Frameshift Antigens for Cancer Vaccine Development

by

Jian Zhang

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

Approved February 2018 by the
Graduate Supervisory Committee:
Stephen Albert Johnston, Chair
Phillip Stafford
Yung Chang
Qiang Chen

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Immunotherapy has been revitalized with the advent of immune checkpoint blockade treatments, and neo-antigens are the targets of immune system in cancer patients who respond to the treatments. The cancer vaccine field is focused on using neo-antigens from unique point mutations of genomic sequence in the cancer patient for making personalized cancer vaccines. However, we choose a different path to find frameshift neo-antigens at the mRNA level and develop broadly effective cancer vaccines based on frameshift antigens.

In this dissertation, I have summarized and characterized all the potential frameshift antigens from microsatellite regions in human, dog and mouse. A list of frameshift antigens was validated by PCR in tumor samples and the mutation rate was calculated for one candidate – SEC62. I develop a method to screen the antibody response against frameshift antigens in human and dog cancer patients by using frameshift peptide arrays. Frameshift antigens selected by positive antibody response in cancer patients or by MHC predictions show protection in different mouse tumor models. A dog version of the cancer vaccine based on frameshift antigens was developed and tested in a small safety trial. The results demonstrate that the vaccine is safe and it can induce strong B and T cell immune responses. Further, I built the human exon junction frameshift database which includes all possible frameshift antigens from mis-splicing events in exon junctions, and I develop a method to find potential frameshift antigens from large cancer immunosignature dataset with these databases. In addition, I test the idea of ‘early cancer diagnosis, early treatment’ in a transgenic mouse cancer model. The results show that

early treatment gives significantly better protection than late treatment and the correct time point for treatment is crucial to give the best clinical benefit. A model for early treatment is developed with these results.

Frameshift neo-antigens from microsatellite regions and mis-splicing events are abundant at mRNA level and they are better antigens than neo-antigens from point mutations in the genomic sequences of cancer patients in terms of high immunogenicity, low probability to cause autoimmune diseases and low cost to develop a broadly effective vaccine. This dissertation demonstrates the feasibility of using frameshift antigens for cancer vaccine development.

ACKNOWLEDGMENTS

My thesis work would not be finished with the help of many people here in ASU. I am very grateful to my advisor Dr. Stephen Albert Johnston who helped me grow up as a researcher and guided me in troubleshooting, who provided funding to my PhD career and all the resources in my experiments. Dr. Johnston taught me to think differently anytime I had negative results and try to extract all the useful information from every piece of data we collected.

I am grateful to my committee member Dr. Phillip Stafford, who gave me suggestions about data analysis and taught me a lot about the immunosignature technology. I am also grateful to Dr. Yung Chang and Dr. Qiang Chen for guidance in my research.

The Center for Innovations in Medicine (CIM) in Biodesign Institute is a wonderful place to work, everyone is very kind and willing to help. I want to give my special thanks to Dr. Luhui Shen, who taught me how to do most experiments in my thesis. And I want to thank Dr. Zbigniew Cichacz and the peptide array core team for my immunosignature studies, I want to thank Lu Wang for providing help in my statistical analysis.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

TABLE Page

TABLE Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER 1

Introduction

1.1 Cancer Statistics

Cancer has been a major public health concern around the world for decades and it is also the second leading cause of death in US after heart disease. It is predicted that in 2017 there will be 1,688,780 new cancer cases and 600,920 cancer deaths in the United States. Even with the tremendous improvement of technology in cancer diagnosis and treatment, overall cancer incidence has increased from 1975 to 2013 while mortality rate has only dropped slightly in the same period. The probability of developing invasive cancer in men's lifetime is 40.8%, for women the number is slightly lower: 37.5% [1]. These numbers indicate that cancer is not only affecting a group or a portion of the people, it has a great impact over every one of us.

With billions of dollars invested in the cancer research and trillions of dollars paid by patients in cancer health care, cancer incidence or mortality rate has not been dramatically reduced. The reason behind this disappointing situation is probably due to the complexity of this disease, and it is also possible that cancer research has been led in a wrong direction.

Cancer is a group of abnormal cells with uncontrolled growth and the potential to invade or spread to other parts of the body. Since cancer cells can originate from almost anywhere in the body, cancer is a collection of diseases. There are over 100 human

cancer types and it is believed that each cancer type has its unique origination and path to develop into a tumor. Even for the same cancer type, each patient has been shown to have unique mutations and may need to be treated uniquely (personal medicine). These complexities and difficulties make it so hard and so expensive to treat cancer effectively. However, there are still six major hallmarks of cancer which are shared across multiple kinds of cancer types and summarized by Hanahan and Weinberg: in order to have uncontrolled cancer cell growth, these cancer cells need the ability to sustain proliferative signaling, resist cell death signaling, evade growth suppressors and enable replicative immortality; to invade and spread to other locations of the body, they need to induce angiogenesis and activate metastasis signaling [2].

For a long period, people attributed cancer to inherited genetic mutations or environmental factors like diet, obesity, tobacco or other carcinogens. But recently a group of scientists in John Hopkins University calculated the correlation between lifetime risk of cancers in 17 tissues worldwide and the lifetime number of stem cell divisions in those tissues, they found that these two numbers are well correlated (median CORR=0.8) and they concluded that random DNA replication errors are responsible for two-thirds of human cancer mutations. While environmental factors are still very crucial for a subset of cancers like esophagus cancer, lung cancer, stomach cancer and cervical cancer. Surprisingly, inherited genetic factors have minimal effects in cancer development of the human population [3]. This conclusion makes the cancer research situation even worse since random DNA replication errors are very difficult to predict and target. The

implication is we will need to design unique drugs/vaccines for each patient - known as precision medicine.

Despite that we are facing so many difficulties with cancer disease, a lot of new treatments have emerged as people are learning from the battle against cancer. Within those cancer treatments, immunotherapy has been the most popular and successful in recent years.

1.2 Current Cancer Immunotherapy and Limitations

The immune system is our weapon to defend the body against all "foreign" invaders including foreign microbes and abnormal cells in our body. The idea of cancer immunotherapy is to utilize the immune system to treat cancer - it can be either activating or suppressing the immune system of cancer patients.

The logic behind cancer immunotherapy is based on the "foreign" invader characteristics of cancer. Any protein which has not been exposed to the immune system can induce an immune response against that protein. Cancer cells need to go through several rounds of mutations to become invasive and the immunogenic mutated proteins related to tumor formation are tumor specific antigens, while at the same time, cancer cells are also producing many errors in proteins unrelated to cancer progression during uncontrolled growth. These are passenger neo-antigens. And the third group of tumor antigens is tumor associated antigens which are over-expressed wildtype proteins. These "foreign" antigens can induce specific anti-tumor immune responses and these specific immune responses are the key for immunotherapy.

There are many types of cancer immunotherapy, and they can be classified into four categories based on the mechanisms of therapy: Cellular Immunotherapy; Antibody Therapy and Cytokine Therapy; Immunotherapy targeting innate immune responses; Cancer Vaccines. Within these four categories, most immunotherapy treatments are specific treatments which target one specific type of cancer or even one cancer patient for a personalized cancer vaccine, while other immunotherapy methods are non-specific. These are used to boost the immune system generally and can be used broadly in different types of cancer.

Dendritic cell therapy and Adoptive T-cell transfer therapy are the two most extensively studied cellular therapies. Sipuleucel-T (Provenge) is one kind of dendritic cell therapy, and it was approved by FDA for treating castrate-resistant prostate cancer in 2010. The treatment method includes three steps: extraction of PBMC (primarily dendritic cells) from patient by leukaphoresis; incubation of dendritic cells with a prostate cancer specific antigen PAP (antigen prostatic acid phosphatase) and immune stimulating factor GM-CSF (granulocyte-macrophage colony stimulating factor); activated dendritic cells are then reinfused into the patient. This treatment has been shown to increase patient survival rate modestly on average but significantly for a minority [4]. Another limitation for this cell-based vaccine is the high cost. The company Dendreon filed for bankruptcy in 2014 due to lack of profit.

Chimeric Antigen Receptor (CAR) T cell therapy is an adoptive T-cell transfer therapy using engineered T-cell receptors. CAR usually consists of three parts: an artificial antibody to bind a common cancer antigen on tumor cells; part of a receptor for

activating cells and multiple activation and co-stimulatory endo-domains to overcome immune tolerance and immune suppression in tumor microenvironment. Carl June's group reported using CAR T cells expressing the anti-CD19 domain and a costimulatory domain from CD137 and the T cell receptor ζ domain to treat three advanced chronic lymphocytic leukemia (CLL) patients successfully. Two of three patients had complete remission and on average one CAR T cell eradicated at least a thousand CLL cells [5]. However, CAR T therapy also has severe limitations. The CAR-T design needs an antibody domain to bind common tumor antigens. Usually the common tumor antigen is a tumor associated antigen and it is widely expressed in normal tissues, so unexpected autoimmune disease is a danger in such treatments. For example, in a trial conducted by NCI researchers, nine cancer patients were treated using anti-MAGE-A3 CAR T cells. Two patients died from comas because another MAGE-A family protein with high similarity to MAGE-A3 has low expression in brain tissue as well [6].

Antibody Therapy has been the most successful immunotherapy for last few decades and the FDA has approved over a hundred monoclonal antibodies for cancer therapeutic treatments [7]. The principle for antibody therapy is relatively simple. Antibodies can be designed to target every hallmark of cancer like angiogenesis/tumor specific antigens/tumor associated antigens/immune suppression etc. Some of these antibodies are targeting tumor antigens for a specific type of cancer and thus can only be used for one cancer type or subtype. We can call them specific treatments. While other antibodies are targeting tumor angiogenesis or immune suppression and can be used widely in multiple cancer types. These are non-specific treatments. As specific mAb treatment examples,

Herceptin and Perjeta are two monoclonal antibodies targeting HER2 product, which is a tumor associated antigen that is overexpressed in 20%-30% of breast cancer tumors [8]. These two antibodies are specifically designed for the treatment of HER2-positive breast cancer. CD20 is another good example of this category. CD20 is a B-cell antigen and it is a good target for treating B cell lymphomas and leukemia. There are over one dozen mAbs targeting CD20 which have been approved by FDA (rituximab, obinutuzumab, tositumomab etc.) or are still under development in clinical trials.

The beauty of non-specific treatments is that instead of searching for specific antigens in cancer patients, we can simply block the cancer progression or immune suppression pathways. One good example is Avastin, which targets the vascular endothelia growth factor A (VEGF-A). VEGF-A is an important growth factor for blood vessels and thus a key element in tumor angiogenesis. Avastin has been approved to treat multiple types of cancer including colorectal cancer, lung cancer, breast cancer, renal cancer and brain cancer. However, Avastin is not perfect as well and there are many limitations. Many health services have restricted the using of Avastin due to the minimal benefit but large expenses for cancer patients [9].

Blocking immune suppression is another non-specific method to treat cancer, and with this method, the most popular group of mAbs - immune checkpoint inhibitors, has been developed recently. Immune checkpoints are regulating molecules which can activate or suppress the immune system. Immunosuppression has been a key obstacle for successful immunotherapy in late stage cancer patients.

Based on the immunoediting theory [10], cancer cells can escape from the immune system after the elimination and equilibrium phase. In the escape phase, cancer cells can get rid of highly immunogenic tumor antigens, down-regulate immune recognition molecules like MHC class I or co-stimulatory molecules, increase the expression of immunosuppression molecules like PD-L1 (Programmed death-ligand 1), galectins, IDO (Indoleamine-2,3-dioxygenase) etc. At the same time, MDSCs (myeloid-derived suppressor cell), M2 macrophages and DCs in tumor microenvironment can express immunosuppressive cytokines like IL-10 and TGF-β. They can also induce the generation of regulatory T cells. T regulatory cells then express inhibitory receptors such as PD-1, CTLA-4 (Cytotoxic T-Lymphocyte Associated Protein 4), Tim-3 and LAG3 to suppress anti-tumor response further [11]. Therefore, it is necessary to overcome the tumor immunosuppression for the anti-tumor T cells to kill the tumor. PD-1/PD-L1 and CTLA-4 are three hot candidates in this field. Ipilimumab which targets CTLA-4 was approved by FDA in 2011 for late-stage melanoma treatment. PD-1 inhibitor Nivolumab was approved in 2014 for treating melanoma, lung cancer, kidney cancer and Hodgkin's lymphoma. PD-L1 inhibitor Atezolizumab was granted accelerated approval for treating bladder cancer. Many other similar mAbs targeting PD-1/PD-L1 and CTLA-4 are still under development and some are recently approved (Durvalumab is approved on May 1 2017 and Avelumab is approved on May 9 2017) [114,115], as well as other immunosuppressive targets [114].

The immune checkpoint inhibitors are very promising drugs but still have at least two limitations: the objective response rate for these mAbs in different cancer types is not

high and the number usually varies from 10%~40% [12-14]; the cost is over $100K for one course of treatment. These two limitations may impede the use of ICI in developing countries.

1.3 Personalized cancer vaccines and neo-antigens

The latest immunotherapy method is cancer vaccine. Vaccines for infectious diseases have proved their great values and these vaccines have eliminated or controlled multiple infectious diseases. Nowadays vaccines are administered to children as standard care in most countries. The world would be much better if we could develop cancer vaccines as well using the same principles. However, the cancer vaccine field has faced many clinical failures during the last few decades. Big pharmaceutical companies and biotech companies used to focus on developing cancer vaccines with tumor associated antigens (TAAs) or tumor germline antigens. These vaccines usually did not pass efficacy trials. Meanwhile, two preventative cancer vaccines which target HBV (Hepatitis B Virus) and HPV (Human papillomavirus) are approved for preventing virus induced cancers (Hep B) or STD (HPV). However, the HPV vaccine is expected to prevent cancers. And only 2 therapeutic cancer vaccines are approved in people: Sipuleucel-T (dendritic cell based cellular vaccine) which was mentioned previously and one oncolytic virus (T-VEC) therapy for treating patients with metastatic melanoma which could not be surgically removed. The mechanism behind T-VEC is that this virus could only replicate in cancer cells but not in healthy cells due to disrupted stress response system in cancer [15]. There is one therapeutic vaccine (Oncept) approved for melanoma in dogs [116].

There are two reasons for the disappointing cancer vaccine trials: overwhelming immunosuppression in tumor microenvironment needs to be overcome; tumor associated antigens are bad antigens as they are not very immunogenic and can cause self-tolerance. It was also suggested that late stage cancer is too difficult to overcome.

However, with the combination of immune checkpoint inhibitors and next generation sequencing technology, this area has attracted huge interests again. The tumor immunosuppression can be overcome by using checkpoint inhibitors as adjuvants, and the trials of using tumor associated antigens (MUC1, HER2) in combination with immune checkpoint inhibitors are being tested again.

Meanwhile, analysis of the results from the immune checkpoint inhibitor trials pointed to neo-antigens as the key element for effective ICI response. One report demonstrated that in melanoma patients with anti-CTLA-4 antibody therapy, long-term clinical benefit was associated with mutational load and specific neoepitopes were associated with long-term benefit [16]. In another case report of a 56-year-old male with melanoma who had been treated anti-CTLA4 antibodies, specific T cells targeting two neo-antigens with point mutations were detected [17]. A recent report further showed that immunogenic neo-antigens were better biomarkers for predicting patients with objective response in ICI trials [18]. All this evidence supports that neo-antigens are better choices for cancer vaccine development over TAAs. But one important disadvantage for this approach is that there are no common neo-antigens between cancer patients from the exome sequencing results [19]. To design cancer vaccines for patients, the individual cancer patient must be sequenced to identify specific genomic mutations that might be

immunogenic. It has been estimated that only a few percent of these mutations will be immunogenic [116].

In summary, the development of personalized cancer vaccine includes 7 steps: isolate DNA/RNA from tumor tissue; Construct DNA and cDNA libraries; whole exome sequencing and RNA sequencing; identify mutations in sequence results; select best epitopes by MHC prediction or other algorithms; formulate vaccines using peptides, DNA vectors or other delivery methods; add adjuvants like immune checkpoint inhibitors or other immune adjuvants [20].

Even with the complexity of the personalized vaccine approach, big pharmaceutical and biotech companies are strongly interested in this area. Advaxis, Moderna, Merck, Amgen, Gritstone Oncology etc. decided to invest hundreds of millions of dollars into this concept. Even with the same idea, each company provides their own platform to construct or deliver the vaccine to boost the immune response. Here I have summarized different techniques and potential advantages or disadvantages for each strategy in Table 1-1. In summary, there are five available strategies now: traditional peptide vaccine; mRNA vaccine in nanoparticles; oncolytic virus vaccine; DNA vaccine; bacteria based vaccine like listeria vaccine [21-24].

Preliminary results for two personalized cancer vaccines have been reported so far and both vaccines were tested in melanoma patients. One group immunized 6 late stage melanoma patients with 13-20 long overlapping peptides (15-30 aa) from mutated sequences. All 6 patients showed positive T cell immune responses against vaccines and 4 patients had no recurrence within 25 months after vaccination. 2 patients with

recurrence were treated with ICI and the tumor regressed afterwards [25]. The second group used RNA vaccination instead of peptides with the same concept. Patients had positive T cell immune responses against vaccination and they experienced significantly less recurrence after vaccination. One patient was not responsive to vaccination due to lack of β2-microglobulin in the cancer cells [26]. Both personalized cancer vaccines showed a cautious promising preliminary result with neo-antigens.

Table 1-1 Personalized Cancer Vaccine Strategies

Company Name	Technology	Advantage and Disadvantages	Reference
Advaxis	ADXS-NEO based technology: use attenuated Listeria monocytogenes transformed with specific plasmids encapsulating many patient's immunogenic neoepitopes	Can induce T-cell responses against weak antigens; can cover hundreds of neo epitopes; potential antigen competetion problem	[21]
Caperna/Moderna	mRNA personalized cancer vaccine with Merck's anti-PD-1 mAb Keytruda	mRNA vaccine is believed to work effectively at very low dose and induce immune response very fast; mRNA vaccine cost is high and not stable	[22]
Inovio and AstraZeneca	SynCon DNA vaccine with IL-12 encoded and checkpoint inhibitors, electroporation delivery	DNA vaccine manufacture cost is low and vaccine is stable, can induce strong T cell response	[23]
PsiOxus and BMS	Oncolytic virus vaccine Enadenotucirev with anti-PD1 mAb Nivolumab	One oncolytic virus therapy T-VEC has been approved in 2015, can induce strong immune response	[24]

Personalized cancer vaccines are very promising, but they also have limitations. Since the vaccine market is aimed, at least initially, at late stage cancer patients, the timescale of making such a vaccine is crucial. Currently, companies need 6-12 weeks to generate the vaccine (15-18 weeks from resection to vaccine delivery in two previous reports) and they are aiming at the reducing the manufacture time to 1 month. But some argue that even 6 weeks is optimistic [27]. Meanwhile, this approach relies on the immunogenic genomic point mutations. Peptides from point mutations are not very immunogenic due to being single amino acid changes and strict antigen selection algorithms need to be applied. Current MHC prediction methods are not accurate and this makes the antigen selection step very difficult. Meanwhile, the tumor needs to have enough mutations for selecting immunogenic neo-antigens. It is estimated that only ~13% of point mutations would bind MHC I and far less could be validated by mass spectrometry [28]. Melanoma or lung cancer patients have thousands of point mutations in their tumors, but patients with pancreatic cancer or neuroblastoma may not have enough mutations for making such a vaccine. The last disadvantage of personalized cancer vaccine is still the cost-benefit problem. The immune checkpoint inhibitor treatment alone can cost over $100K for a course of treatment. When combined with genomic sequencing and vaccine design, the price will be even more (most estimates are ~$50K). This would mean that this approach would not be widely available to those with cancer.

It is widely agreed that the combination of neo-antigen cancer vaccines with checkpoint inhibitors is the best strategy to treat cancer. But instead of making the complexed personalized cancer vaccine, to develop a cancer vaccine with low cost and high

efficiency is more promising and urgent for the public. The key element in making such a cancer vaccine is to find common, tumor specific, neo-antigen and we need to find a new path for searching for specific for neo-antigens. To date, all the research effort has been focused on finding neo-antigens produced by genomic mutations. It is known that in every step of the biological process from DNA replication to gene transcription to mRNA splicing to protein synthesis to protein assembly, cells are making a lot of random errors. The natural error rate of these biological processes is astounding, so the cells have evolved numerous quality control system at every biological process: proofreading of exonuclease, DNA repair system, nonsense mediated decay at mRNA, proteasome to remove aberrant proteins and beyond all these correcting mechanisms, cells have apoptosis signaling pathway to kill themselves in an emergent situation. Even with so many correction mechanisms to reduce the errors, random errors are still inevitable. The random errors in DNA replication have already been shown to be the most important factor in tumor formation [3]. And at each step of information transfer far more new errors are introduced into the system due to the errors of enzymes or radiation, free radicals or other unknown reasons, the error rate increases exponentially from the initial step to the final step. Based on this theory, it will be much easier to find tumor neo-antigens in mRNAs or in proteins, which are downstream of the DNA replication, yet very little effort has been put into these two categories. While the search of neo-antigens in proteins was difficult before, it will be a very promising direction for finding neo-antigens as the MALDI technology improves. mRNA is relatively easier and more straightforward to analyze. My thesis is focused on uncovering one type of tumor neo-

antigens in mRNA: the frameshift mutations in coding mono-repeat microsatellite regions and mis-splicing events.

1.4 Coding Mono-repeat Microsatellite Regions and Microsatellite Instability in Cancers

Microsatellite DNA is a large family of repetitive DNA sequences consisting of DNA motifs with 1-5 base pairs. It is also known as simple sequence repeats (SSRs). Microsatellites are primarily used as a genetic marker for kinship analysis and genetic relatedness between subspecies [29]. Microsatellite instability (MSI) has also been found to play an important role in neurological diseases [118] and cancer.

Microsatellite mutations are different compared to point mutations used in personalized cancer vaccines. Point mutations (PMs) change only one nucleotide, missense PM changes one amino acid and neutral PM do not change the amino acid sequence, while microsatellite mutations result in the gain or loss of repeat units. If the repeat unit consists of 3 nucleotides, the mutations will cause amplification or contraction of certain amino acids, but when the repeat unit is 1/2/4/5 nucleotides, the mutations can cause frameshift mutations. Translation will be shifted in frame to produce frameshift peptides. The mutation mechanisms can also be totally different from point mutations. Replication slippage of the DNA polymerase in the lagging strand during DNA replication can be an important mechanism in long microsatellite mutations [30].

