). The intensity of these peptides increased 5.7 weeks before the first tumor occurred and decreased later. The intensity changes of these peptides may be caused by the primary tumor. In order to show that the changes are tumor-related, the same 24 peptides were also plotted in the age-matched wild-type mice. The autocorrelated pattern appeared only in the transgenic mice with the development of tumors but not in the wild-type controls, which indicated its tumor specificity. The changing pattern implies that tumor antigens may have appeared early, before the primary tumor triggered a strong immune response, and that those antigens disappeared when the tumor matured.
In this study, an individual dynamic monitoring of antibodyome was conducted in a murine breast cancer model. Autocorrelation function was calculated for each peptide in two transgenic mice and two wild-type mice. The Durbin-Watson test was used to find significant peptides with autocorrelation structure. Selected autocorrelated peptides showed dynamic and tumor-specific changes with tumor development. The pattern of changes can be tracked back five weeks before the tumor was palpable. Although this study used only four mice and the results have not been repeated, the results might attract the interest of other researchers in this field. The method used in this study can also be applied to genomic, transcriptomic, proteomic, and metabolomic data.
AP2.3 Robustness of immunosignature

AP2.3.1 Introduction

A good, robust biomarker should have low variance in a population, even when that population is diverse. On the one hand, since many biomarkers have high variance in both healthy and diseased populations, they are likely to fail in clinical trials, as they cannot differentiate the overlap between the healthy and diseased populations. This kind of biomarkers I called limited biomarkers. They will lead to high false-positive rate. A robust biomarker, on the other hand, should have low variance in both healthy and diseased populations (Figure AP2.10) so that a sample can be classified as either disease-positive or disease-negative with high accuracy. We use correlation to measure the heterogeneity in a population. If the correlation among samples in a population is high, that means that there is low variance in that population.
AP2.3.2 Result and discussion:

In this study, serum samples from four transgenic (Tg) and eight age-matched wild-type (WT) control mice at a time point before the first breast tumor was detected were run in CIM V2 arrays using the standard CIM clinical protocol.
(Table AP2.4). Sixty-seven peptides were significant by t-test (p<0.05), with a 1.3-fold change as the cutoff. The selected peptides were specific immunosignatures for breast cancer at two weeks before the first tumor. Here, we wanted to test whether these peptides were robust or not.

The correlations of 67 peptides in either the transgenic group or the control group were high and had a low variance (Figure AP2.11). These were called in-group correlations. The calculated correlation between one transgenic sample and one wild-type sample and all possible combinations were called between-group correlations. The correlation of 67 peptides was lower than the in-group correlation and had a much wider range. This showed that the selected 67 peptides had lower variance in in-group correlations and had higher variance in between-group correlations (p<0.01 between among and WT).

![Correlation by 67 peptides](image)

Figure AP2. 11: Correlation of selected peptides in transgenic (Tg) and wild-type (WT) control mice. Correlations of 67 peptides in any two samples of Tg group are label with “Tg”, same as “WT”. Correlations between one transgenic sample and one wild-type sample and all possible combinations are labelled with “among”.

148
In order to show that the difference between in-group and between-group correlations were not due to internal array bias, I calculated the correlations for both using all peptides on the 10K array (Figure AP2.12). This should reflect the overall antibody profiles of samples. The result was that in-group and between-group correlations were at around same level, and their variance was not as great as in the correlation of the 67 selected peptides.

Figure AP2. 12: Correlation of all peptides on the 10K array

In sum, I tested the robustness of selected peptides and showed that robustness could be a criterion for evaluating the performance of an immunosignature.
Permissions have been obtained to reprint copyrighted material.

American Cancer Society has granted permission automatically to use Figure 1. 1 Biology of Breast Tissue and Figure 1. 2: Illustration of Mammogram and MRI and Figure 1. 3: Five-year Relative Survival Rates (%) by Stage in US.

Molecular & Cellular Proteomics has granted permission automatically to use Figure 1. 4: Distribution of Human Plasma Protein.

Figure 1. 6: Introduction of CTLA-4 and PD-1 Treatment has obtained permission for Elsevier with license number 3654470549052.

Use of Figure 1. 7: Biology and Mechanism of Anti-PD-L1 Antibody Treatment has been approved by Leann Sandifar, Vice President – Operations, from Patient Resource LCC on June 11th.
BIO SKETCH

I grew up surrounded by outstanding healthcare professionals and biomedical researchers. As a child, I was curious about the fancy equipment that doctors used to perform surgery and amazed by how easily the drugs a doctor prescribed could cure a patient’s disease. Seeing patients relieved of their pain and the happiness of parents when their child was cured made me feel that doing something that could improve people’s health would be a meaningful way to spend my life. I chose biotechnology as my undergraduate major at Wuhan University in China. The concept that, in the future, a person’s health condition could be monitored conveniently and that diseases could be detected before symptoms occur became clear to me. I realized that early diagnosis and prevention were more crucial than treatment at the late stage of disease. Since that time, I have been interested in all kinds of new technologies to diagnose a disease. I came to ASU because it was one of only a few universities that have many professors focusing on developing diagnostic tools for human diseases. My work with Dr. Johnston is to find new strategies to diagnose breast cancer early and to improve cancer outcomes through novel treatment. It has been an amazing journey, and I feel sincerely grateful to Dr. Johnston and ASU for giving me this opportunity. Upon graduation, I plan to pursue a career in building novel technology to improve people’s health.