There are over 20 unstable microsatellite repeats which can cause neurological diseases. Most of microsatellite repeats are tri-nucleotide repeats (CAG, CTG, CGG etc.), and they are located in the coding region or 5'UTR and 3'UTR of different genes. The pathogenic mechanisms can be protein loss-of-function, gain-of-function or RNA gain-of-function, with the mechanism of some diseases still unknown. The most famous example is the expansion of trinucleotide $(CAG)_n$, which will generate insoluble poly Q aggregates and cause neurological disorders [31].

Unlike microsatellite instability in neurological diseases, microsatellite instability in cancer is related to the impairments of the mismatch repair (MMR) system. DNA mismatch repair is a system to recognize and repair errors in DNA replication and recombination. DNA polymerase has an internal error rate of $10^{-4} – 10^{-5}$ at the nucleotide insertion step. With the help of the proofreading exonuclease the error rate can be reduced to $\sim 10^{-7}$. Escaped mistakes will then be corrected by the mismatch repair system, which will further reduce the error rate by 50 to 1000-fold [32]. So, it is apparent that the mismatch repair system is essential for maintaining DNA integrity during replication, and any mutation-induced inactivation or epigenetic silencing of the system will cause microsatellite instability [33], which is common in various types of cancer including stomach cancer, esophageal cancer, head and neck squamous cell carcinoma, non-small cell lung cancer, colorectal cancer and melanoma etc.

Another important feature in cancer microsatellite instability is that mono-nucleotide microsatellite mutations are detected much more frequently. One report has analyzed microsatellite patterns in exome and whole-genome sequencing data of 147 colorectal

cancer samples and 130 endometrial cancer samples from TCGA (The Cancer Genome Atlas). Out of total 10K microsatellite instability (MSI) events in colorectal genome sequences, di-nucleotide and tri-nucleotide mutations only account for 7.6% of the total mutations, while 92.4% MSI events are mono-nucleotide mutations. A strong bias pattern of mononucleotide mutations also exists in the genomic sequences. Based on their findings, one nucleotide and two nucleotides deletions events have much higher frequencies than one nucleotide insertions in both colorectal cancer and endometrial cancer. Microsatellite frameshift mutations do not solely exist in MSI positive cancer patients. For example, TGFBR2 gene has a 10A homopolymer and the mutation rate of two nucleotide deletion was over 50% in MSI-positive samples, but one nucleotide deletion mutation was detected in MSI-negative sample with lower mutation rate (~10%) [24]. This indicates that microsatellite mutations in genomic sequences exist in MSI negative or even in the normal population but with much lower mutation rate. Another genomic sequencing report had similar findings. They detected over 2,000 MSI events in one MSS (microsatellite stable) cancer patient, which was much less mutations compared to the numbers in MSI cancer patients but still a considerable amount of mutations [25]. MSI events are also much more enriched in 5'UTR, 3'UTR, non-coding sequences than in coding regions, which may be related to high selection pressure of mutations in coding regions. Another interesting finding is that most of the MS frameshift mutated genes are under-expressed (loss of function) with only a very small portion of frameshift mutated genes over-expressed. These over-expression genes may offer a selective advantage in cancer development [34].

Microsatellite Instability is also an important biomarker for predicting survival in MSI cancers. Higher MSI burden is significantly associated with better survival of cancer patients. The pattern of microsatellite instability distributions in different genes can be used as a classifier to separate different types of MSI cancers [35].

With the development of immune checkpoint inhibitor (ICI) therapy, MSI can be an important biomarker to predict clinical benefit after ICI treatment. It has been shown that the number of neo-antigens is associated with clinical benefit in ICI treatment [16]. Since MSI cancer patients have large numbers of frameshift neo-antigens from the microsatellite regions, obviously MSI status should correlate with clinical benefit of ICI treatment as well. A clinical trial has confirmed this hypothesis, MSI status can be used as a predictive marker for response to PD-1 (pembrolizumab) blockade treatment in stage IV cancer patients. In the trial, 40% of MMR-deficient colorectal cancer patients (N=10) and 71% of MMR-deficient non-colorectal cancer patients (N=7) had object response, while 0% of MMR-proficient patients (N=18) had object response [36]. Recently the FDA, in a milestone decision, approved the use of Merck's PD-1/PD-L1 inhibitor Keytruda for MSI-H patients with unresectable or metastatic solid tumors regardless of cancer types.

Besides in human malignancies, Microsatellite Instability is also detected in benign diseases especially in benign lung diseases. D Spandidos detected 23% of chronic obstructive pulmonary disease (COPD) samples were positive in at least one microsatellite marker [37]. Microsatellite instability was also detected in asthma, sarcoidosis, idiopathic pulmonary fibrosis [38]. Usually patients with these lung diseases

will have higher probability of developing into malignant tumor, these findings may indicate that microsatellite instability can be a marker for precursor cancer or microsatellite instability can exist in different kinds of diseases which we have never thought of.

1.5 Microsatellite Frameshift neo-antigens as cancer vaccine source

Till now, microsatellite instability research has been focused on finding mutations in genomic sequences. People have found that microsatellite regions are very unstable and microsatellite mutations in genomic sequences can still be frequently detected in various MSI tumors, MSS tumors and even in benign lung diseases. However, if we consider the fact that the DNA polymerase error rate is around 10^{-10} to 10^{-8} per generation [39], while the RNA polymerase error rate is 10,000-fold higher and it is around 10^{-6} to 10^{-5} [40], each mRNA molecule is then translated into 2000-4000 protein molecules [41] and the error rate is amplified over another 1000-fold. It is obvious that microsatellite mutations should be more enriched in mRNA and proteins than in DNA sequences, and it is very likely that common microsatellite frameshift neo-antigens can be found from mRNA sequences.

Table 1-2. Advantages of RNA-based Frameshift Antigens over DNA –based Personalized Cancer Antigens

	Antigen Source	Antigen Abundance	Antigen Type	Immunogenicity & Autoimmunity	Cost

Personalized Cancer Neo-antigens	DNA	10^{-10} to 10^{-8}	Point Mutations	Low May cause autoimmunity	over $100K
Microsatellite Frameshift Neo-antigens	mRNA	10^{-6} to 10^{-5} 10,000-fold enrichment	Frameshift Antigens	High Low chance of causing autoimmunity	~$200 depend on delivery system

I have mentioned that Biotech and Pharmaceutical companies have decided to pursue the hardest way to find neo-antigens: massive genome sequencings for point mutations. The advantages of using frameshift mutations in microsatellite regions have been summarized in Table 1-2. There are 3 major advantages: 1. Microsatellite frameshift antigens from mRNA regions are more abundant due to the high mutation rate; 2. Frameshift antigens usually consist of peptides from several amino acids to hundreds of amino acids, which are likely to have high immunogenicity and low chance of causing autoimmunity; 3. The cost of making a vaccine from microsatellite frameshift antigens is much lower without the need of sequencing and analyzing the whole genome of the patient.

1.6 Frameshift antigens from mis-splicing events

Microsatellite regions are one type of error-prone DNA sequences where frameshift mutations may frequently occur at the mRNA level during transcription. But it is not the

only source. As mentioned before, random errors in the transcription and splicing process are more abundant than genomic mutations, microsatellite FS mutations are from the transcription process and another type of FS mutations can be caused by random errors in the splicing which are mis-splicing FS mutations.

There are 8.8 exons per human gene and alternate splicing increases the diversity of proteins [119]. It is reported that 95% of multi-exon genes in human are alternatively spliced [120] and it is difficult for spliceosome to accurately splice every time with such complexity. Mis-splicing events generate unexpected proteins and in most cases quality control machineries in the cell can handle these aberrant proteins. However, mis-splicing events are commonly detected in multiple diseases. One study estimated that over 60% of human disease-causing mutations alter the splicing process instead of changing the coding sequences [121]. For example, amyotrophic lateral sclerosis (ALS) and retinitis pigmentosa are caused by mutations of important genes in the splicing process [122].

Mis-splicing events are widely detected in cancer cell lines and in human tumor EST (Expressed Sequence Tag) libraries [123-125]. However, little attention has been paid to investigate the frameshift variants from mis-splicing events and the applications of these frameshift peptides. Previous graduate students in our group have analyzed mis-splicing frameshift mutations which are enriched in tumor EST libraries and results indicate that there are over 400 frameshift peptides from mis-splicing within the same gene (majority) or across different genes (gene fusions), tens of mis-splicing events were validated in the tumor cDNA samples or cancer cell line cDNA samples. Several mis-splicing FS

peptides are also detected in the mouse tumors and they could offer protection from tumor challenge in both mouse breast cancer and melanoma models [126,127].

These preliminary results proved the feasibility of using mis-splicing frameshift antigens for cancer vaccine development. However, tumor EST libraries only provide limited data for detecting all possible mis-splicing events in cancer patients. The development of high-throughput peptide array technology makes the unbiased screening of all mis-splicing frameshift peptides possible and we are currently working with Roche Nimblegen for this goal. I have built a human exon junction FS peptide database which covers all potential mis-splicing frameshift peptides. In this thesis, I have described the characteristics of this database and its potential applications in Chapter 6.

My thesis includes three parts: the first part is focused on building databases for MS frameshift antigens in different species, demonstrating the biological principles of using MS frameshift antigens for cancer vaccine development, developing algorithms to select MS FS antigens and test in mouse models; the second part involves developing a dog prophylactic cancer vaccine and a small dog safety trial with our vaccine; the third part expands our antigen searching to all possible FS antigens in human genes and explore the potential of "early diagnosis, early treatment" in mouse transgenic tumor model with immune checkpoint inhibitors.

CHAPTER 2

Characterization of FS Peptides from MS Regions and Its Advantages to be Used as Cancer Neo-antigens

2.1 Abstract

With emerging evidence supporting that cancer neo-antigens are the key elements for successful immune checkpoint inhibitor treatment [16, 42], the idea of developing a cancer vaccine with neo-antigens thrives again after decades. Currently the most popular neo-antigen source is the personalized genomic point mutations which has the limitations of high cost and low efficiency. A recent report indicated that FS antigens from insertions and deletions (INDELs) of microsatellite regions (MS) and mis-splicing events were much more abundant at mRNA level than in DNA, and frameshift (FS) peptides from these mutations offered protection in different mouse tumor models (in review). However, thousands of MS regions exist in the human coding sequences and these potential FS peptides have not been investigated systematically. Here I provided a comprehensive analysis of predicted FS peptides from mono-repeat MS regions. We constructed a predicted MS FS peptides database for human, dog and mouse. There are over 7,000 MS FS peptides with a minimum peptide length of 8aa in the human coding sequences. In addition, we summarized the characteristics of MS distribution and natural selection pressures over MS regions and FS peptides. Long FS peptides were highly enriched in MS FS mutations and a group of extremely long FS peptides (>200 amino acids) showed unique amino acid composition and codon usage patterns. In addition,

human MS FS peptides were predicted to be better cancer neo-antigens than point mutations with much higher MHC I and MHC II binding affinity. These data provided a rich resource for cancer vaccine development with MS FS peptides.

2.2 Introduction

Cancer vaccines have suffered several decades of failures in clinical trials using tumor associated self-antigens. Neo-antigen is the remaining choice for developing preventive or therapeutic cancer vaccines. With the appearance of immune checkpoint inhibitors (ICI) and their great success in cancer clinical trials and in practice, the whole cancer vaccine field is revitalized. Objective Response rate for anti-CTLA4 and anti-PDL1 treatments is around 10%~40% in clinical trials and the rate is higher when used in combinations [12-14]. More importantly, mutation load is reported to be closely associated with the clinical benefit of ICI treatments [16, 42] and T cells are directed to target neo-antigens in ICI treatments [43]. So, it would be ideal if cancer vaccines composed of neo-antigens and ICI treatments can also be used in combination therapy. However, with the analysis massive genomic sequencing data, it is clear that common neo-antigens do not exist at the genomic level. One report demonstrated that there is no CD4/CD8 neo-antigen epitope which has frequency higher than 0.2% in a cohort of 20,000 human tumor samples [44]. While it is nearly impossible for finding common neo-antigens based on genomic sequencing results, our lab showed that FS mutations from MS regions and mis-splicing events are much more abundant at the mRNA level than DNA level, and FS peptides from these mutations can offer protection in multiple

mouse tumor models (in review and unpublished data). Therefore, we have constructed databases which include all possible FS peptides from mono-repeat MS regions in human, dog and mouse, and we provided a comprehensive analysis of these MS regions and FS peptides in this chapter.

MS regions are known to be very unstable even at the genomic level and MS instability is associated with multiple neurological diseases [31], 18 types of cancer diseases [35], and also in multiple benign lung diseases [38]. MS regions have high mutation rates due to the slippage of DNA polymerase in DNA replication. One specific gene ACTBP2 with an MS repeat length of 15A, was reported to have a mutation rate as high as $7*10^{-3}$ per locus per generation in humans [45]. Similar to the findings of a high correlation between the mutation load and ICI treatment clinical benefit, high MS mutation burden in certain cancers is also reported to be associated with longer survival of cancer patients compared to low MS mutation burden [35]. MSI-H cancer patients were reported to benefit from PD-1 blockade treatments [36] and the FDA granted accelerated approval for using MSI as biomarker for PD-1/PD-L1 inhibitor treatments. These data indicate that an anti-MS FS peptide immune response is an important factor of anti-tumor activity in MSI-H patients.

Past research on MS FS mutations has been focused on characterizing a small number of genomic mutations in MSI-H cancers patients [33, 46-48], and the evolution patterns of MS regions across species [49, 50]. No report has characterized the FS peptide sequences from potential MS FS mutations. MS regions have long been used as a genetic marker but not as a focus as a potential cancer neo-antigen source. With the growing

interest in cancer neo-antigen discovery, it is important to analyze these potential FS antigens from MS regions. In this chapter, I have provided a comprehensive analysis of the potential MS FS antigens in human, dog and mouse. In addition, specific patterns of these MS regions and FS antigens are also summarized and the data provides important information for cancer vaccine development.

2.3 Results

2.3.1 Distribution of mono-repeat MS in coding regions of human, dog and mouse

Mono-repeat MS regions are more prevalent in non-coding regions but they are also in coding regions. Only MS FS mutations in coding regions will be translated into FS peptide sequences and used as potential neo-antigens. Mono-repeat MS regions with a minimum repeat length of 7 and predicted one nucleotide insertion/deletion FS peptides were included in the analysis of this report.

One example of a MS containing gene is human RNPC3. The mRNA and peptide sequence is shown in Fig 2-1. This gene has 12 continuous adenine repeats and a predicted 1A insertion generates insertion FS peptide (KV) with 2 amino acids, while 1A deletion generates deletion FS peptide (GLMTLSKMIKKK …) with 17 amino acids.

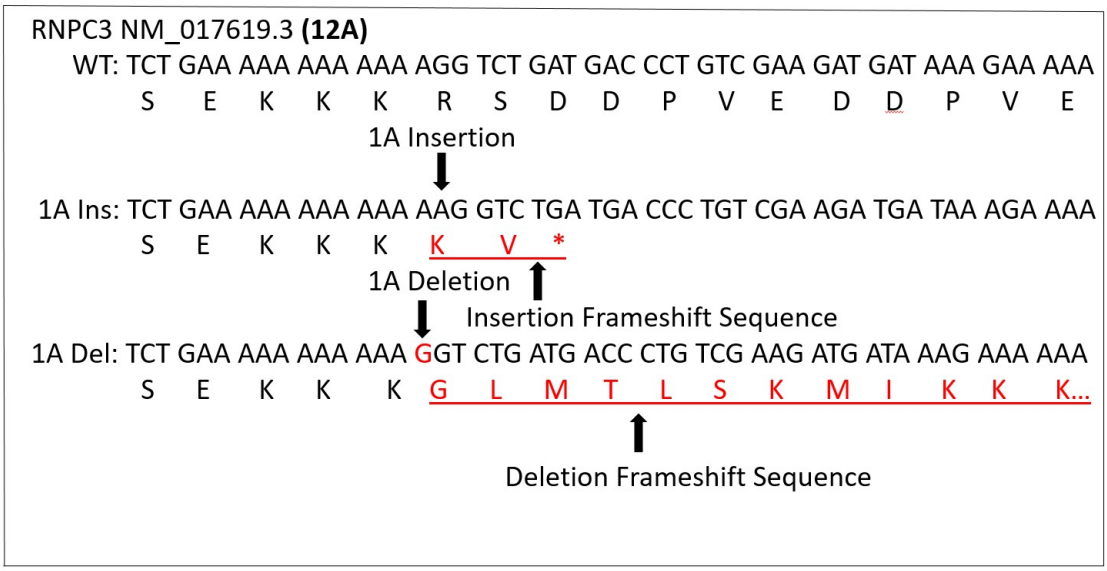


Figure 2-1 Mono-repeat MS region of human RNPC3 gene. Human RNPC3 gene (NM_017619.3) has 12A homopolymer, predicted one nucleotide insertion and deletion events can generate predicted insertion FS peptide (KV) and deletion FS peptide (GLMTLSKMIKKK...).

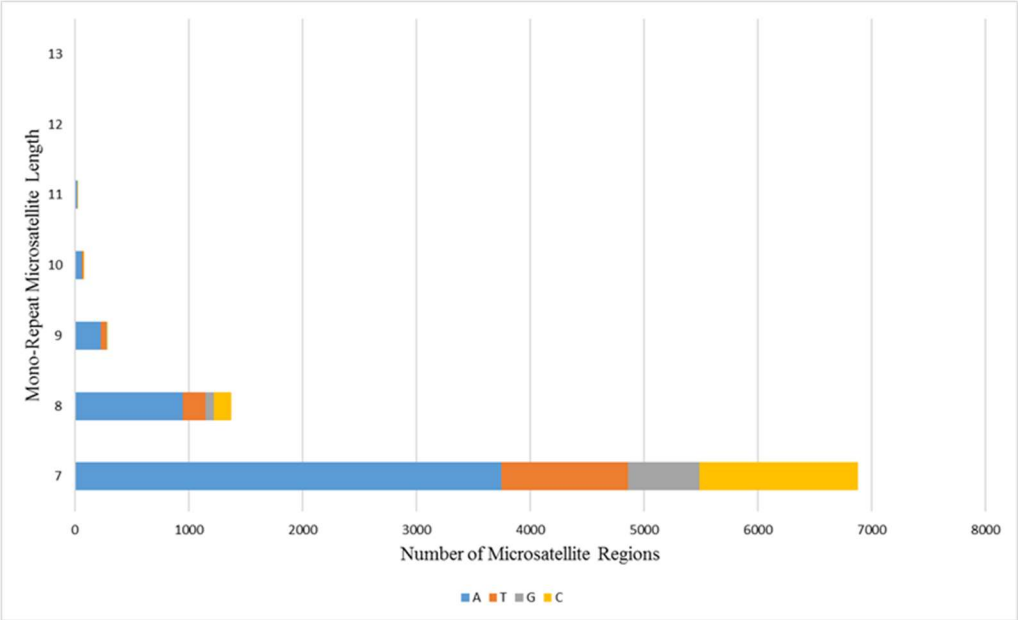
To analyze all possible MS regions and potential FS peptides, human and mouse coding sequences were acquired from the Consensus Coding Sequence database [51]. Dog coding sequences were acquired from NCBI RefSeq database [52].

In total, there are 8,617 human mono-repeat MSs in coding regions, 10,182 in dog and 4,837 in mouse. The distribution of human mono-repeat MS regions in Fig 2-2 shows that the number of MSs increases exponentially as MS length decreases. Nearly 80% of all the MSs have only 7 mono-repeats. Adenine (A) homopolymer is the dominant MS type and accounts for 58% of all the MSs while T and C homopolymers are in the middle range with 16% to 18% of all MSs, and G homopolymer has the least number of candidates with 8% of the total MSs. Mouse MS distribution has a similar MS type and

MS length distribution as in human (Fig2-3B), while dog has more G MS regions (13%) and C MS regions (30%). In addition, A and T MS regions are the dominant MS type in long MS regions (10 repeats or above), which is same as in short MS regions, while C repeats are the dominant MS type in dog long MS regions (11 to 15 repeats) (Fig2-3D).

With the differences in distribution of MS regions, it would be interesting to know whether these MS regions and FS peptides are under natural selection pressure.

A

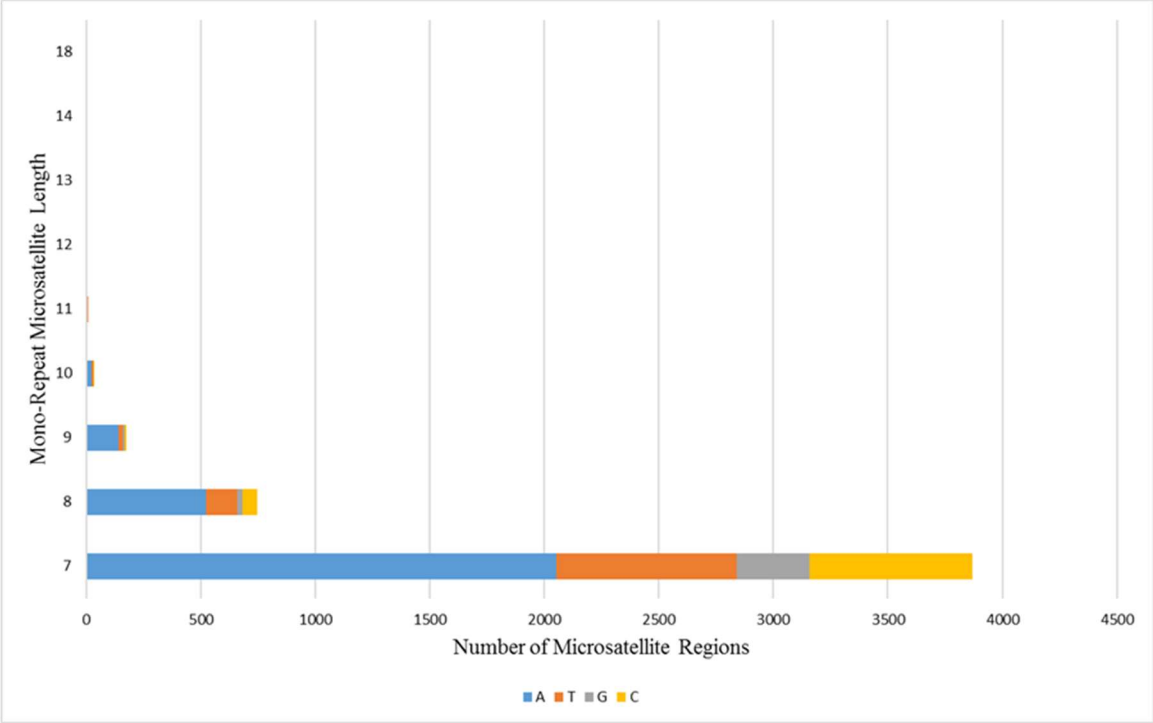


B

MS Type / MS LEN	A	T	G	C	SUM
7	43.42%	12.86%	7.22%	16.17%	79.68%
8	10.99%	2.30%	0.80%	1.77%	15.86%
9	2.58%	0.54%	0.08%	0.12%	3.32%
10	0.71%	0.12%	0.00%	0.01%	0.83%
11	0.21%	0.01%	0.00%	0.01%	0.23%
12	0.02%	0.00%	0.00%	0.01%	0.03%
13	0.01%	0.02%	0.00%	0.00%	0.03%
SUM	57.94%	15.86%	8.10%	18.10%	100.00%

Figure 2-2 Distribution of MS Regions in Human coding regions. A. Number of MSs by mono-repeat length and type. B. Percentage of MSs by mono-repeat length and type.

A

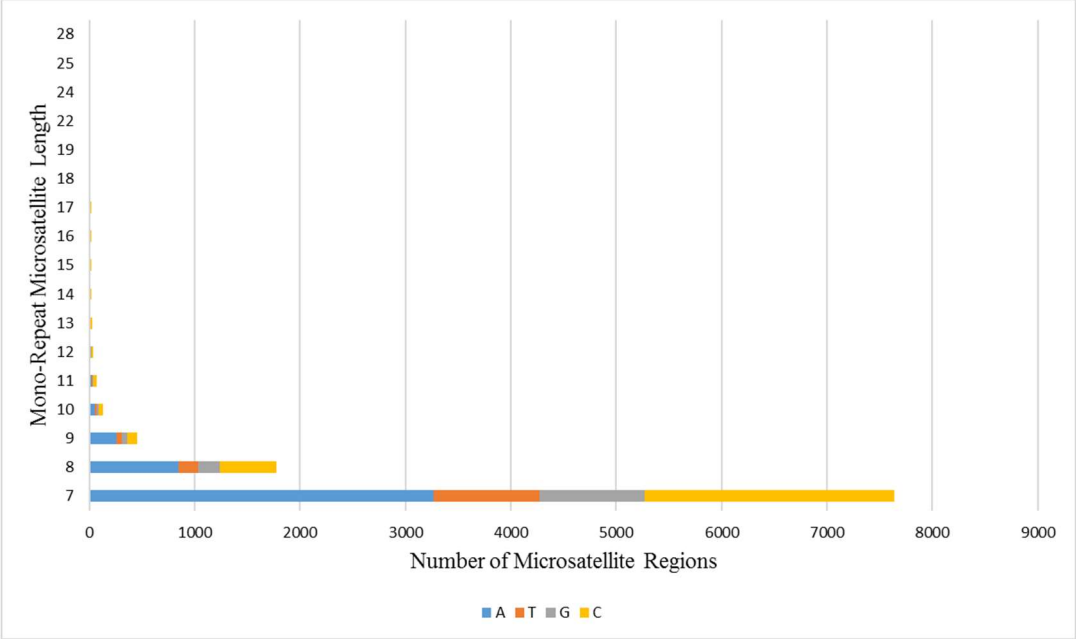


B

MS Type / MS Len	A	T	G	C	SUM
7	42.46%	16.21%	6.57%	14.74%	79.99%
8	10.83%	2.79%	0.50%	1.32%	15.44%
9	2.92%	0.41%	0.10%	0.17%	3.60%
10	0.52%	0.08%	0.00%	0.02%	0.62%
11	0.10%	0.02%	0.00%	0.00%	0.12%

12	0.02%	0.00%	0.02%	0.00%	0.04%
13	0.02%	0.08%	0.00%	0.00%	0.10%
14	0.00%	0.04%	0.00%	0.02%	0.06%
18	0.00%	0.02%	0.00%	0.00%	0.02%
SUM	56.87%	19.66%	7.19%	16.27%	100.00%

C



D

MS Length\MS Type	A	T	G	C	SUM
7	32.10%	9.82%	9.84%	23.29%	75.04%
8	8.33%	1.86%	1.98%	5.25%	17.42%
9	2.52%	0.50%	0.47%	0.92%	4.42%
10	0.54%	0.12%	0.23%	0.36%	1.25%
11	0.19%	0.08%	0.08%	**0.31%**	0.66%
12	0.06%	0.04%	0.11%	**0.11%**	0.31%
13	0.11%	0.01%	0.02%	**0.10%**	0.24%
14	0.06%	0.01%	0.04%	**0.08%**	0.19%
15	0.04%	0.01%	0.02%	**0.08%**	0.15%
16	0.06%	0.02%	0.01%	0.03%	0.12%
17	0.03%	0.01%	0.02%	0.04%	0.10%
18	0.02%	0.01%	0.00%	0.00%	0.03%
19	0.01%	0.00%	0.00%	0.00%	0.01%
22	0.00%	0.02%	0.00%	0.01%	0.03%
24	0.00%	0.02%	0.00%	0.00%	0.02%
25	0.00%	0.01%	0.00%	0.00%	0.01%
28	0.01%	0.00%	0.00%	0.00%	0.01%
SUM	44.07%	12.53%	12.82%	30.58%	100.00%

Figure 2-3 Distribution of Mono-repeat MSs in Dog and Mouse coding sequences. A-B, Number of MSs and percentage distribution by mono-repeat length and type in Mouse coding sequences. C-D, Number of MSs and percentage distribution by mono-repeat length and type in Dog coding sequences.

2.3.2 MS type has a significant impact on selection pressure

MS regions were reported to be highly unstable only at the genomic level in small populations like MSI-H cancer patients, who have mutations in the mismatch repair systems. MS regions could be under selection pressure constantly even in the normal population or they might not be randomly distributed. At least two factors could be related to selection pressure: MS type and MS length. Adenine (A) and Thymine (T) mono-repeats are generally thought to be more mutagenic with only two hydrogen bonds, while Guanine (G) and Cytosine (C) mono-repeats are more stable with three hydrogen bonds. In addition, long MSs are known to have much higher mutation rate than short MSs [45].

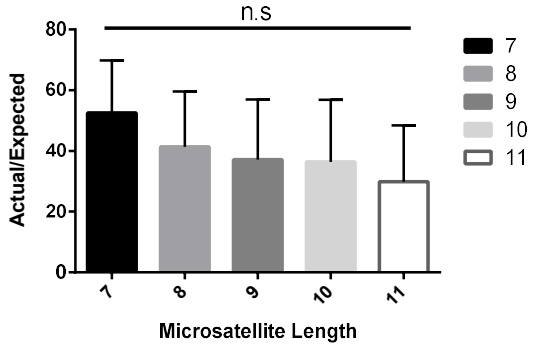
Selection pressure was calculated with the ratio of actual MS numbers to predicted MS numbers (no selection pressure would be randomly distributed) in the human coding sequences. A high ratio indicates low selection pressure while low ration means high selection pressure. The results showed that all four MS types were under selection pressure to some degree, but A and T MS regions had much lower selection pressure with 75.8% and 66.1% of actual MS numbers compared to predicted MS numbers. On the contrary, G and C MS regions only had 4.2% and 11.8% of actual MSs compared to predicted numbers. G/C MS regions had significantly lower ratio (higher selection pressure) than A/T MS regions (Fig2-4).

Another important factor was repeat length. It turned out that selection pressure went much higher when repeat length increased from 7 (52.6%) to 11 (29.9%), but it was not

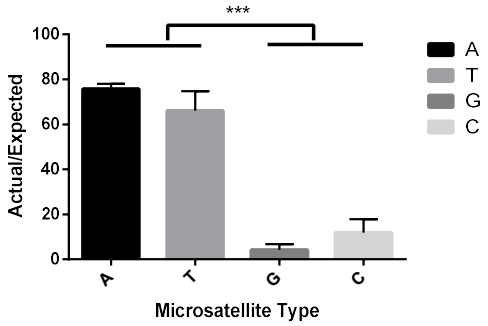
statistically significant. Selection pressure over A repeats did not change when repeat length increased from 7 to 11 (79.6% and 81.2% correspondingly). However, selection pressure did increase a lot for T repeats (83.5% to 32.9%), G repeats (13.3% to 0%) and C repeats (33.9% to 5.6%).

In addition to the characteristics of MS regions, we analyzed the distribution and patterns of MS FS peptides predicted from the MS regions.

A



B



C

MS Type	Actual#	Expected#	ratio	MS Type	Actual#	Expected#	ratio
7A	3750	4710	79.62%	**10G**		85	0.00%
7T	1111	1331	83.50%	**10C**	1	70	1.43%
7G	624	4706	13.26%	**11A**	18	22	81.18%
7C	1397	4111	33.98%	**11T**	1	3	32.87%

8A	949	1234	76.92%	**11G**		22	0.00%
8T	199	291	68.40%	**11C**	1	18	5.58%
8G	69	1232	5.60%	**12A**	2	6	34.44%
8C	153	1056	14.49%	**12T**		1	0.00%
9A	223	323	69.01%	**12G**		6	0.00%
9T	47	64	73.88%	**12C**	1	5	21.74%
9G	7	323	2.17%	**13A**	1	2	65.73%
9C	10	271	3.69%	**13T**	2	0	N/A
10A	61	85	72.07%	**13G**		2	0.00%
10T	10	14	71.88%	**13C**		1	0.00%

Figure 2-4 Selection pressure in Human MS regions. Expected MSs were calculated by the frequency of A, T, G, C in human coding sequences. A-B. percentage of actual MS number over predicted MS number was used to measure the selection pressure of MSs, x axis were MS length and type. C. Detail table of the actual MS numbers and expected MS numbers for different MS types and lengths.

2.3.3 Distribution of and selection pressure on Predicted MS FS Peptides

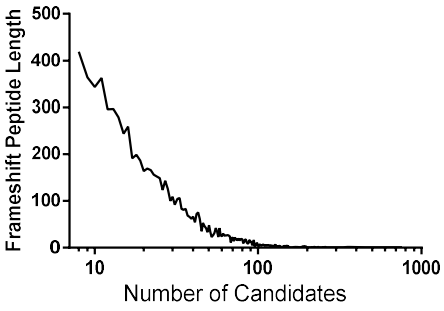
Coding FS peptides from MS regions are important for making cancer vaccines, since FS peptides are generally thought not to be functional proteins and we would expect them not to be long peptides if stop codons (TGA, TAA, TAG) are distributed randomly. In the meantime, relatively long FS peptides have the advantage of containing multiple high affinity MHC class I and II epitopes.

The human coding FS peptide sequences were used in our analysis. In the distribution of FS peptide length, the number of FS candidates decreased exponentially as FS peptide length increases (Fig2-5). Most FS peptides had less than 50 amino acids. But there were still over 300 FS candidates which had over 100 amino acids, with the longest FS peptide over 400 amino acids.

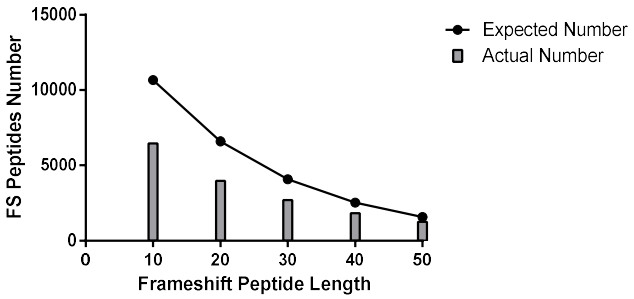
Long FS peptides could be under selection pressure considering they were more immunogenic and may have deleterious functions when expressed. The expected FS peptide length was calculated with a random distribution of stop codons in mRNA sequences and the selection pressure was quantified with the ratio of actual FS peptide length distribution over predicted distribution.

Results indicated that short FS peptides (<50 amino acids) are under selection pressure. There were 60% of the expected 20-mer FS peptides in the real human MS FS peptides. While long FS peptides (100 amino acids or longer) were more highly enriched in these MS FS peptides than expected. There are over 4-fold of actual FS peptides with 100 aa or longer than expected and the fold change increased as FS peptide length increases. The number of FS peptides with 50 aa or longer was about equivalent between actual and predicted number (Fig2-5B).

A



B



C

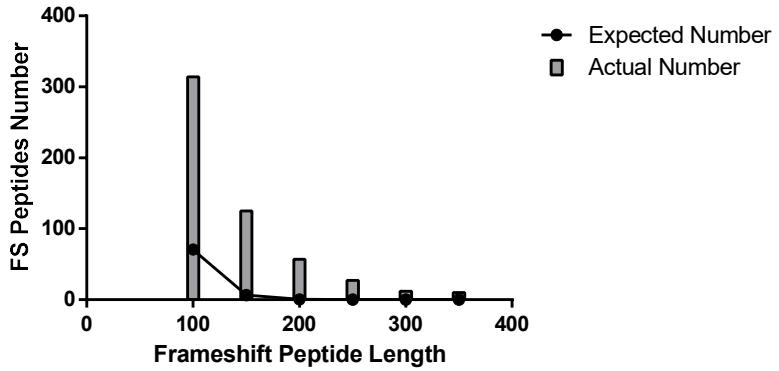


Figure 2-5 Distribution of Human MS FS peptide length. A. Distribution of human MS FS peptide length. B-C Selection pressure over the FS peptide length. Expected FS peptide length was calculated with random distribution of stop codons and selection pressure was quantified with ratio of actual FS peptide length distribution over predicted distribution.

At the genomic level, one nucleotide insertion and deletion events are biased. One nucleotide deletion is the dominant mutation in MSI cancer patients [34]. Therefore, it is also interesting to see that whether there is a bias in the peptide length of insertion and deletion peptides. FS peptides from one nucleotide deletion events are significantly

longer than one nucleotide insertion events by a few more amino acids (Figure 2-6). A small group of FS peptides from one nucleotide insertion events had extremely long peptide lengths with 300 amino acids or more, while there was no FS peptide with 300 aa in one nucleotide deletion events.

As mentioned previously, long FS peptides were more highly enriched in MS regions than expected and it would be interesting to know whether these FS peptides had specific characteristics.

A B

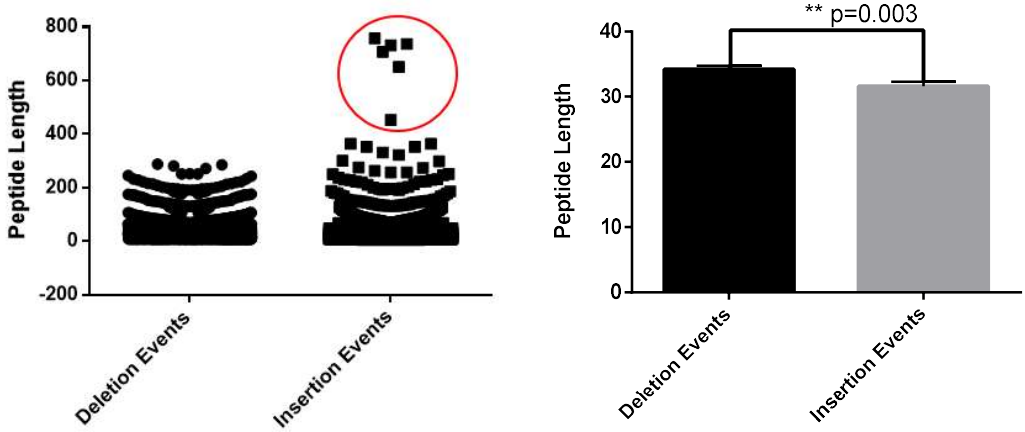


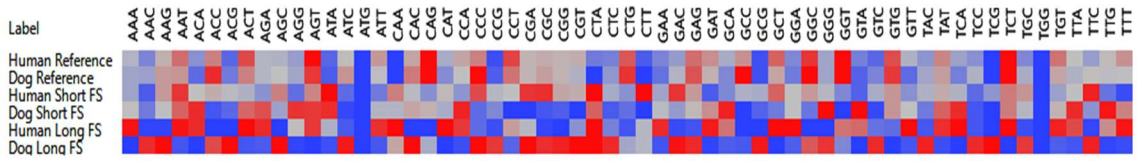
Figure 2-6 Insertion and Deletion FS Peptide Length Bias. A-B Comparison of peptide length of insertion and deletion events, a group of extremely long FS length is marked with red circle. Non-paired student's t-test is used to compare two groups.

2.3.4 Long FS peptides are a unique group of peptide sequences

Long FS peptides from MS regions with 200 aa or longer should not exist in human sequences based on probability calculations. However, this group was enriched in both human and dog microsatellite regions. One possibility would be that these long FS peptides were functional proteins as WT protein sequences. To test this idea and compare these FS peptides to the WT proteome, codon usage and amino acid compositions were compared among these three groups: long FS, short FS and WT peptide sequences.

Results of codon usage analysis showed that human WT codon usage and dog WT codon usage were almost identical with correlation of 0.995. Human and dog short FS peptides shared a similar pattern of codon usage compared to WT codon usage (correlation=0.92~0.93). Both human and dog long FS peptides had quite different codon usage patterns from short FS peptides (correlation 0.85~0.78) or the wildtype proteome sequences (correlation 0.88~0.89) (Fig2-7). Interestingly, human long FS peptides codon usage had very high similarity to dog short FS peptides (correlation 0.94) and very low similarity to dog long FS peptides (correlation 0.69).

A



B. Codon usage correlation between 6 groups

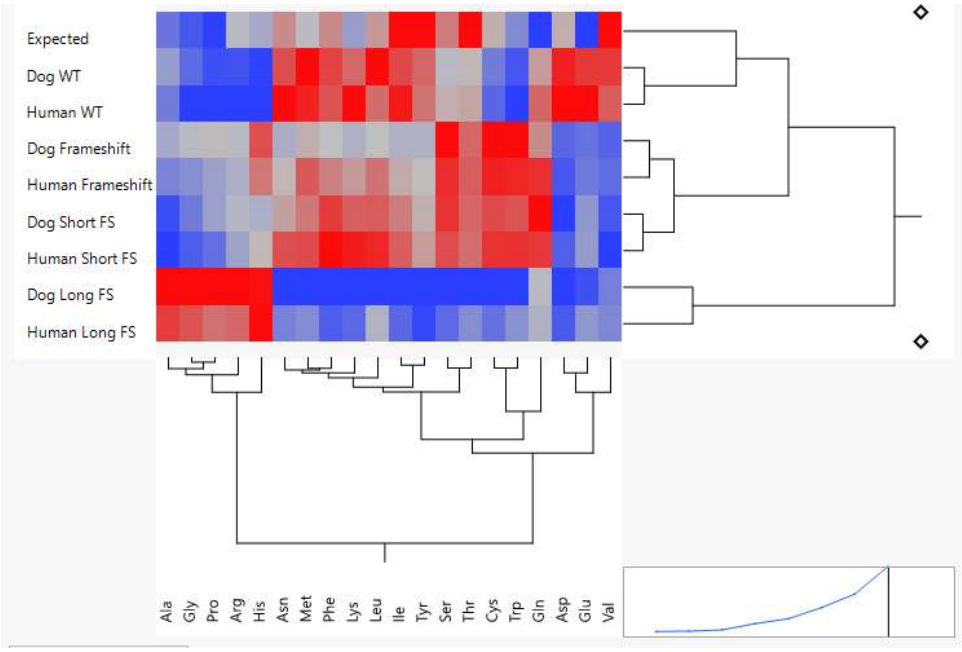
	Human Reference	***Human Short FS***	***Human Long FS***	***Dog Reference***	***Dog Short FS***
Human Reference	1.000				
Human Short FS	0.931	1.000			
Human Long FS	**0.884**	**0.847**	1.000		
Dog Reference	0.995	0.937	0.852	1.000	
Dog Short FS	0.938	0.915	0.942	0.922	1.000
Dog Long FS	**0.866**	0.911	**0.688**	**0.897**	**0.783**

Figure 2-7 Codon usage bias between long FS peptide versus short FS peptides. A-B, 64 dog long FS peptides with 300 aa or longer and 53 human long FS peptides with 200aa or longer were included in "long FS" group, FS peptides with 20 aa or shorter were

included in “short FS” group, human and dog wildtype proteome sequences were used as “WT”. Heat map of codon usage distribution was shown in A and correlation table of codon usage between different groups was shown in B.

Amino acid compositions were compared by the same method. The amino acid composition of human and dog WT sequences were very similar as well (correlation=0.981). WT protein amino acid composition was not randomly distributed (correlation to random was 0.7~0.79). However, the aa composition of FS peptides should be more randomly distributed if they were not under selection, and in fact whole human/dog MS FS database and short human/dog FS peptides had very high correlation to the random distribution (correlation ~0.9), which indicates that they were likely not under selection. However, both human and dog long FS peptides had unique amino acid compositions which were very different from other groups (correlation to WT (0.5~0.6), random distribution (0.6~0.7) or short FS sequences (~0.6)) (Fig2-8). Despite that different codons were used between human and dog long FS peptides (Fig2-7), they were highly conserved in amino acid composition (correlation=0.98). Alanine, Glycine, Proline, Arginine and Histidine were enriched in long FS peptides, while other amino acids were less abundant in these long FS peptides than short FS peptides or WT sequences.

A



B

	Expected	*Dog WT*	*Dog FS*	*Dog Short FS*	*Dog Long FS*	*Human WT*	*Human FS*	*Human Short FS*	*Human Long FS*
Expected	1.000								
Dog WT	0.792	1.000							
Dog Frameshift	0.885	0.749	1.000						
Dog Short FS	0.890	0.789	0.955	1.000					
Dog Long FS	**0.652**	**0.530**	**0.807**	**0.668**	1.000				
Human WT	0.713	0.981	0.663	0.743	0.398	1.000			
Human Frameshift	0.901	0.785	0.985	0.989	0.722	0.720	1.000		
Human Short FS	0.865	0.795	0.912	0.989	0.573	0.767	0.968	1.000	
Human Long FS	**0.728**	**0.631**	**0.865**	**0.740**	**0.986**	**0.502**	**0.796**	**0.659**	1.000

Figure 2-8 Amino acid composition bias between long FS peptides and short FS peptides. Expected: aa composition by random; Dog/Human WT: WT proteome aa composition; Dog/Human FS: whole MS FS database aa composition. A. Two-way hierarchical clustering of amino acid composition in human and dog peptide sequences, expected number is calculated based on number of codons for each amino acid. B. Correlation matrix of amino acid composition between different groups.

The unique characteristics of long FS peptides indicated they could have unknown functions. The simplest way to answer this question would be to see whether these FS peptides were used as functional proteins in another species. Out of a total of 53 FS peptides with 200 amino acids or longer in human, 40% of these candidates (total 21) have conserved peptide sequences in the wild type protein sequences of other species. While the percentage went down to 9% for mid-length FS peptides (31-50 aa) and 3% to short FS peptides (17-30 aa). However, when the percentage was normalized to FS peptide length, the rate was about the same between long FS peptides and short FS peptides, while mid-length FS peptides had higher chance of being found in WT proteomes of other species (Fig2-9). These results demonstrated that long FS peptides were not preferably used as functional proteins in other species, which was consistent with previous results that long human and dog FS peptides did not share the same codon usage or amino acid composition with WT sequences (Fig2-8). This group of 53 long FS peptides containing genes were not enriched in a specific biological process or molecular function based on ontology analysis, but most of these long FS containing genes had repetitive functional domains. These proteins included 12 DNA-binding proteins like zinc finger proteins and forkhead box proteins, and transmembrane proteins or other proteins.

In addition to these interesting characteristics, it was important to know why FS antigens would be better than the commonly used point mutations based neo-antigens.

A B

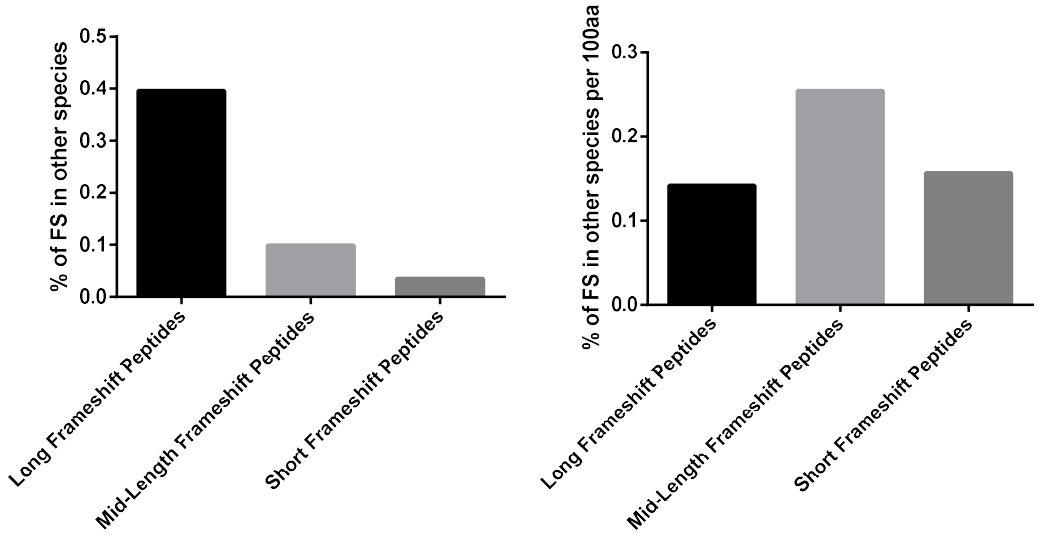


Figure 2-9 Long Human FS peptides were not preferably existed in wildtype protein sequences of other species. A. Human FS peptides were blasted against NCBI Refseq protein sequences of other species, percentage of FS peptides which had homolog wildtype sequences in another species was calculated. 53 long FS peptides (200 aa or longer), 1332 mid-length FS peptides (31aa~50aa) and 2133 short FS peptides (17aa~30aa) were used in the analysis. B. Percentage of FS peptides in WT sequences of another species was normalized to 100 amino acids of FS peptide length. E-value cutoff is 0.01.

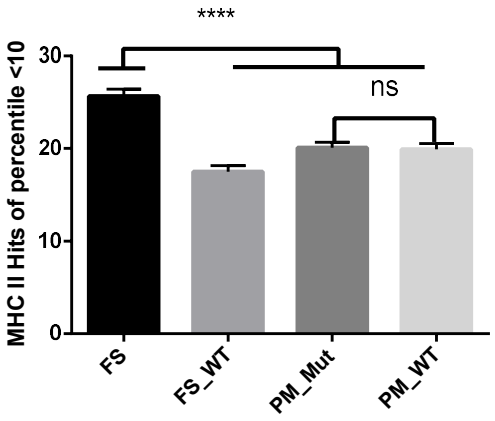
2.3.5 Advantages of using MS FS peptides over point mutations as neo-antigens

Patient-specific point mutations at the genomic level are the focus for use in personalized cancer vaccines. However, there are three potential advantages of using MS FS peptides instead of point mutated peptides: high immunogenicity; low chance of causing autoimmunity and low cost of making a vaccine from these FS peptides.

To compare the immunogenicity of FS peptides and peptides from point mutations, both MHC Class I and Class II predictions were performed between MS FS peptides and point

mutated peptides. 1,242 point mutated peptides and corresponding wild type peptide sequences were acquired from a previous published cancer genome sequencing report [53]. The same number of MS FS peptides with the same peptide length as the point mutations were randomly selected from human MS FS peptides. Results showed that MS FS peptides had significantly more MHC II prediction hits (high immunogenicity) with a prediction rank percentile of 10 or lower when compared to their WT counterparts or point mutation peptides, while overall the point mutation peptides had almost the identical MHC II prediction hits compared to wildtype sequences. Very similar results were observed in MHC I predictions. The best prediction score for each peptide was used in the analysis. The MS FS peptides had significantly lower percentile rank (high immunogenicity) than paired wildtype sequences or point mutation peptides, while there was still no difference between point mutations peptides and paired wildtype sequences. These results indicated that MS FS peptides were much more immunogenic on average than genomic point mutated peptides.

A



B

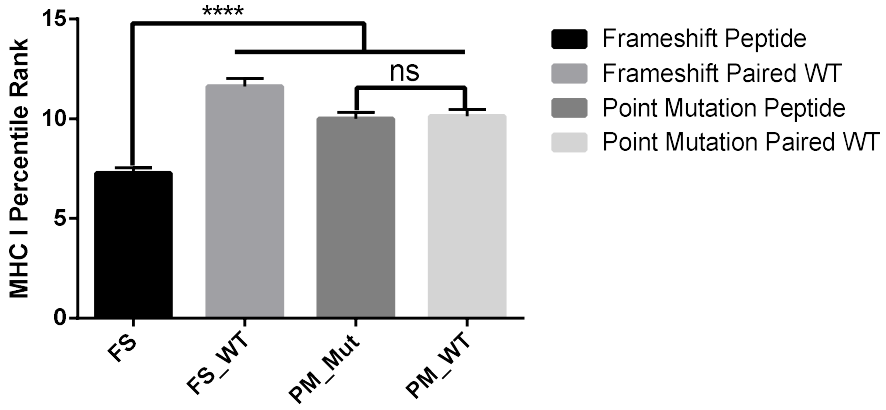


Figure 2-10 MHC Class I and II prediction for MS FS peptides and point mutation peptides. FS: MS FS peptide, FS_WT: FS peptide paired WT peptide, PM_Mut: point mutated peptide, PM_WT: point mutation paired WT peptide. Reference human HLA types are used. A. 1,242-point mutation peptides were somatic cancer genome missense point mutations,20 aa peptides were generated from the point mutation site, same number of 20-aa MS FS peptides and paired WT peptides were used for comparison, y axis was number of hits with percentile rank < 10 in MHC II prediction; B. 1,259 10-aa peptides were generated using the same method as in A for MHC I prediction, y axis was the best prediction percentile rank for each peptide, reference human HLA types are used.

Another advantage of MS FS peptides is that many MS regions are conserved across human, dog and mouse. It is easy, therefore, to transform vaccine candidates in mouse into dog or human trials directly. Humans and dogs share 1,124 homolog MS candidates, dog and mouse share 345 homolog MS candidates, and 269 MS candidates can be found in all three species. Vaccines developed with these 269 MS candidates can be tested in all three species without needing to change the components (Figure 2-11). This criterion is used in antigen selection for the dog cancer vaccine which will mentioned in Chapter 5.

In addition, FS peptides have a low chance of causing autoimmune diseases when compared to point mutations. Point mutated peptides only have 1 amino acid change and immune responses against these peptides can potentially spread to target the wildtype peptide sequence and cause autoimmune diseases, while FS peptides do not have this disadvantage. The cost of making a personalized cancer vaccine from point mutations was tremendous with genome/exome sequencing, antigen selection and vaccine development. While MS FS mutations at the mRNA level were abundant as reported and these predicted peptide sequences may make producing a personal vaccine much less expensive.

Table 2-1 Advantages of Microsatellite Frameshift Antigens over Personalized Cancer Antigens

	Antigen Source	Antigen Abundance	Antigen Type	Immunogenicity & Autoimmunity	Cost
Personalized Cancer Neo-antigens	DNA	10^{-10} to 10^{-8}	Point Mutations	Low May cause autoimmunity	over $100K
Microsatellite Frameshift Neo-antigens	mRNA	At least 10^{-6} to 10^{-5}	Frameshift Antigens	High No autoimmunity issues	~$200 depend on delivery system

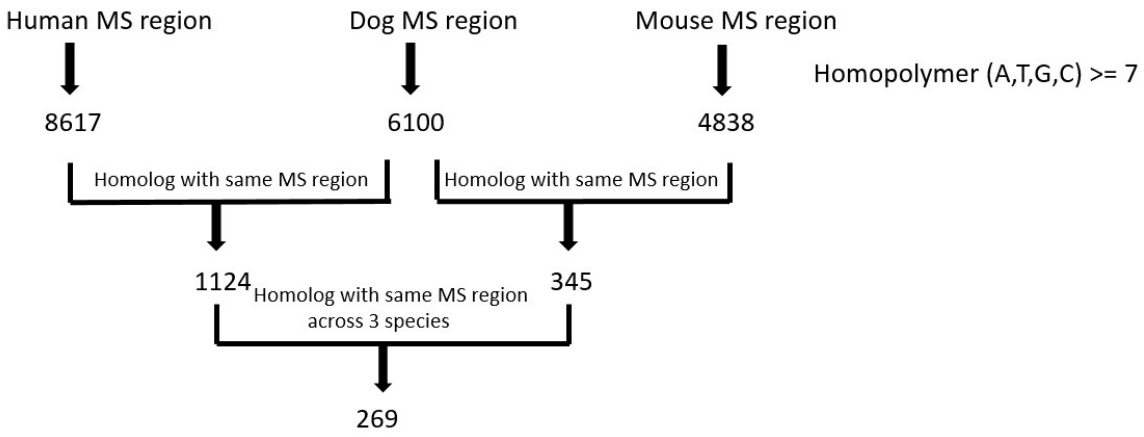


Figure 2-11. Microsatellite Homologs across human, dog and mouse. Homolog genes with same MS region were included in the analysis.

2.4 Discussion

Cancer neo-antigens are the key element for cancer vaccine development. Neo-antigens from patient specific genomic point mutations, the focus of current efforts, have severe limitations in high cost and a long production cycle. Here we have provided a comprehensive analysis of the mono-repeat MSs and predicted FS peptides from these MS regions. We have constructed an MS FS peptide sequences database for human, dog and mouse. Analysis showed that MS type was related to selection pressure and long FS peptides were highly enriched in human MS regions. A group of long FS peptides had unique codon usage and amino acid composition which could be related to repetitive domains of the proteins. We further demonstrated that MS FS peptides had higher MHC I and MHC II binding affinity compared to point mutations and low chance of causing

autoimmunity. These data support that at least FS derived from RNA-MS FS antigens could be a rich resource of potential cancer neo-antigens for cancer vaccine development.

In the MS regions analysis, selection pressure was much higher on the G/C repeats than the A/T repeats, which could be related to the stable triple hydrogen bond in G/C pairs compared to the double hydrogen bond in A/T repeats. The selection pressure on the MS length only went slightly higher as the MS length increased but the difference was not statistically significant. This suggests that MS regions could be relatively stable at the genomic level when the mismatch repair system was functioning properly. This is consistent with the findings that MSI-H cancer patients only accounted for a small percentage of all the cancer patient population. However, MS mutations at the mRNA level were much more abundant as reported.

Long frameshift peptides are a unique set of peptide sequences. They are specifically enriched in the microsatellite FS peptides while short frameshift peptides are under negative selection pressure. These long FS peptides have unique amino acid composition and codon usage patterns which is quite different from wildtype sequences and short frameshift peptide sequences. Five amino acids (Glycine, Arginine, Alanine, Histidine and Proline) are highly enriched in these long frameshift peptides. Two amino acids (Arginine, Histidine) are positively charged with the other three neutral amino acids. The positive charge of these long frameshift peptides may offer selective advantages. We first thought that these long frameshift peptides might be functional protein variants of the wildtype proteins. It turned out these three groups have about the same chance of being

in WT proteomes of other species, which indicates these long frameshift peptides are probably not known functional proteins.

With thousands of potential MS candidates available, we could not use all of them at once for vaccine development. The algorithms to select personalized MS FS antigens and broadly effective MS FS antigens remains to be investigated. B cell immune response and T cell immune response screening of these potential MS FS antigens may uncover some of these commonly reactive antigens in cancer patients.

In conclusion, thousands of MS FS peptides have the potential of being used as cancer neo-antigens. With the advantages of high immunogenicity, low chance of causing autoimmunity and low cost, the idea of developing prophylactic and therapeutic cancer vaccines with these predicted MS FS antigens can be helpful in reducing the cancer incidence and mortality rate in developing countries.

2.5 Materials and Methods

2.5.1 FS peptide database

Human and mouse mRNA sequences were acquired from the NCBI CCDS database [51], dog mRNA sequences from the NCBI RefSeq database [52]. Mono-repeat MS regions with a minimum of 7 nucleotides were predicted in coding sequences. One nucleotide predicted insertion and deletion peptides were predicted after the MS regions.

2.5.2 Selection pressure calculation

The expected number of MSs for a specific length and type in coding sequences was calculated by the following equation:

$$Expected\ Number(A,T,G,C) = total\ mRNA\ nucleotides * Frequency(A,T,G,C)^{Microsatellite\ Length}$$

The expected number of a certain FS peptide length was calculated by the frequency of non-stop codons (61/64) and FS peptide length:

$$Expected\ Number(Peptide\ Length) = total\ microsatellite\ number * (\frac{61}{64})^{Peptide\ Length}$$

2.5.3 Codon usage analysis

Human and dog reference codon usage numbers were acquired from the codon usage database [54]. FS peptide codon usage was analyzed with the original mRNA sequences and a codon usage heatmap was generated using JMP Pro 13 (SAS, NC).

2.5.4 Amino acid composition and peptide analysis

Amino acid composition was calculated by counting the number of each amino acid in the FS peptide and WT peptide sequences using Excel. Expected frequency of amino acids was calculated by the percentage of codons for each amino acid. Two-way hierarchical clustering with the Ward method was performed using JMP Pro 13. To find conserved FS peptide sequences which were used in wildtype proteins of other species, human and dog FS peptides were blasted against whole RefSeq protein database [55]. Blast hits were counted with the cutoff e-value of 0.01, then blast hits were normalized to the original FS peptide length.

2.5.5 MHC Prediction analysis

Human MHC II Reference sets [56] and MHC I Reference sets [57] were acquired from IEDB. The Consensus prediction method [58-60] was used for MHC I and II prediction. For MHC Class I prediction, the best prediction score for each peptide was used. In MHC Class II prediction, the number of predicted hits with percentile rank less than 10 were counted for each peptide. Point mutated peptides in cancer patients and corresponding WT peptides were analyzed from previous published cancer genome sequencing results [53]. 1,242-point mutation peptides were somatic cancer genome missense point mutations, 20 aa peptides were generated from the point mutation site (10 aa before the mutation site and 9 aa after). The same number of 20-aa MS FS peptides and paired WT peptides were used for comparison. 1,259 10-aa point mutated peptides (5 aa before the mutation site and 4 aa after) and WT peptides were generated using the same method for MHC I prediction.

CHAPTER 3

Microsatellite Frameshift Peptides are More Abundant at the mRNA Level and Offer Protection in the 4T1 Mouse Tumor Model

3.1 Abstract

Microsatellite instability (MSI) is well characterized as a phenotype for certain cancer types (colon cancer, stomach cancer etc.) due to the impaired DNA mismatch repair machinery [33, 61, 62]. But for most other cancer types, microsatellite mutations are rarely detected at the genomic level [63] and believed not relevant to cancer progression. However, microsatellite frameshift mutations at the RNA level have not been thoroughly investigated, largely because it is difficult to pick out partial single nucleotide insertion/deletion from the background noise in RNA sequencing technology. Here, we detected mono-repeat microsatellite frameshift mutations in tumor cDNA samples and show that mutations at the mRNA level are much more abundant than at the DNA level. Microsatellite frameshift mutation rates were strongly related to microsatellite repeat length. Frameshift peptides for one specific microsatellite candidate, SEC62, could be detected in both the human HEK293 cell line and mouse 4T1 cancer cell line with a GFP reporter system. We also developed a method to select the best 9 mouse MS FS peptides and this pool showed protection in the 4T1 mouse breast cancer model. Both T and B cell responses were correlated to the tumor protection. Our results demonstrated that instead of using MS as a biomarker, a large group of microsatellite frameshift peptides are also a potential cancer vaccine source.

3.2 Introduction

Microsatellite Instability (MSI) is caused by impaired DNA mismatch repair (MMR). The phenotype is characterized by hypermutability in microsatellite regions of MSI-H cancer patients. Microsatellite Instability is not very common in the general cancer population. In an exome sequencing study of 18 microsatellite instability cancer types, 15 cancer types had less than 5% of the patients with the MSI-H signature, while only endometrial cancer, colon cancer and stomach cancer have around 20%-30% MSI-H cancer patients [35].

Microsatellite instability is usually used as biomarker for cancer subtypes [48, 64]. Recently, it was reported that high microsatellite mutation load is beneficial to cancer patients compared to microsatellite stable cancer patients. High MSI burden is correlated with significantly better survival of cancer patients [35, 65], and microsatellite instability shows promise in checkpoint blockade immunotherapy [66-68]. These results indicate that the immune system can target these microsatellite mutations derived frameshift antigens and killing the tumor cells.

Both antibody responses and T cell responses against microsatellite frameshift antigens were detected in MSI-H cancer patients [69-71]. T cell responses against microsatellite frameshift peptides in microsatellite stable (MSS) cancer patients was much lower compared to MSI-H patients but still detectable. More interestingly, 2%~15% of the healthy controls also had antibody response against frameshift peptides [70], which was also lower than the percentage in the MSI-H patients. These data suggest that microsatellite mutations are expressed and presented in the healthy population with lower

level as well. Apparently, these mutations are more prevalent at the protein level than people have thought.

The mutation rate at the mRNA level is 10,000-fold higher than it is at the DNA level [40] and each mRNA molecule is translated into 2,000 – 4,000 proteins [41]. It is apparent that microsatellite frameshift mutations should be much more prevalent at the mRNA level or protein level in cancer patients. And when the fact of high mutation rate at mRNA level is combined with the evidence that the immune response against these MS frameshift peptides is beneficial to MSI cancer patients [5-11], these microsatellite frameshift peptides could be ideal cancer vaccine antigen sources. However, the application of these frameshift peptides as cancer vaccine has not been studied.

Here we have examined microsatellite frameshift mutations at the mRNA level and DNA level using dog tumor cDNA samples and 4 human breast cancer cell lines. We found that frameshift mutations are much more abundant at the mRNA level than the DNA level. Furthermore, we have constructed a plasmid with an 11A MS region and a GFP reporter. The mutation rate for this 11A MS in human HEK293 cell line is as high as 25.5%. In addition, a pool of 9 MS FS candidates was tested in the mouse 4T1 breast cancer model and the vaccine showed protection. Together these results proved that MS FS peptides at the mRNA level could be unrecognized source of cancer vaccine components.

3.3 Results

3.3.1 Microsatellite variants are much more abundant at the mRNA level than mutations at the genomic level

Microsatellite sequences are usually repeated sequences of 1-6 base pair units. Mono-repeat microsatellite mutations are the most abundant category in MSI-H cancer patients [35]. Here we only used mono-repeat microsatellite sequences in coding regions with microsatellite repeat length larger or equal to 7 repeats. One example of such a repeat (dog gene Sec62, XM_846664.4) is presented in Figure 3-1. Wild type Sec62 gene has 11 consecutive Adenine and a one nucleotide insertion or deletion can shift the reading frame, thus making insertion frameshift peptide RWGKGRA and deletion frameshift peptide MGKRKSLK….

To measure microsatellite mutations at the mRNA level, 20 dog microsatellite candidates with variable MS repeat length were sequenced in dog tumor cDNA samples from 4 major dog cancer types (Fibrosarcoma, Mast Cell Tumor, Osteosarcoma and Hemangiosarcoma). Sequencing results showed that the microsatellite regions rarely mutate at mRNA level when repeat length ranged from 7 to 9 with the exception that there was one nucleotide insertion mutation in the 9G MS region of B3GAT gene. As the mono repeat length increased to 10 or longer, most microsatellite regions had either insertions or deletions in the cDNA sequences. 10A in EIF2B3 and 11A in MS038 had both one nucleotide insertions and deletions. When the MS repeat length was not very large (11 repeats or less), mutations are detected in the cDNA sequences but rarely in genomic DNA sequences. One example was dog gene Sec62 We sequenced both

genomic sequences and cDNA sequences for the candidate in tumor cDNA samples and the corresponding DNA samples from same tissue. Mutations at cDNA level were much more abundant (1A insertions and deletions) than at the genomic level. There was a very small percentage of one nucleotide insertions in the genomic sequences which could rarely be seen (Fig3-1B&C). But as the repeat length increased to 15, as in MS305 with 15T microsatellite region, a significant amount of insertion and deletion mutations were detected at genomic sequences as well (Fig3-1D).

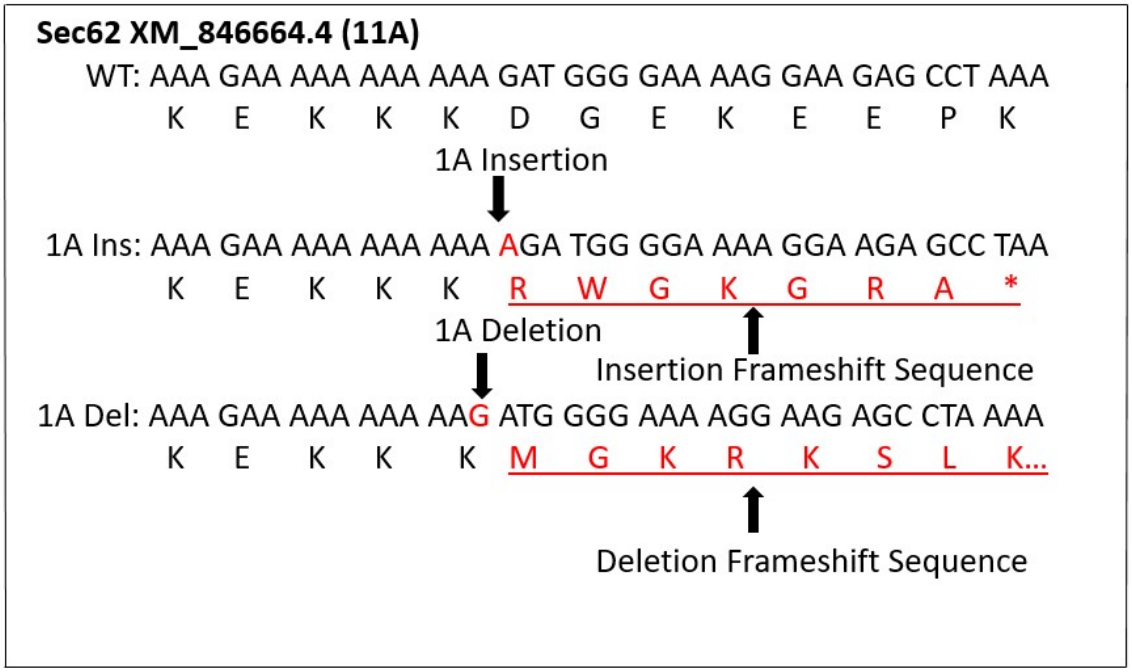


Figure 3-1 Example of mono repeat microsatellite in coding sequences. Dog Sec62 gene (XM_846664.4) include 11 adenine repeat, one nucleotide insertion and deletion frameshift peptides are marked as red.

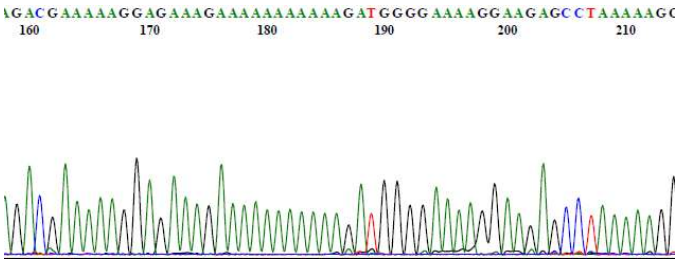
A similar approach to detect microsatellite mutations at the mRNA level and in genomic sequences was applied to human microsatellite candidates. Genomic DNA and cDNA of

four human breast cancer cell lines (HTB26, CRL2315, CRL2326 and CRL1504) were used in the sequencing. When microsatellite repeats were of length 10 or less, microsatellite mutations could be detected in cDNA sequences (one nucleotide insertion in MS443 and MS75, one nucleotide deletion in MS807) but rarely in genomic sequences. While for microsatellite candidates with 11 repeats, one nucleotide deletion events could also be easily detected in genomic sequences of candidates MS866, MS20 and MS87 (Fig3-3). Interestingly, only one nucleotide deletion was detected in the genomic sequences but not one nucleotide insertion, which was consistent with findings that one nucleotide deletion was the dominant mutation type of microsatellites in MSI-H cancer genome sequencing results [34].

A

Candidate	MS Type	Mutation at cDNA
MS136	7A	No
MS459	7A	No
MS683	7T	No
MS383	8A	No
SWAP70	8A	No
MS880	8A	No
MS507	8C	No
MS497	8G	No
HNRNPH1	8T	No
MS226	9A	No
ARV1	9A	No
OPHN1	9A	No
MS265	9G	Yes (1G Ins)
ZC3HAV1	9T	Yes (1T Ins/Del)
MS514	10A	Yes (1A Ins)
MS383	10A	Yes (1A Del)
CEP290	10A	Yes (1A Ins)
EIF2B3	10A	Yes (1A Ins/Del)
MS038	11A	Yes (1A Ins/Del)
EFHC2	11A	Yes (1A Ins)

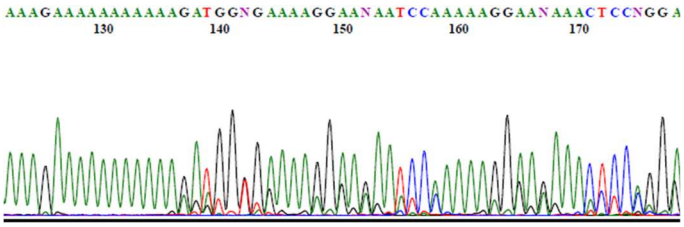
B



Genomic Sequences: Sec62, 11A, No mutation

C

cDNA Sequences: Sec62, 11A, 1A Ins / Del



D Genomic Sequence for MS305, 15T

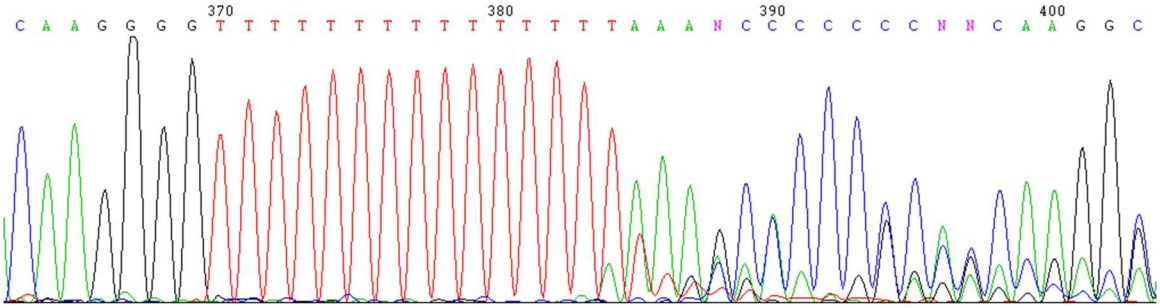
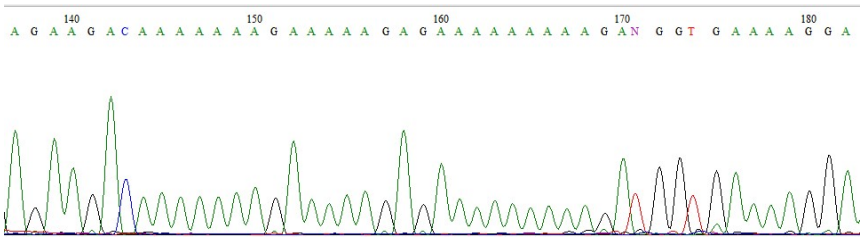


Figure 3-2. Microsatellite frameshift mutations at Dog tumor cDNA samples. A. Summary sequencing results of 20 microsatellite candidates in 3 dog cancer types (MCT, OSA, HAS); B&C. Genomic sequence and cDNA sequence of dog Sec62 gene in 5 dog tumor samples (MCT, OSA, HAS). D. Genomic sequence of dog candidate MS305 with 15T microsatellite region in dog tumor DNA samples.

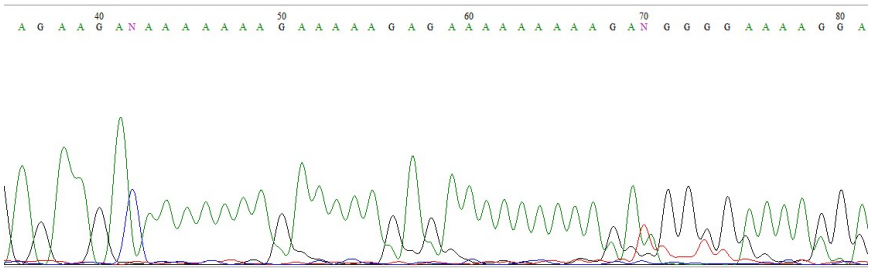
A

Candidate	MS Type	Mutations
MS93	8T	No mutation at cDNA&DNA level
MS24	9A	No mutation at cDNA&DNA level
MS990	9T	No mutation at cDNA&DNA level
MS443	9C	1C Insertion at cDNA level
MS687	9G	No mutation at cDNA&DNA level
MS75	9A	1A Insertion at cDNA level
MS806	10T	1T Deletion at cDNA level
MS866	11A	1A Deletion at DNA level
MS20	11A	1A Deletion at DNA level
MS87	11A	1A Deletion at DNA level

B. Human Sec62 9A NM_003262.3, Genomic DNA



C. Human Sec62 9A cDNA, 1A Insertion



D. Human MS866 11A, genomic sequence

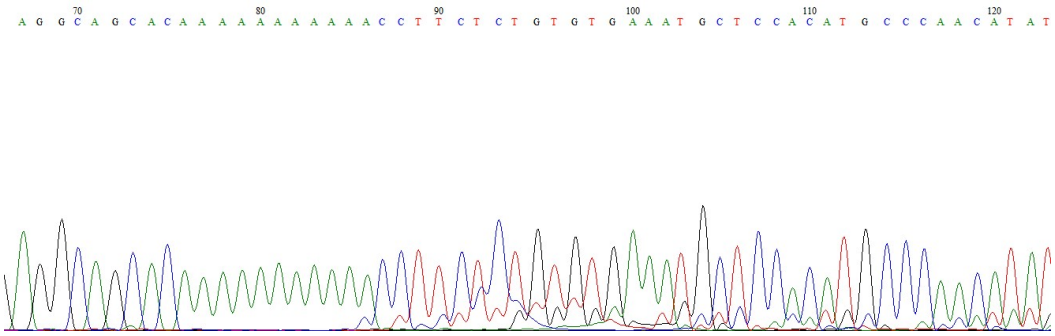


Figure 3-3. Microsatellite frameshift mutations at Human breast cancer cell lines. A. Summary results of 10 microsatellite candidates sequenced in HTB26, CRL2315, CRL2326 and CRL1504 breast cancer cell lines. B&C. Genomic sequence and cDNA sequences of human Sec62 gene. D. Genomic sequence for human MS866 candidate with 11A MS region.

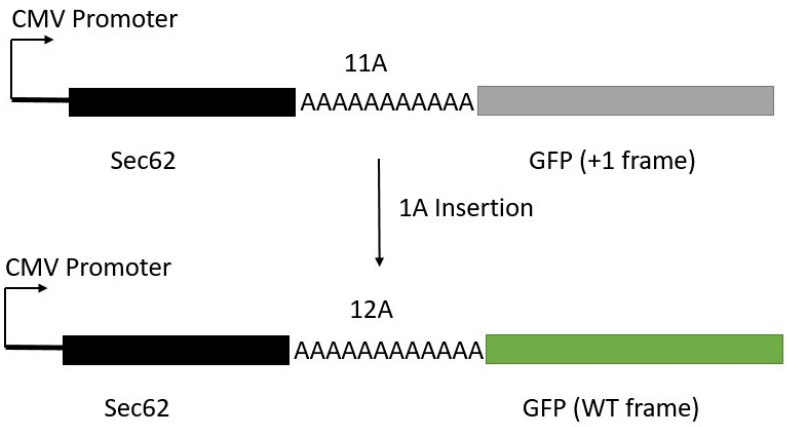
3.3.2 Translation of microsatellite frameshift mutation can be detected in both the human HEK293 cell line and the mouse 4T1 breast cancer cell line

Frameshift mutations in microsatellite regions usually cause premature termination codons (PTCs) and they are believed to be degraded by nonsense-mediated mRNA decay system (NMD pathway). However, some truncated frameshift proteins can escape the NMD system or bypass the NMD system if PTCs are located within 50-55 nucleotides of

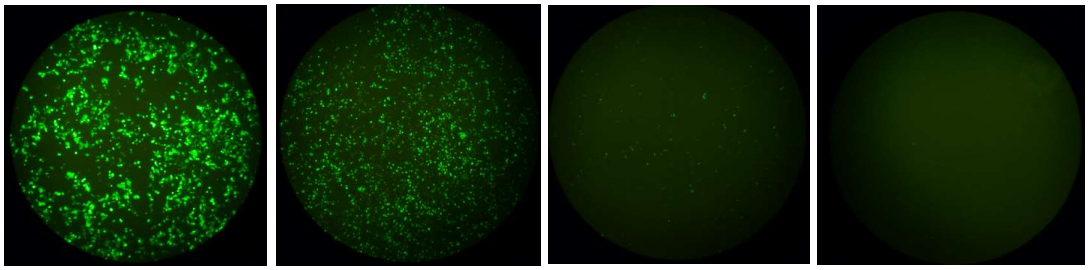
the last exon junction [72]. Frameshift proteins for these microsatellite regions can be detected. Another report estimated the overall NMD effect on INDELs was only 14% [18], suggesting that NMD would not prevent the translation of MS FS peptides.

To test whether frameshift proteins are translated, we used a different approach by constructing a plasmid which contained the coding sequences of Sec62 with 11A MS region followed by GFP in the +1 reading frame (an insertion frameshift mutation will induce GFP expression). Sec62 with 12A region and GFP in the WT frame were used as positive controls to calculate the mutation rate. The 11A microsatellite region was disrupted in Sec62-NM (No Microsatellite) construct and used as negative control.

A.



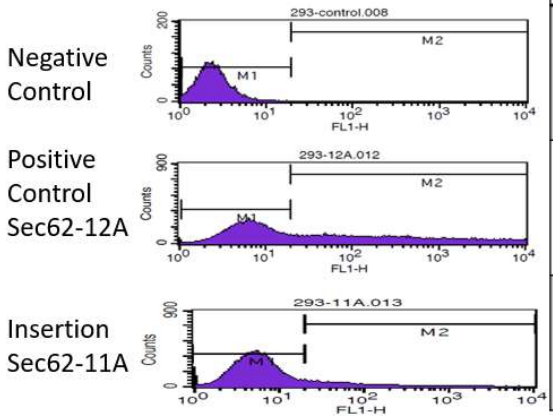
B.



GFP Positive Control Sec62-12A Sec62-11A Sec62-NM

Positive Control Negative Control

C.



	M1	M2
Negative Control	100%	0%
Positive Control Sec62-12A	49.90%	50.10%
Insertion Sec62-11A	87.23%	12.77%

Figure 3-4. Detection of frameshift protein in HEK293 cell line. A. Construct of plasmids, GFP is in +1 reading frame of Sec62-11A construct and WT frame of Sec62-12A construct. B. HEK293 cell line was transfected with plasmids, GFP fluorescence in Sec62-11A plasmid represented frameshift insertion events. C. Mutation events were counted in FACs, Sec62-12A was used as positive control.

When the human HEK293 cell line was transfected with Sec62-11A, GFP fluorescence from one nucleotide insertion events was detected as shown in Figure 3-4B. One nucleotide deletion events were rarely detected with only 0.45% of all cells were GFP positive (Figure S3-4). However, if the 11A microsatellite region was disrupted (Sec62-NM, used as negative control), the fluorescence disappeared completely, which indicated that intact 11A microsatellite region was essential for the INDELs. To measure the mutation rate, 80,000 cells from each group were collected after transfection and counted by flow cytometry. 50.10% of cells were fluorescence positive in Sec62-12A as positive control, while 12.77% of cells were fluorescence positive in Sec62-11A. The mutation rate based on this experiment is 25.5% for 11A microsatellite region, which was about the same level of mutation rate as measured by estimate from the cDNA sequence traces shown in Figure 3-2C. The same experiment was repeated in the mouse breast cancer cell line 4T1 (data in Figure S3-1, S3-2). 1A insertion and 1A deletion were both detected in the 4T1 cell line. However, the mutation rate was much lower in the 4T1 cell line than the HEK293 cell line due to a low transfection rate. Then same experiment was repeated in other mouse and human cancel cell lines like MCF7 and HCC1143. However, mutations could not be detected due to extremely low transfection rate since GFP was rarely detected even in positive controls (Figure S3-3).

3.3.3 Frameshift peptides derived from microsatellite mutations show protection in the mouse 4T1 cancer injection model

Frameshift mutations at microsatellite regions could be detected at the cDNA level based as shown above. However, to test whether frameshift peptides are a good cancer vaccine source, we need to test whether these FSPs protect mice from developing tumors.

Thousands of FSPs from mouse microsatellite regions are potential vaccine candidates. To select the best vaccine candidates, we designed a method to screen the microsatellite frameshift candidates (Fig3-5). As we demonstrated above, microsatellite repeat length is correlated with frameshift mutation rate. Microsatellites with repeat length of 10 or above usually had high frameshift mutation rate at the cDNA level as shown in previous data. MHC I&II predictions for BALB/c mouse were used to select candidates with the best MHC I&II epitopes. The top 9 candidates were selected with this method with their detailed information was summarized in Table 3-1.

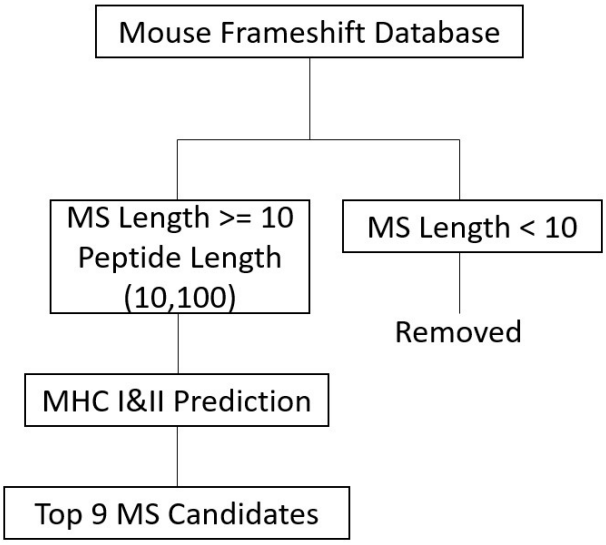


Figure 3-5. Method of selecting best microsatellite frameshift peptides. Minimum microsatellite repeat length was 10, frameshift peptide was longer than 10 amino acids and less than 100 amino acids, combination of MHC I&II prediction scores were used to select the best 9 candidates.

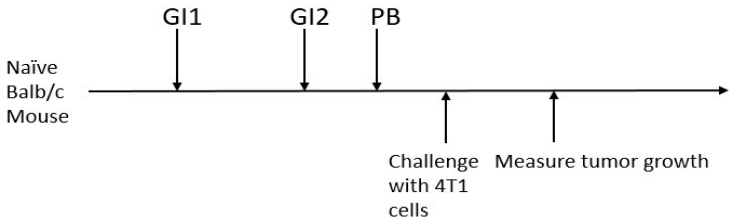
Table 3-1. Information of 9 Microsatellite Frameshift Peptides

Gene Name	mRNA Accession	MS Type	Frameshift Peptide
GM5464	NM_001034881.3	14_T	VTAGYQEEEMEASACGAKGPGLAPWP PSWLALQDSLLCVVVALADLRRKSCC
SLC35F5	NM_028787.4	10_T	LLCVVFGKFVIPRSTFRHTGCHSEYFVF NFWTFYFNPCSCISE
FAM71A	NM_001109759.1	10_A	RQKKIRPPKKKRSIQGQRQKPPRDHRC ECDQLFCFFWWGGNP
CCDC112	NM_001160399.1	11_A	RVYSKLENQKAAKEGGNTQVKRKGGH RASAFSKQSRR
NEMF	NM_025441.3	11_A	KAKEQAAAEAAEEQAAACRCGSQPVS LCQCQKIL
ANKRD45	NM_028664.1	18_T	FGVNGARRNSRIGEFRKVTIFLTARV
RFC3	NM_027009.2	10_A	LKSAPLQATTTLKLIPVMRGTATEL
CALR4	NM_001033226.4	13_T	FFSLSLSFLHRWMDKTVGTI
CHD2	NM_001081345.2	10_A	VLPNLPSQSSTF

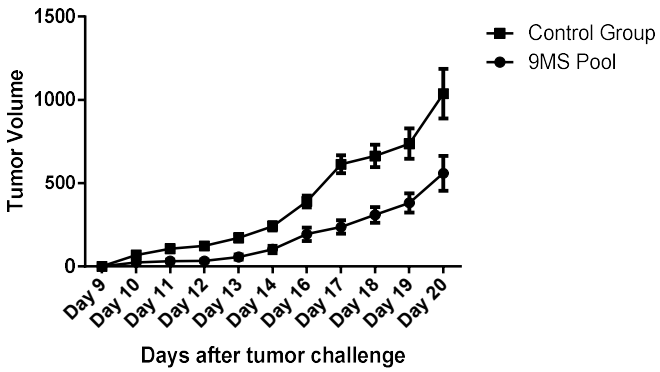
These 9 microsatellite frameshift peptides were pooled to immunize naive BALB/c mice and the mice then challenged with 5,000 4T1 mouse breast cancer cells [73]. 4T1 mouse breast tumor model is a well characterized transplantable mouse breast tumor model which can metastasize spontaneously. The immunization regime included two rounds of genetic immunizations (GI1, GI2) with a gene gun and then one dose of KLH conjugated peptide boost (PB). Empty plasmid and KLH only were used in genetic immunization and peptide boost correspondingly for the control group (Fig3-6A).

Tumor volume was measured over time after tumor injection. As shown in Fig3-6B, tumor volume was significantly lower in the vaccine group compared to the control group from day 11 after the tumor injection and the tumor progression was successfully suppressed. Tumor weights were measured at the termination time point. The vaccine group had significantly lower tumor weight than the control group as well. These results indicated that this 9 microsatellite frameshift peptides pool delayed tumor growth significantly in this mouse breast tumor model.

A.



B.



C

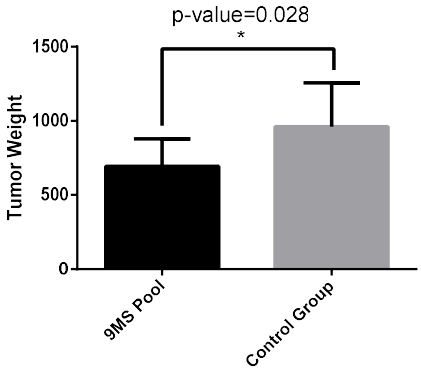


Figure 3-6 Nine Microsatellite frameshift peptides pool vaccination offer protection in mouse 4T1 injection model. A. immunization regime, GI1: 1st genetic immunization, GI2: 2nd genetic immunization, PB: peptide boost. B. tumor volume of 9MS pool vaccination group versus control group, each group had 10 mice. C. tumor weight comparison of vaccine group versus control group at termination point.

3.3.4 T Cell and B Cell Immune Responses Were Related to the Tumor Protection

T cells are known to be important in anti-tumor immune responses. To test whether the T-cell immune response was related to the tumor protection in this mouse model, 10 mice in the vaccine group were separated into two groups, 5 mice with lowest tumor volume were assigned into Low Tumor Volume (LTV) Group and 5 mice with highest tumor volume were assigned into High Tumor Volume (HTV) Group. Spleen cells were collected at termination point and pooled in the corresponding groups. IFN-gamma releasing T cells were counted after spleen cells were stimulated with immunized peptides pool. Results showed that the LTV group had many more spots than the HTV group. The control group had very few IFN releasing T cells. These data indicate that the T-cell response was directly correlated to tumor progression.

A. B.

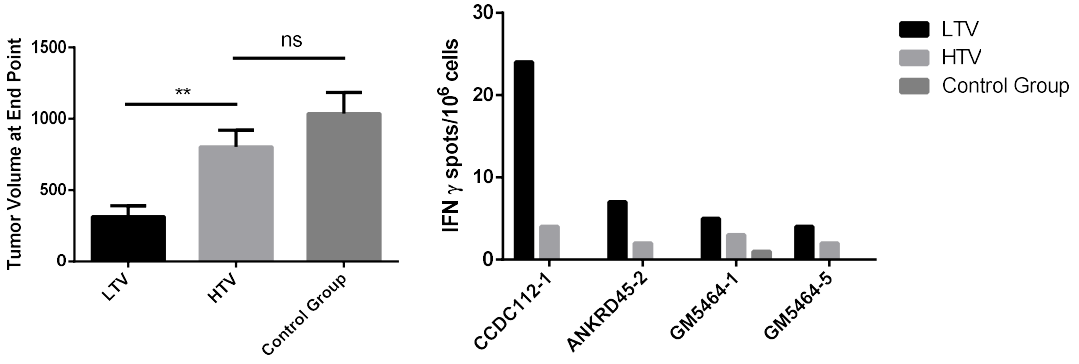


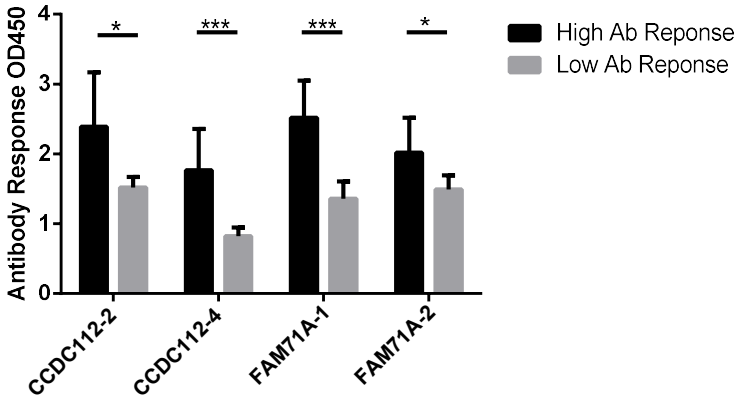
Figure 3-7. T cell immune response was related to tumor protection level. A. 10 mice in vaccination group were separated in LTV (low tumor volume) and HTV (high tumor volume group). B. IFN γ releasing ELISPOT was used to measure T cell immune response in LTV, HTV and control group. X axis represents the partial frameshift peptides which were used in the immunization, each frameshift antigen was divided into 17-mer non-overlapped frameshift peptides, CCDC112-1 represented the first peptide from CCDC112 frameshift antigen shown in Table 3-1 .

To test whether the B cell immune response was also related to tumor protection, serum samples after peptide boost were collected and the antibody response against each immunized frameshift peptide was measured by ELISA. Within the vaccine group, 4 mice had significantly higher antibody responses than the rest of the 6 mice and they were assigned to High Antibody Response (HAR) group. The other 6 mice were assigned to Low Antibody Response (LAR) group (Fig3-8A). Then we compared tumor volume between HAR group and LAR group at the termination point. Results show that HAR group had lower tumor volume than the LAR group (Fig3-8B). The p-value for one tail non-paired T-test was 0.052.

The CCDC112 frameshift peptide shows positive response in both the T-cell response and B-cell response. The T/B immune response against different portions of CCDC112 was correlated to tumor protection. Therefore, the CCDC112 frameshift peptides from the vaccine pool could have played an important role in the tumor protection. In addition, the ANKRD45 frameshift peptide in the T cell response and FAM71A frameshift peptide in B cell response were also correlated to tumor protection.

In summary, both T cell and B cell immune response assay showed that positive immune response was important in tumor protection.

A.



B.

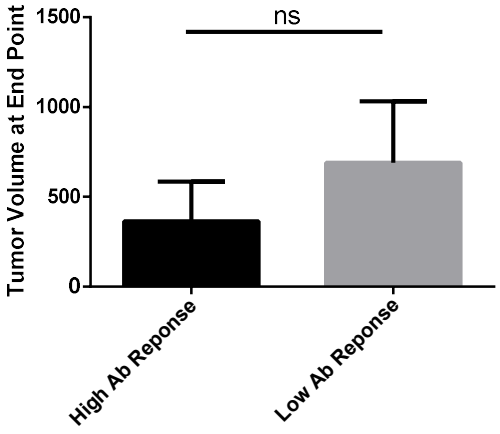


Figure 3-8. High antibody response (HAR) group had lower but not significant tumor volume than Low antibody response (LAR) group. A. Antibody response was measured in ELISA at 1:200 dilution, 4 mice in vaccine group were assigned to HAR and 6 mice were assigned to LAR, antibody response in HAR was significantly higher than LAR. B. Tumor volume of HAR and LAR group at termination point, HAR group had lower tumor volume but not statistically significant than LAR group, p-value=0.052 with one tail non-paired t test.

3.4 Discussion

In this study, we showed that frameshift mutations in mono-repeat microsatellite regions could be detected in tumor cDNA samples and the mutations at the mRNA level were much more abundant than at the DNA level. Furthermore, we showed in one example of Sec62-11A-GFP (GFP in +1 frame), the insertion frameshift protein was successfully translated and the mutation rate for this 11A was up to 25.5%. Furthermore, we selected the 9 best mouse microsatellite frameshift peptides based on MS repeat length, frameshift peptide length and MHC prediction scores to test as a vaccine. These 9 FSPs were pooled and vaccinated in BALB/c mice. This vaccine delayed tumor growth significantly in the mouse 4T1 breast cancer injection model. The tumor protection was also closely related to T cell and B cell immune responses. Our findings prove that microsatellite frameshift peptides can offer protection in the mouse tumor model, and these FSPs could be a potential antigen source in human cancer vaccine development.

Microsatellite frameshift mutations were usually used as biomarkers of microsatellite instability in MMR deficient cancer patients. Microsatellite regions were very unstable at the genomic level and MSI was found in benign lung diseases as well [35, 38]. Other reports have provided evidence as well. For example, microsatellite frameshift mutation in the 10A MS region in TGFBR2gene was detected at a lower level at MSI-Negative cancer patients [34]. And microsatellite instability events were also found in Microsatellite Stable (MSS) cancer patients at a lower level (1,500 1bp instability events vs 6,000 in MSI patient) [35].

The sequencing results of 20 dog MS candidates and 10 human MS candidates showed that MS frameshift mutations at the mRNA level could be easily detected when the MS repeat length is 10 or above, and the frameshift mutations were also detected at the genomic level when MS repeat length was even longer (repeat length is 15 in dog tumor samples and 11 in human breast cancer cell lines). Others report similar findings that long microsatellites are extremely mutagenic. For example, one report showed that the gene ACTBP2, with a microsatellite length of 15, was reported to have a mutation rate as high as $7*10^{-3}$ per locus per generation in humans [45]. Interestingly, genomic mutations of microsatellite regions in human cancer cell lines were biased to one nucleotide deletions, which was consistent with cancer genome sequencing results from other reports [34]. However, more insertions than deletions were in the cDNA sequencing results, which might be related to the characteristics of DNA polymerase and RNA polymerase. In addition, some MS candidates showed both insertion and deletion events while other candidates only had one type of mutation, which could be caused by the detection limit of the PCR experiment. Frameshift mutations of most MS candidates at the genomic level included both insertion and deletion events with different frequency (APPENDIX A Table A.1).

The expression of most microsatellite frameshift peptides is usually suppressed [72] with only a small percentage over-expressed [34]. The Sec62-11A-GFP (+1 frame) construct showed that the mutation rate was as high as 25.5% by counting the GFP+ cells. This evidence proved that the NMD system was not very efficient in cell lines at least. However, the limitation of the design for determining this number may not represent the

real mutation rate in vivo, since the whole GFP sequence was inserted after the 11A region instead of the wildtype Sec62 sequence. A small portion of frameshift proteins will transform GFP-negative cells to GFP-positive cells and 25.5% is the maximum mutation rate possible. The real mutation rate in vivo remains to be investigated.

To select the best MS FS candidates, MHC I&II predictions were used in combination with MS repeat length and FS peptide length. The MHC II prediction was included because recent cancer therapy reports showed that the MHC II could be more important [74]. Many other criteria like gene/protein function, other characteristics of FSPs were not included in our analysis and might provide important information in selecting FS antigens.

In conclusion, we demonstrated that microsatellite frameshift mutations were more prevalent at mRNA level and they could offer tumor protection in 4T1 mouse breast cancer injection model. Instead of using microsatellite frameshift mutations as biomarkers, we proved that MS frameshift peptides could also be great antigen source. Thousands of microsatellite regions are available in the human coding sequences, and this huge list of potential MS frameshift peptides may boost the cancer vaccine development.

3.5 Materials and Methods

3.5.1 cDNA and DNA samples

Dog tumor RNA snap samples were collected by Dr. Douglas Thamm at Colorado State University. Allprep DNA/RNA mini kit (Qiagen, CA) was used to extract genomic DNA and RNA from tumor samples. Tumor samples from Fibrosarcoma, Mast Cell Tumor, Osteosarcoma and Hemangiosarcoma were used. Each candidate was sequenced in 3 dog tumor samples.

Human cDNA samples and DNA samples were extracted from 4 human breast cancer cell lines: HTB26, CRL2315, CRL2326 and CRL1504. PureLink genomic DNA mini kit and PureLink RNA mini kit (Thermo Fisher Scientific, MA) were used to extract DNA and RNA. Primers for these 31 candidates are listed in the table S3-1, Phusion High-Fidelity PCR Kit (New England Biolabs, WA) was used for PCR experiments and the standard protocol in the manual was used.

3.5.2 Peptides Selection

BALB/c mouse MHC alleles H2-Kd, Dd, Ld for MHC I prediction and H2-IAd, H2-IEd for MHC II prediction were used for analysis. The MHC Consensus prediction method [59, 60] was used for MHC I and II prediction and the prediction was performed with IEDB analysis resource. Peptide sequences were predicted frameshift variant peptides from corresponding microsatellite regions. The minimum microsatellite repeat length was 10 and minimum FS peptide length was 10 amino acids. Candidates with best combined MHC I/II prediction scores were selected (sorted by rank). Peptides were synthesized by Sigma.

3.5.3 Mouse Vaccination

20 naïve Balb/c mice were ordered from Jackson Laboratory (Bar Harbor, ME). Genetic immunization was conducted with the Helios Gene Gun System (Life Science Research, CA) using published protocols [75-78]. The protocols for making gene vaccine bullets was described elsewhere [79]. The first genetic immunization was at week 7, each mouse was immunized with two types of bullets and each bullet included 100ng plasmid, 0.25μg LTA-LTB mix adjuvant and 2.5μg CpG. 4 weeks later, mice were vaccinated with the second genetic immunization with 1 μg plasmid and equal amount of LTA-LTB, CpG as in the first immunization. The control group was immunized with empty plasmid and the same dosage of adjuvants. Peptides were conjugated with KLH using Imject Maleimid-Activated mcKLH Spin Kit (Thermo Fisher Scientific, MA) using the standard protocol. In the peptide boost, mice in the vaccine group were immunized with 15 μg KLH-conjugated peptide pool and 15 μg CpG, and the control group was immunized with 15 μg KLH only and same dosage of CpG.

3.5.4 ELISA

Mouse serum was collected after peptide boost and the antibody response was measured by ELISA. Briefly, 96-well plates (Thermo Fisher Scientific, MA) were coated with 50ul 10 μg/ml peptide overnight at 4 ºC, then it was blocked with 3% BSA/PBST for 1hr at 37 ºC. After incubation with sera for 1hr at 37 ºC, the wells were probed with HRP Anti-mouse goat IgG for 1hr at 37 ºC, then the plate was developed using TMB substrate solution and terminated with 0.5N HCl. Plates were assessed using SpectraMax 190 Molecular Devices at OD 450nm.

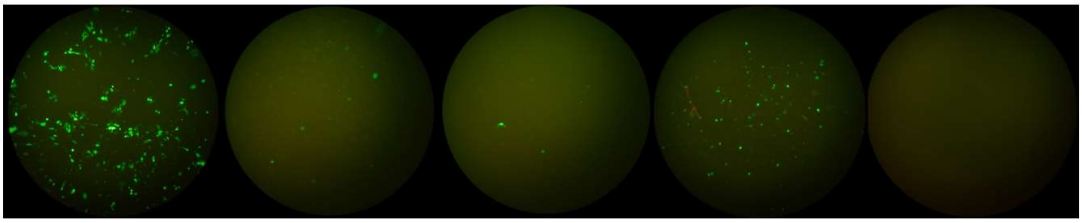
3.5.5 ELISPOT

Mouse spleen cells were collected at the termination point. Spleen cells were pooled into three groups: low-tumor volume group (5 mice with lowest tumor volume), high-tumor volume group (5 mice with highest tumor volume) and control group. ELISPOT assay was used to calculate the frequency of T cells releasing interferon γ. BD Mouse IFN-gamma ELISPOT set (BD Bioscience, CA) was used and standard protocol of the manual was applied. Splenocytes (10^6 cells per well) were incubated for 48 hours with the vaccinate frameshift peptides (20 μg/ml). Concanavalin A (2 μg/ml) was used as a positive control. After the assay was completed, the plate was scanned and positive spots were counted.

3.5.6 Cell lines transfection

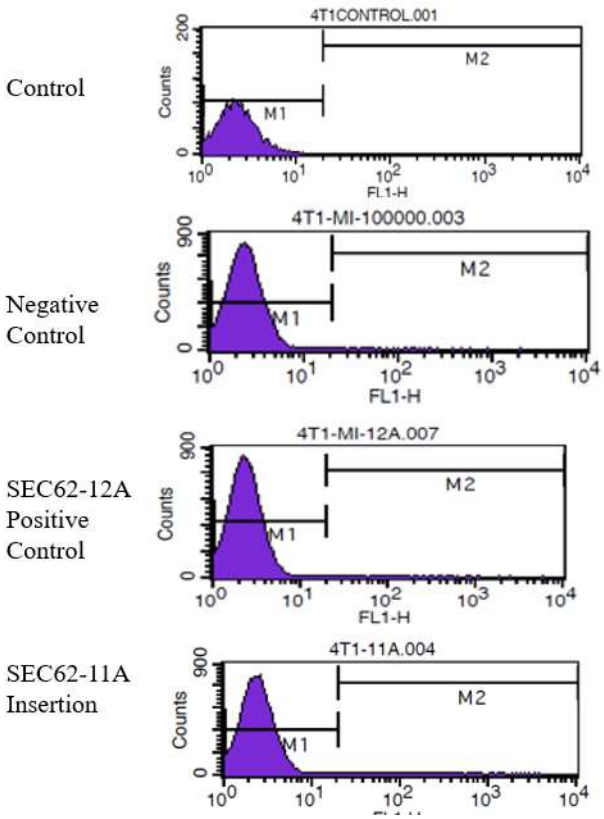
Cell lines (HEK293, 4T1, MCF7, HCC1143) were purchased from ATCC and cultured with standard protocols. Lipofectamine 2000 Transfection Reagent (Thermo Fisher Scientific, MA) was used to transfect plasmids into cell lines for overnight. Cells were then prepared in FACS buffer and quantified with flow cytometry.

Supplemental Data



GFP-Postive Control | Sec62-10A deletion mutation results in GFP expression | Sec62-11A insertion mutation results in GFP expression | Sec62-12A WT-GFP expression | Sec62-Negative Control

Figure S3-1. Frameshift mutations of Sec62-11A in 4T1 mouse breast cancer cell line.



	M1	M2
Control	96.39%	0%
Negative Control	99.91%	0.09%
SEC62-12A Positive Control	99.81%	0.19%
SEC62-11A Insertion	99.62%	0.32%

Figure S3-2 Insertion mutations were detected in Mouse 4T1 cell line.

M2 represents GFP positive population, 0.32% of cells with SEC62-11A construct were GFP positive and only 0.09% of cells in negative control were GFP positive.

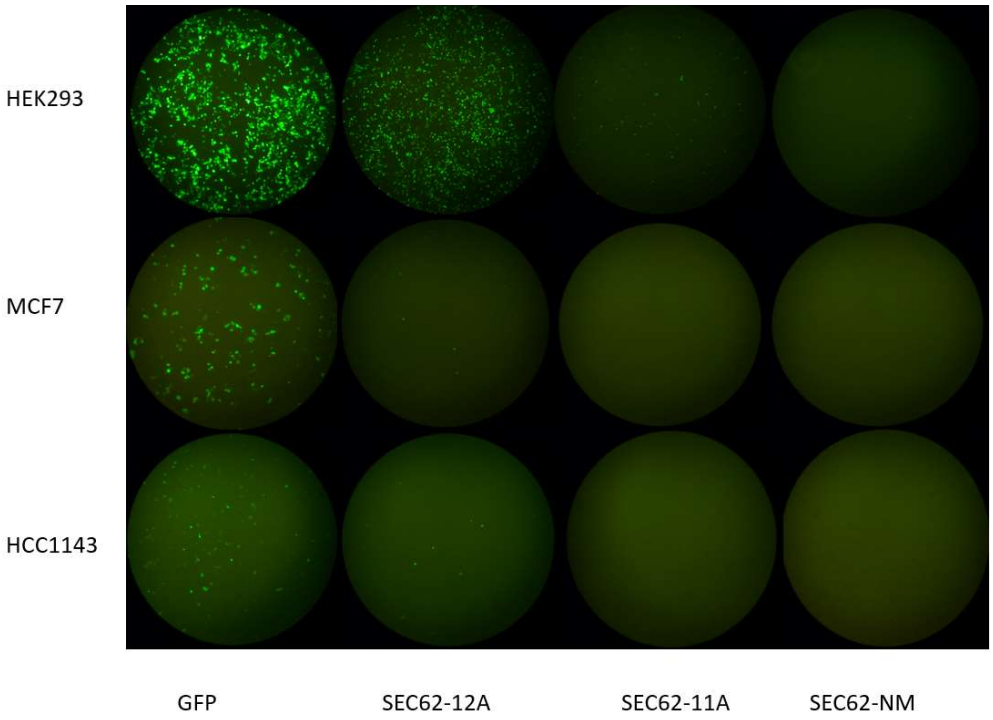
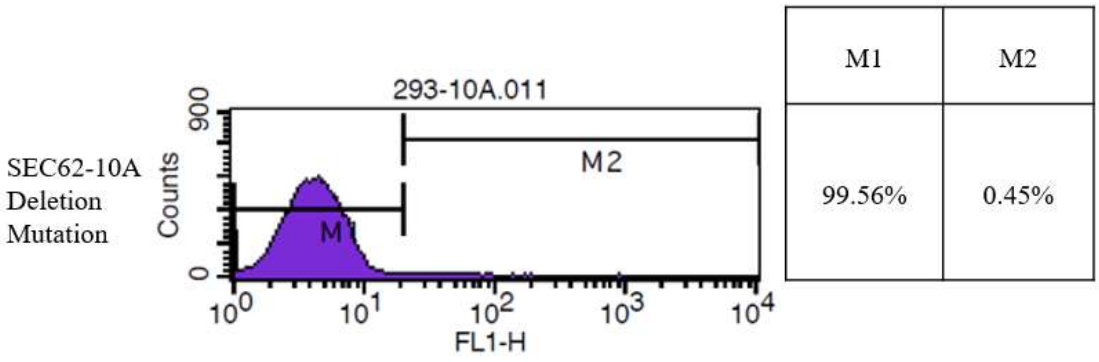


Figure S3-3 Insertion mutations were not detected in MCF7 and HCC1143 cell lines due to low transfection efficiency. There were very few GFP positive cells even in SEC62-12A positive controls from MCF7 and HCC1143 cell lines.



M1	M2
99.56%	0.45%

Figure S3-4 Deletion mutations were very rare in HEK293 cells in SEC62 11A microsatellite region. Sec62-10A construct was built for detecting 1A deletion events, only 0.45% of cells had deletion mutations and were GFP positive.

Table S3-1 PCR Primers for MS Candidates

Candidate ID	Forward Primer	Reverse Primer
345777443	CGGTGTCACTGCTCTTGG	GTAAGAGGTGAGAAGCCG
345777443	GTTACGGGCCGGTGTCAC	CCCTTTTTGCTGTGTGGC
345778567	TTGGAGAAAGGCAGGCAG	TCCGAGAACTCCGAGATC
345778567	CAGCCGCAGAGAAAGGAG	CTGCTCCGAGAACTCCGA
345780774	AGACAGGTGCCCGAGTTG	ACCACAACTCCGAGGCCT
345781087	CAGGAATTAGTGCTAGTTGC	AGACCAAATTGAGTGTGCC
345782500	AGTGAGCATCTATGTTGTG	GGGTCGAACGTGTGTCTC
345783265	GCCTCCTCTTCACACACC	GCCACACTGAGACACCAC
345789305	CAACCTCCTGGACTACAAG	TGAACACGTACACCAACTG
345790242	GATGAACAACAGTGGCCTGA	TGCACTAGTTTCAGGAGATTC
345794507	GATGTCACCCCTCTGCTCT	GGATGACCCCTTTGGACA
345794507	GACTTGCTTCGTGGTACTC	CATGTCAGCCTCGCTCTC
345796041	AGGAGAGAAGGCAGGAAG	AGCTTTTTCCTTGTTATGTG
345796516	CAGTCAGTGTGGTGGGTG	GGGAGGAAAGAGTTGGTC
345796542	AAGGGAGTCTGTGGTTGA	CAAAGAGGGAAGAGAGTGG
345796987	TCCCTCTTCCTCCACCTC	ATCTCCCCCTCTTCTTCAG
345797136	TGCTGATGGGGTATTGAAG	TGGTCGGGAAGGTTAGTG
345801690	TTGTTTTGTGATGGAGGC	TCTTATAGGATGACTTGTTGG
345802155	GCCTTCAGCCTTGTTGCG	GCCGCCACCGTCGCTGCC
345803383	TAGATCTTAGTGACTGTGTTGAC	TGGAAGTGTTTGGGATTAG
345807358	CGGAGGAGCAGTACAAAGTG	CGGTGGCTGGTCTTGTAC
359322296	AGAGCGCAGAAAGTTCCA	CGGAGGCTCTCAGAGGTA
359323976	GCGTTGCGGGCCCCCTCC	CGCTGCTCTGCTCTGCTC
ms050	CCTGAGAACACAAACATTGAT	TACCAGCACAACACCTCCA
MS053	TGGAGCAGCAGAAAAGGC	GTCACGGATGAAGAGGACTT
MS20	GAGCAAGATTCATCAGATAC	TCTATCATCATTGTTCATCAG
MS20_cDNA	TCCCATGTTCTCCTGCCC	TGGGAGAGTCGGCCTGTC
MS226dogc	ACCTTCGAGCGTCTCCTAGA	CCTTGAACTCTGGGAAGACG
MS24	TCTTGGGTCTCTTGTTCTTCAG	TTTTCTTGTCTCATGTGCTGTC
MS24_cDNA	TTGGAGGAAGAGCTTAGG	CACTAAGGAAATCAAAGGC
MS246dogc	TCGCAAAATTTCCAAAAAGG	ACAAACTGCATCCTGGGAGT
MS266dogc	GACCTGCGCGACTCCTTT	GGCGAAGGTAAGGGTGAGG
MS292	AGATGAGGAGGTCGCTGAGA	CTGGTGGGTCTCTTGAGGAT
MS305dogc	CATGCTCAACCTCCTGGACT	AAGAGCTCACCGAAGAACGA
ms355	ACTGGGGGAACTGATTCACA	CAAGTAGTCCGCTCGGGTCT
MS361	CGACAAAGTGGTGAAGAAGG	ATCAGGGCTGCAAAAACATT
MS434	CCCTACCCTGTCCCCTGT	GAAAAAGGGGTCTCGGTCTC
MS434C	GAGGCCACCCTGCTCTTG	GAGGGACCCCTTCCTCCT
MS443	CCACCCCTGAGATTCTGTCT	GGCTCACCATGACAGCATAA
MS443C	CCTTCAACCAGAGCTCCCTA	GACAGGATGGCAAGGATGAT
MS459	GATCGAATGCTGCCAGAACT	TTGGGCCTGAAGATCAGACT
MS470	AAGGACACAGTGACACAGAGG	GTGCTGATATGGTGAGGAAG
MS497	TGCGTGGAAGACAAGGAC	GACATAATTCTTAGTGGGGTAC
MS500Cdogc	GGTCCTCGGGGTCTGTCC	CTCTCGTAGGGGAGGGATG
MS500dogc	AGACCCGGCTCACTGATACA	GAGCACGGGAGAATCAGAAA
MS514dogc	GTGTGGGTCTTGAACCGAAG	TCCAGCTGCTTATTTTCTTCG
MS683	GCCACCTTCCTTGAAGACAC	TCATGACCATCTGTGCCTTC
MS687	GATTTCCACTGGGTGAATGG	AGGTGGGAGGGAAAGTGTTC
MS75	TGCCATACCTGTTTTTCCC	AGTTATCTCAGGTAGGTGTTGC
MS75_cDNA	AAAGGAAAAGCTGAAAGTGGAA	GCAACAGCAAGGAGAAGAATAC
MS773	CCAAATAGGAGCCTGCTGTC	TCTCCTTCAGAGCCTCGTTC
MS806	GGAAGGTGGCAGAAACAGAG	TCTTCATGCTGGGGATAAGG
MS806C	GGCTTCACAATTCAGGAAGG	AATACAAGCTCATCCGGGACT
MS866	GCGCCTACAGTGCCTACTCT	AGAGCCTGATAGGGTGTTGC
MS866C	TCCAGCTGGTGGTGTCTGTA	AGGAGGATGACTGCACCAAA
MS87	GTGCGCCTACAGTGCCTAC	CCTCAAACACTGCATGAAGC
MS87_cDNA	TCAAGTTGGCTGCTAGAGTC	GTTGGCCAGGAGGATGAC
MS880	GGGCCATCCAGTATATTCCA	CGTACAATACACCTGTGGGC
MS93	TGCACAGCACAGAGGTCCT	CAGCAAACAGGAATGGGC
MS93_cDNA	CCAACTGGTGCTGCAAGG	TGGGAGAGTCGGCCTGTC
MS990	CAAATCAGGCCTCACTGTCA	TCTGACGACTCAAGCTGTGG
MS990C	TGCTCATCAGCACTGTAGGC	GCTGCCCACAAAACAATGTC
MS991	GGGGAAGAAGGATGAGGAAC	TTAATTCTCCCACGGCAAAG
MS991C	GTTGGGGACACTTGCTCAGT	TCCGTGTCTGTTTCCTCCTT

CHAPTER 4

Using Frameshift Peptide Arrays for Cancer Neo-Antigens Screening

4.1 Abstract

Neo-antigens are important for an anti-tumor immune response based on recent clinical studies of immune checkpoint inhibitors (ICI). Besides point mutations at the DNA level, we reported recently that frameshift peptides (FSPs) from insertions or deletions (INDELs) of microsatellite regions (MSs) and exon mis-splicing FS mutations were detected in the mRNA but not in the corresponding DNA (in submission). These FSPs offered protection in mouse tumor models. FSPs shown to be better neo antigens than point mutations in the sense of high immunogenicity and low probability of causing autoimmunity. However, there was not a general method available to select the best FSPs from thousands of potential candidates. Here we provided a platform of using frameshift peptide arrays for cancer neo-antigen screening which quantifies the anti-FSP humoral immune response in cancer patients. The results of screening 9 types of dog cancer serum indicated that cancer serum samples had significantly higher antibody responses against FSPs than normal samples. Common reactive FSPs and cancer specific immune responses were both detected. In addition, non-reactive FSPs and the corresponding reactive FSPs were selected from the screening. The mouse homolog FSP sequences were used for testing in mouse tumor models. Non-reactive FSPs did not offer protection in mouse melanoma and breast cancer models while reactive FSP offered protection in both models. The tumor protection was linearly correlated to Ab response to the FSP.

These data provide a rationale for using frameshift peptide array as the platform for cancer neo-antigen screening.

4.2 Introduction

Neoantigens have been shown to be the primary targets of effective, anti-tumor T cells during ICI therapy [128,129]. The combination of neo-antigen vaccination with immunotherapy has been hypothesized to boost the anti-tumor immune response and further improve the treatment benefit and the number of patients responding to treatment over ICI alone. Research efforts were directed to searching for genomic mutations and massive genome sequencing technology was used for finding personalized point mutations [130]. Two initial studies of personalized melanoma cancer vaccines showed that this approach induced strong immune responses and reduced recurrence significantly [131,132]. Meanwhile, we have proposed an alternative approach for neo-antigen based vaccine development. Frameshift mutations based cancer neo-antigens can be better than point mutations. Frameshift mutations from microsatellite regions and mis-splicing events [80] are much more abundant at the mRNA level as shown in previous chapters and they are easily missed in genome sequencing. The feasibility of using FSPs for cancer vaccine development was validated using the mouse 4T1 breast cancer model. However, with thousands of frameshift candidates available, a general method for frameshift antigen screening is necessary. Here we were test whether a frameshift peptide array platform is effective or not in neo-antigen screening.

Frameshift mutations are usually generated from transcription through microsatellites and mis-splicing of RNA INDELs. High levels of microsatellite frameshift mutations at the

DNA level are observed in mismatch repair deficient cancer patients (MSI), but frameshift mutations at mRNA level are rarely investigated. Frameshift mutations are much more abundant at the mRNA level due to the 10,000-fold higher error rate from in RNA compared to DNA. Chapter 3 showed evidence that FSPs from microsatellite regions and mis-splicing events [80] worked independently and they offered tumor protection in injectable and transgenic mouse tumor models.

Here we tested the idea of using a dog frameshift peptide array for neo-antigen screening. Mono-repeat microsatellite regions and INDELs were predicted from the dog coding DNA sequences, while mis-splicing INDELs were found from human tumor EST libraries [81] and dog homolog sequences were used in the FS array. We constructed a frameshift peptide array with 830 20-mer peptides representing 377 frameshift antigens. 116 dog cancer serum samples from 9 major dog cancer types and 52 dog healthy serum samples were screened for positive cancer neo-antigens. Common reactive FSPs and non-reactive FSPs were then tested in the mouse melanoma and breast cancer models.

4.3 Results

4.3.1 Dog Frameshift Peptide Array Platform

830 peptides from 377 predicted frameshift antigens were synthesized commercially and printed on NSB9 amine surface slides. The space between each peptide on the NSB9 surface is 3nm and the NSB9 slide is specifically designed for detecting high affinity antibody-peptide binding, as opposed to the immunosignature arrays [133]. Each frameshift antigen was represented with 1~4 17-mer non-overlapping frameshift peptides, depending on the frameshift peptide length, with a three-amino acid linker (GSC) at C-

terminus. 328 of these 377 antigens were predicted INDEL frameshift peptides from mono-repeat microsatellite regions (minimum repeat length was 7). The rest of the 49 frameshift antigens were from mis-splicing events. Mis-splicing frameshift antigens were discovered in human tumor EST libraries and the dog homolog peptide sequences were then used for the frameshift peptide array.

Each NSB9 amine slide included 24 peptide arrays. The principle of this frameshift peptide array was like a conventional ELISA but with much higher sensitivity. Cancer patient sera was incubated with the frameshift peptide array overnight and fluorescence labeled secondary antibody was then applied. The slide was scanned afterwards and the anti-frameshift peptide response was transformed to fluorescence intensity (Fig4-1).

To test this platform, we screened dog cancer serum samples and normal samples on this array.

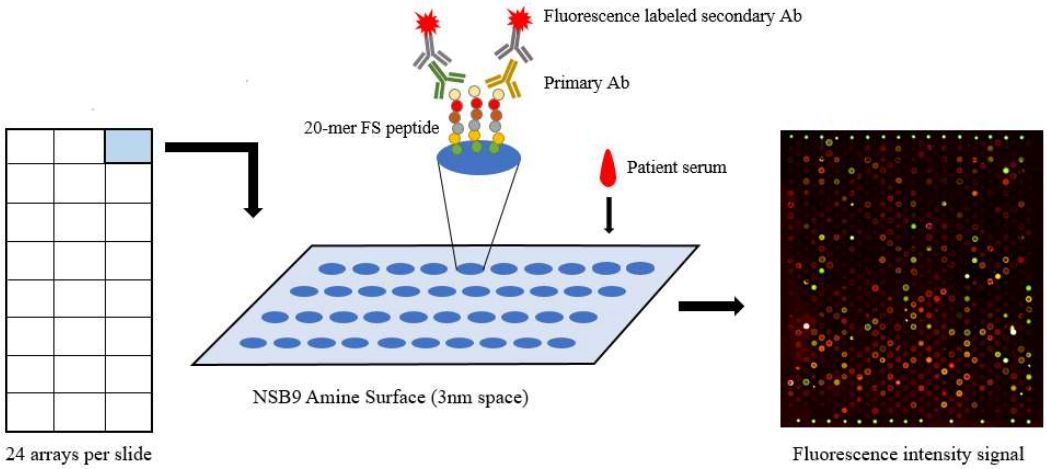


Figure 4-1. The process of measuring anti-FS immune response on frameshift peptide array. Cancer patient serum was incubated overnight with frameshift peptide

array, fluorescence labeled secondary antibody was then used for transforming antibody-FSP binding signal to fluorescence, each slide contains 24 arrays.

4.3.2 Screening dog cancer serum samples with frameshift peptide arrays

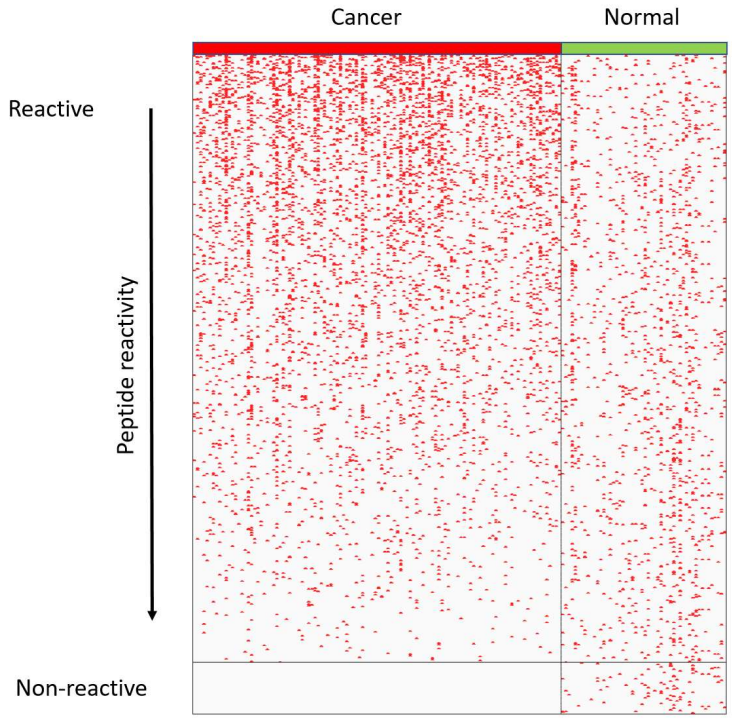
9 types dog cancer serum with a total of 116 samples were screened with the dog frameshift peptide array. These 9 types of cancer included carcinoma, fibrosarcoma, hemangiosarcoma, lymphoma, mast cell tumor, osteosarcoma, histiocytic sarcoma, synovial cell sarcoma and malignant histiocytosis. 52 healthy dog serum samples were used as control. The sample description is summarized in table S4-1.

Fluorescence intensity for each array (sample) was normalized to median florescence of the array. A cutoff value (Cutoff = Average (Normal) + 2*STD (Normal)) was set for each peptide. Samples with normalized values which were larger than the cutoff were counted as positive for this peptide and samples with lower value were negative. The distribution of positive samples for each peptide was shown in Figure 4-2. There were fewer positive samples in the normal group generally and these samples were distributed relatively evenly. Overall the cancer samples had a significantly higher positive rate than normal samples (p-value <0.0001). About half of the peptides had a similar positive rate between cancer group and normal group, while there was small group of FSPs (122 FSPs) which were highly positive in the cancer group and they had over 10% positive rate in the cancer group. A small group of highly positive rate peptides is listed in Table S4-2. Most of the listed FSPs were from the MS regions with positive rate as high as 15%~19%. Interestingly, 65 FSPs had no reactivity in the cancer group at all, while these non-reactive FSPs had the same reactivity in normal group as other FSPs. These non-

reactive peptides and common reactive peptides were then tested with mouse tumor models.

Since 9 types of cancer serum were tested with the array, one interesting question would be whether cancer type-specific and common anti-FS reactivity could be observed.

A



B

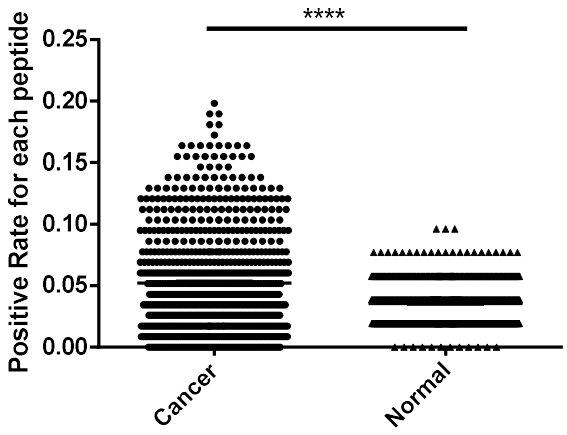


Figure 4-2 Dog cancer serum samples had higher positive rate than normal samples. A. Overall distribution of FSP positive samples in cancer group and normal group. Reactivity was sorted by descending order of cancer group, each red dot represented one positive sample for corresponding FS peptide; B. Positive rate distribution of each FS peptide was compared between cancer group and normal group of all FSPs, p-value<0.0001.

4.3.3 Anti-FS reactivity in different cancer types

Out of the total 9 cancer types, 6 cancer types had 10 or more samples in the experiment (Table S4-1) and these 6 cancer types were used for the cancer type analysis. The positive rate distribution of all the FSPs was compared between different cancer types.

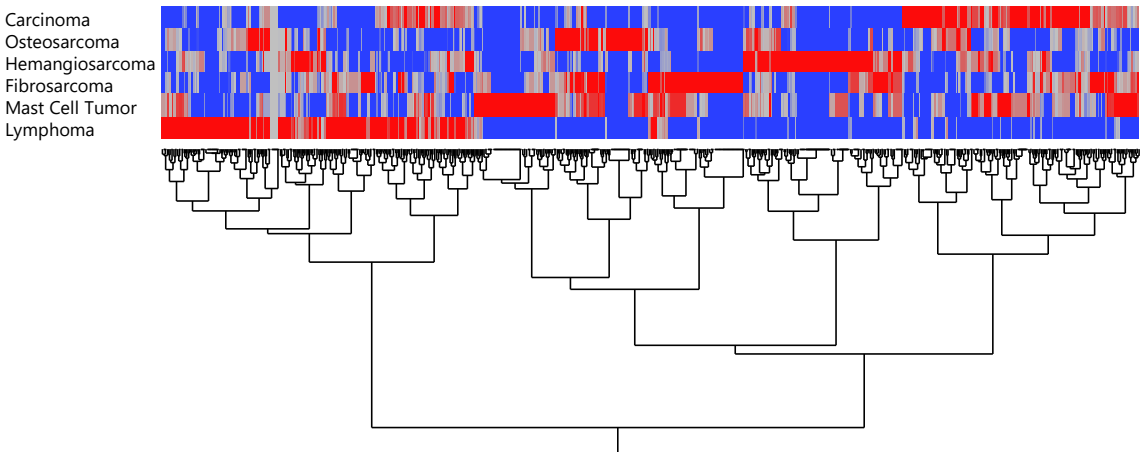
Frameshift mutations in both MS and mis-splicing are depend on the level of RNA production. Therefore, the abundance of FSPs was hypothesized to be related to the gene expression level. Therefore, we would assume that the anti-FS immune response of each cancer type was related to the expression profile of the origin of the cancer cells and that correlations between different cancer types should be relatively low.

It turned out that the positive rate correlations between cancer groups were relatively low (~0.3), which was consistent with our assumptions. However, common reactivity across 3~4 cancer types was not rare (Fig4-3A), which indicated a broadly effective cancer vaccine would be feasible if we had included enough FSPs to cover different cancer types.

In addition, lymphoma was clustered as a unique group and it had a specific anti-FSP profile which was totally different from the other 5 cancer types. The correlations of positive rate profile between lymphoma and other cancer types were the lowest and they range from 0.15~0.26 (Fig4-3B), which further validated our assumptions since the gene expression profile of lymphoma was most different from other 5 cancer types.

Next step we would like to test whether common reactive FSPs and non-reactive FSPs could offer protection in mouse tumor models.

A



B

	Carcinoma	*Fibrosarcoma*	*Hemangiosarcoma*	*Lymphoma*	*Mast Cell Tumor*	*Osteosarcoma*
Carcinoma	1.0000					
Fibrosarcoma	0.2915	1.0000				
Hemangiosarcoma	0.3352	0.2220	1.0000			
Lymphoma	**0.2664**	**0.2206**	**0.2603**	1.0000		
Mast Cell Tumor	0.3982	0.4694	0.3298	**0.1528**	1.0000	
Osteosarcoma	0.3064	0.3337	0.3060	**0.1912**	0.3607	1.0000

Figure 4-3. Positive rate distribution across different cancer types. A. Two-way hierarchical clustering of positive rate of each FS peptide by each cancer type. Lymphoma was clustered to a unique group and the rest of the five cancer types as one group. B. Correlation matrix of positive rates by each cancer group (carcinoma).

4.3.4 Only Reactive FSPs provided protection in mouse tumor models

As mentioned previously, there was a group of 65 non-reactive FSPs which had 0% positive rate in the cancer group, while they had the same reactivity in the non-cancer group as other FSPs. At the same time, there were more than 400 FSPs which had more reactivity in cancer group than non-cancer group. We hypothesized that the reactive peptides in the cancer group should be good vaccine antigens. There were two reasons behind this hypothesis: 1. Recent ICI clinical trials showed that ICIs were very effective for late stage microsatellite instability high (MSI-H) patients [66] and the FDA granted accelerated approval for using MSI-H as a biomarker for ICI treatments. ICI treatment boosted the anti-MS FSPs immune response for reactive FSPs but not for non-reactive FSPs in these patients, and the boosted anti-reactive FSPs immune response could kill tumor cells effectively; 2. The immunoediting process suppressed the anti-tumor immune response and was constantly removing highly immunogenic tumor cells. The fact that a high anti-FSPs immune response could be detected at late stage indicated that it was

difficult for tumor cells to get rid of these FSPs. However, it was not clear to us what the antigenic potential of the non-reactive FSPs would be.

To test this hypothesis concerning the reactive FSPs and gain information on the non-reactive FSPs, we tested these reactive and non-reactive FSPs in the mouse tumor models. Within these 65 non-reactive FSPs, we could find 5 highly conserved mouse homolog peptides, 2 mouse MS FSPs which had the same MS regions at the same location of the homolog genes, 3 mis-splicing FSPs included 2 peptides resulting from mis-splicing events in gene-fusions and 1 from mis-splicing events within the same gene.

Most frameshift antigens included in the FS array had more than 34 amino acids and they were presented with 2 or more non-overlapping 17-mer peptides with a GSC linker. The antibody response for different FSPs from the same FS antigen usually had different positive rates in the cancer group. To make a direct comparison of reactive peptides and non-reactive peptides, we searched for reactive FSPs from the same 5 frameshift antigens which had one peptide in the non-reactive section. We could find one reactive peptide (RAGNFVTVEIQSLVPKK) from the C1S gene which had 7.7% positive rate in cancer group, the other FSP (SLPILFGSLRKQYMYSK) from the same antigen had no reactivity in cancer group at all (Fig4-4A & Table 4-1).

Table 4-1. List of non-reactive peptides and reactive peptide

Gene Name	Type	Cancer PR	Type of FS	Dog (Array) Sequence	Mouse homolog sequence
PCBP2	Non-reactive	0%	MS	HDAWQHRIQCRFGCICSGSC	HDPWQHRIQCRFGCICS
C1S	Non-reactive	0%	MS	SLPILFGSLRKQYMYSKGSC	ALPILIGSPIRQNTCLKMW
FANCI	Non-reactive	0%	Mis-splicing	VSPGVSELRRNSKKYGKGSC	LSPGMSELQRNSKHCGK
CCDC13_HHATL	Non-reactive	0%	Mis-splicing	PLVPAAAAWSLCGPLCGGSC	PMVPAAAAWPLYGPLCG
DDIT3_MARS	Non-reactive	0%	Mis-splicing	LPLGVSRGFPSAKASCFGSC	LPLGVSMGFPSAKANCF
C1S	Reactive	7.76%	MS	RAGNFVTVEIQSLVPKKGSC	RVGSFVTMEIPSPVLKK

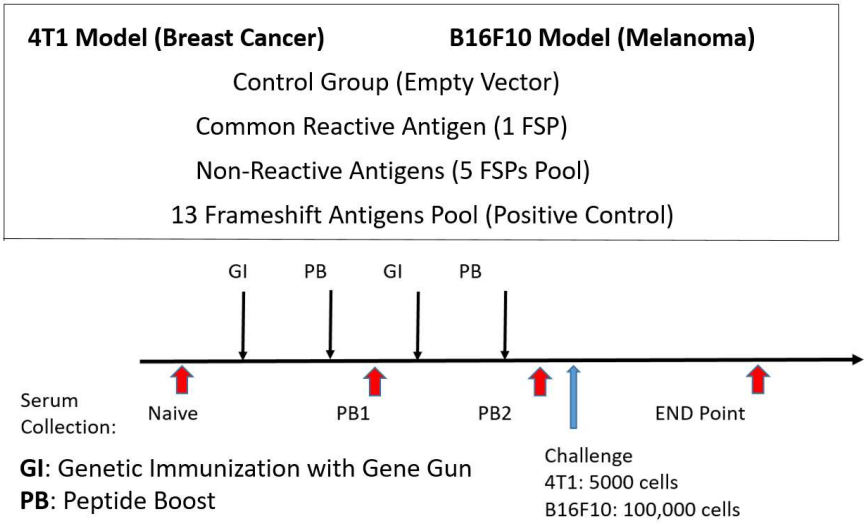
The mouse melanoma injectable model (B16F10 cell line) and breast cancer injectable model (4T1 cell line) were used for testing one reactive peptide and the 5 non-reactive peptide pool. A pool of 13 previously tested frameshift antigens was used as a positive control (Table S4-3). All 4 groups received two rounds of genetic immunization with the gene gun on the ear and a peptide boost via sub-cutaneous injection. The reactive peptide vaccine group had significantly slower tumor growth in the melanoma model than the non-reactive FSPs pool group and control group (p-value <0.01). The same results were found in the 4T1 breast cancer model - the reactive FSPs offered tumor protection while non-reactive FSPs pool did not. The tumor volume between the non-reactive FSPs pool group and the control group was not significantly different. The 13-frameshift antigen pool showed a similar level of protection as the reactive FSP group. These data indicated that reactive FSPs are better cancer neo-antigens in these two completely different mouse cancer models, and it is feasible to select cancer FS antigens for multiple cancer types based on the screening results of frameshift peptide array.

Since these FSPs were first selected based on antibody response against FSPs in dog cancer serum samples, one interesting question would be whether we could see

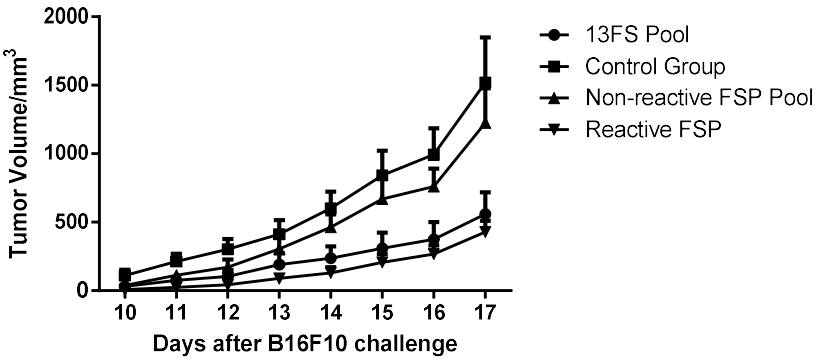
correlation between the antibody response level against the reactive FSP and tumor protection level.

A

B



C



D

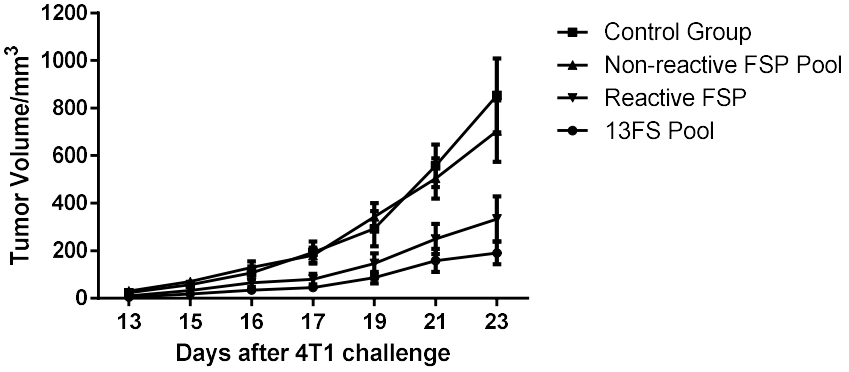


Figure 4-4. Reactive FSPs showed protection in the mouse melanoma and breast cancer models. A. Reactive and non-reactive FSPs from the same microsatellite frameshift antigen of C1S. B. 4 groups of mice were used: the control group was immunized with control plasmid, 13FS Pool was used as positive control, non-reactive FSPs pool and reactive FSP. All 4 groups received 2 rounds of genetic immunization with gene gun and peptide boost via subcutaneous injection. The B16F10 cell line was used for the melanoma model and the 4T1 cell line was used for breast cancer model. Each group had 10 mice. C. Reactive FSPs slowed tumor growth significantly compared to the non-reactive FSPs pool and the control group in the melanoma model (p-value <0.01). D. The Reactive FSPs offered tumor protection in mouse breast cancer model as well. The tumor volume was significantly lower than control group and non-reactive FSPs pool group (p-value<0.05).

4.3.5 Tumor protection was linearly correlated to antibody response

Reactive FSPs and non-reactive FSPs were selected based on the antibody response against the FSPs in dog cancer patients. We hypothesized that these antibody responses had anti-tumor activity or it was indirectly related to other anti-tumor mechanisms. To test the idea, the antibody response against vaccinated FSPs was measured in the mouse breast cancer model and correlated to tumor volume at the endpoint of the experiment.

Antibody responses were measured after PB2 (before tumor challenge) and at end point (after tumor challenge), both time points were included in the analysis since both timepoints could have impact on tumor protection. R-COMP peptide was encoded in a plasmid as an internal positive control for genetic immunization. Antibody response against r-COMP was not correlated to tumor volume in any of the 13FS pool group, non-reactive FSPs pool group or reactive FSP group, which indicated that anti-r-COMP immune response was not essential for tumor protection. However, the antibody response against 13FS pool and reactive FSP was significantly linearly correlated to tumor volume at the end point (p-value=0.03 for 13FS Pool, 0.0024 for Reactive FSP) , while antibody response against non-reactive FSPs pool did not correlate to tumor volume (p-value=0.18). These data confirmed that tumor protection was linearly correlated to antibody response against selected reactive FSPs but not non-reactive FSPs, and the antibody response against reactive FSPs had tumor-killing activity or it was indirectly related to other tumor-killing activity mechanisms.

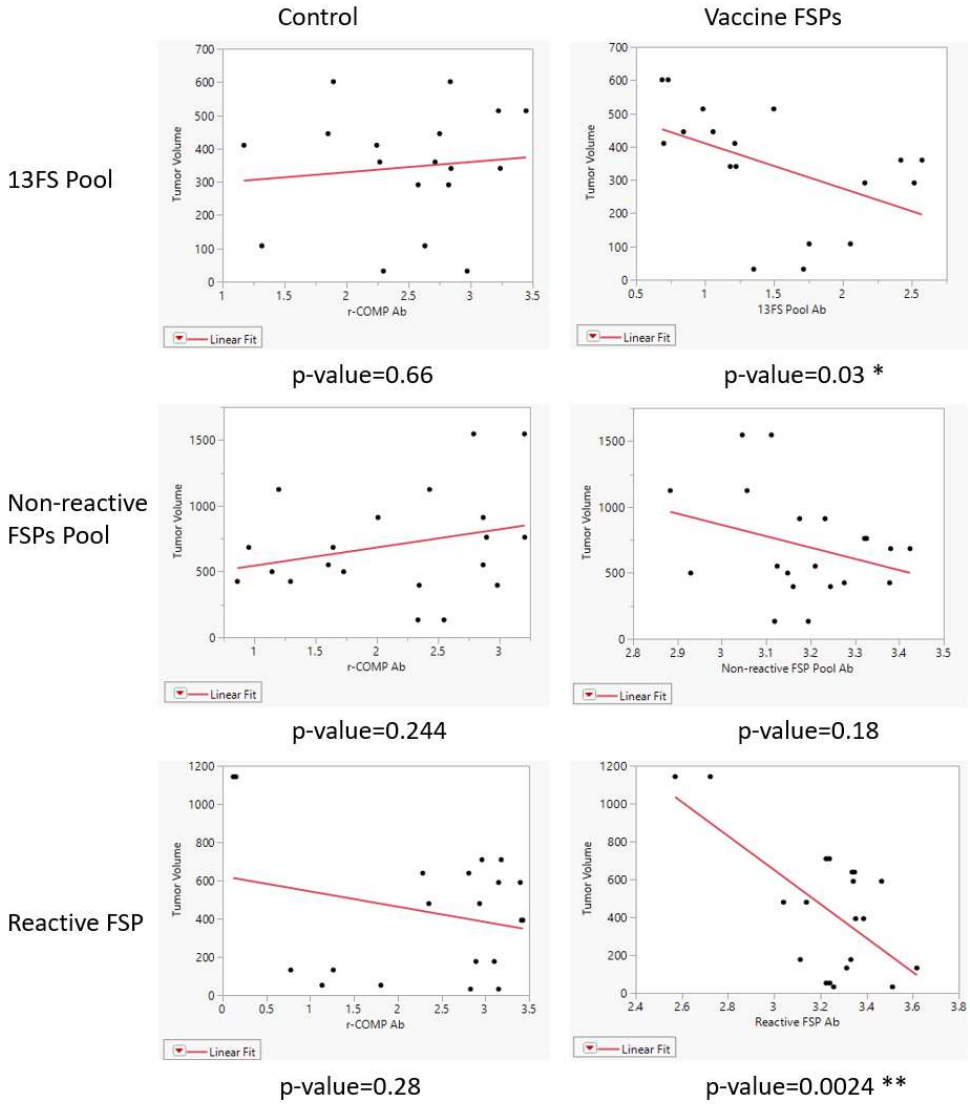


Figure 4-5. Tumor volume was linearly correlated to antibody response against vaccinated reactive FSPs in the 4T1 breast cancer model. Tumor volume at the end point was used for analysis. The antibody response was measured by ELISA after PB2 and at the end point. r-COMP peptide was encoded in a plamsid and used as an internal control for genetic immunization. The p-value was calculated for linear fit of antibody response against tumor volume.

4.4 Discussion

We first provided a platform for screening potential cancer frameshift antigens. Over 300 frameshift antigens were represented in 830 FSPs on the frameshift peptide array and this platform could be easily expanded with more antigens and peptides. We had screened 116 dog cancer serum samples representing 9 cancer types and 52 normal serum samples. Cancer-type specific and common anti-FSP immune responses were observed. Lymphoma was clustered as a unique group from the other cancer types. Then we selected non-reactive FSPs and corresponding reactive FSP based on the array data. We further validated that the selected reactive FSP could offer tumor protection in the mouse melanoma and breast cancer models while the non-reactive FSPs could not. We further showed that this tumor protection was linearly correlated to the antibody response against the immunized reactive FSPs but not the non-reactive FSPs. All these data suggest that using a frameshift peptide array for cancer FS antigen screening is feasible and efficient.

With ~400K potential frameshift antigens at the mRNA level in cancer patients, we can only select a small proportion for developing cancer vaccines, usually only 20 candidates were included in a single vaccine. Using the frameshift peptide array to screen the B cell immune response in cancer patients could be a solution for selecting FS antigens. A potential criticism about using positive frameshift antigens discovered from late stage cancer patients is based on the logic that the antibody response did not stop the tumor progression in the first place. However, the success of using immune checkpoint inhibitors in late stage MSI-H patients confirms that tumor can be killed by activating late stage anti-FS immune responses. In the meantime, unlimited rounds of an immunoediting

process could mean the surviving positive frameshift antigens are difficult for tumor to get rid of.

Another interesting phenomenon was the group of 65 non-reactive FSPs. These non-reactive FSPs are not by chance based on statistical analysis, and most of these non-reactive FSPs exist in the frameshift antigens which included a reactive FSP counterpart as well (Fig4-4A). We hypothesized that the immune reaction to these frameshift antigens was not unique to cancer. It may also occur during acute virus or bacterial infections due to disrupted transcriptional and translational systems, but the presentation of these frameshift antigens was only transient in these infectious diseases and induced non-mature antibody response against these antigens which target the non-reactive FSP part of the frameshift antigen. However, when one patient developed a tumor, the immune system was constantly trained with these frameshift antigens and the antibody response matured towards targeting the reactive FSP in the same antigen. This maturing antibody response had anti-tumor activity, and it could be detected in late stage cancer patient from the frameshift peptide array. The non-reactive FSP could not induce an effective anti-tumor immune response based on this hypothesis, but the reactive FSPs in the same antigen could, which was consistent with our findings (Figure S4-1). This hypothesis remains to be validated.

In conclusion, these data demonstrated that the frameshift peptide array platform could be used for cancer neo-antigen screening and reactive FSPs selected from the screening can protect mice from tumor challenge while non-reactive FSPs could not. Besides the usage of this platform for broadly effective cancer antigen screening, the platform was also

expandable and it could be used widely in discovering cancer subtype specific neo-antigens and personalized frameshift antigens in the future. The array may also have use in diagnosis of cancer.

4.5 Materials and Methods

4.5.1 Frameshift Peptide Array

830 frameshift peptides were non-overlapping 17-mer peptides with GSC linker at the C-terminus representing 377 frameshift antigens from mis-splicing events and indels in microsatellite regions. All the peptides were synthesized by ChinaPeptides Co. (Shanghai, China). The peptide array was printed as previously described by Applied Microarrays, Inc (Tempe, AZ) on NSB-9 amine slides from NSB Postech (Seoul, Korea). Serum samples were 1:200 diluted with 3% BSA in PBST and 200 µl diluted serum was incubated with the peptide array at R.T overnight. After washing with PBST for 3 times, the slide was incubated with secondary antibody for 1 hour at R.T. The slide was washed and then scanned with the Agilent C Scanner (Agilent Technologies, CA). Fluorescence intensity was analyzed with GenePix Pro 6.0 (Axon Instruments, CA). Each array was normalized to the median. The positive rate was calculated by setting a cutoff as average of normal the samples plus two standard deviation of this average.

4.5.2 Dog Serum Samples

116 dog cancer serum samples and 52 non-cancer dog serum samples were collected by CSU and detailed cancer type information is in Table S4-1.

4.5.3 Genetic Immunization

The DNA fragments encoding the FS peptides were cloned as a C-terminal fusion into the genetic immunization vectors pCMVi-UB and pCMVi-LSrCOMPTT with the Bgl II and Hind III sites and mixed in a 1:1 ratio as the genetic vaccine plasmids. Three adjuvants were encoded by genetic immunization vectors. LTAB is an abbreviation of immunization with a 1:5 ratio by weight of two plasmids, pCMVi-LTA and pCMVi-LTB, expressing the heat-labile enterotoxins LTA and LTB from Escherichia coli. Vectors pCMVi-UB, pCMVi-LSrCOMPTT, pCMVi-LTA (also called pCMVi-LS-LTA-R192G) and pCMVi-LTB were available from the PSI:Biology-Materials Repository DNASU (dnasu.org) at Arizona State University. Additional adjuvants were the poly I:C single strand nucleotides from Sigma. The first genetic immunization was at week 7. Each mouse was immunized with two sets of bullets and each set included 1 μg plasmid, 0.25μg LTA-LTB mix adjuvant and 2.5μg CpG. 4 weeks later, mice were vaccinated with the second genetic immunization with the same dosage. The control group was immunized with empty plasmid and the same dosage of adjuvants. Peptides were conjugated with KLH using Imject Maleimid-Activated mcKLH Spin Kit (Thermo Fisher Scientific, MA) using a standard protocol. In the peptide boost, mice in the vaccine group were immunized with 15 μg KLH-conjugated peptide pool and 15 μg CpG, and the control group was immunized with 15 μg KLH only and the same dosage of CpG.

4.5.4 Gene Gun and Bullets

The gene gun protocol and bullets for genetic vaccination was described in Chapter 3.

4.5.5 Mouse tumor models

4T1 and B16F10 cell lines were purchased from ATCC and cultured with ATCC recommended protocols. Naïve C57BL/6 and BALB/c mice were purchased from Jackson Laboratory. 100,000 B16F10 cells were injected into C57BL/6 mice 2 weeks after the final peptide boost, and 5,000 4T1 cells were injected into BALB/c mice 2 weeks after final PB as well. Each group had 10 mice.

4.5.6 ELISA

Mouse serum was collected after peptide boost and the antibody response was measured by ELISA. Briefly, serum samples were diluted 1:200, 96-well plates (Thermo Fisher Scientific, MA) were coated with 50ul 10 μg/ml peptide overnight at 4 ºC, then it was blocked with 3% BSA/PBST for 1hr at 37 ºC. After incubation with sera for 1hr at 37 ºC, wells were probed with HRP Anti-mouse goat IgG for 1hr at 37 ºC, then the plate was developed using TMB substrate solution and terminated with 0.5N HCl. Plates were assessed using SpectraMax 190 Molecular Devices at OD 450nm.

Supplemental Data

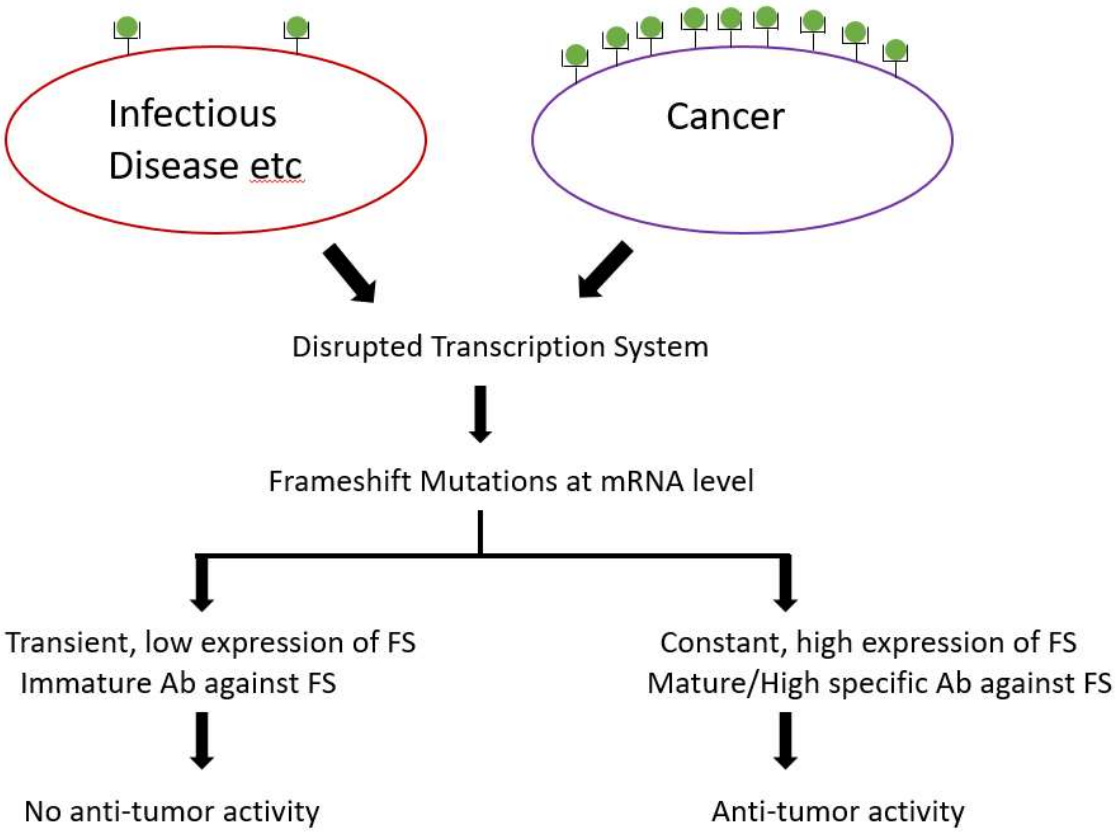


Figure S4-1 A simple model to explain why non-reactive FS cannot offer protection in mouse tumor models.

Table S4-1 Description of Dog tumor samples

Cancer Type	Number	Collection Site
Mast Cell Tumor	22	CSU
Fibrosarcoma	20	CSU
Osteosarcoma	20	CSU
Hemangiosarcoma	17	CSU
Carcinoma	13	CSU
Lymphoma	10	CSU
Sarcoma	8	CSU
Malignant histiocytosis	4	CSU
Synovial cell sarcoma	2	CSU

Table S4-2 High positive rate peptides in the cancer group

Gene	Type	Peptide	Cancer PR	Normal PR
PLEKHM2	MS	GSSAVAVGGPTPRIGPTGSC	19.83%	5.77%
DPH2	Mis-splicing	LPCSSLTSYWEMLWLWLGSC	18.97%	1.92%
HGS	MS	KKSTTRIPMWPCTPWRSGSC	18.97%	3.85%
BAX	MS	HPSWPWTRCLRMRPPRSGSC	18.10%	5.77%
SEC62	MS	KFFWMEMRYLCGSMTQFGSC	18.10%	1.92%
SEC62	MS	QKKRKKKKDKVETQNRSGSC	17.24%	1.92%
OTUB2	MS	QSPASGPGSLSLGWNGSGSC	16.38%	3.85%
CHST2	MS	LLPPSGRPPPLQPPPPPGSC	16.38%	1.92%
WNT2B	MS	PLLPHSPRSSRTPLLPHGSC	16.38%	5.77%
ZNF282	MS	RRSATRSRRCAGKWPREGSC	16.38%	1.92%
DCLK2	MS	PVPWQAPEAGPPEAGLPGSC	16.38%	1.92%
FAM83H	MS	RPSRSAGPARCPPCPSAGSC	16.38%	5.77%
POU3F2	MS	RLARSPCAPEAPGEPRAGSC	16.38%	5.77%
MSH3	MS	LALWECSLPQARLCLIVGSC	15.52%	1.92%
EEF1A	Mis-splicing	GGHLCSSQRYNGSKICRGSC	15.52%	1.92%
KRT8	Mis-splicing	SSMRSCRAWLGSTGMTCGSC	15.52%	1.92%

Table S4-3 Components of 13 Frameshift Antigen Pool

Gene Name	TYPE	mRNA Accession	MS Type	Frameshift Peptide
GM5464	MS	NM_001034881.3	14_T	VTAGYQEEEMEASACGAKGPGLAPWPPSWLALQDSLLCVVVALADLRRKSCC
SLC35F5	MS	NM_028787.4	10_T	LLCVVFGKFVIPRSTFRHTGCHSEYFVFNFWTFYFNPCSCISE
FAM71A	MS	NM_001109759.1	10_A	RQKKIRPPKKKRSIQGQRQKPPRDHRCECDQLFCFFWWGGNP
CCDC112	MS	NM_001160399.1	11_A	RVYSKLENQKAAKEGGNTQVKRKGGHRASAFSKQSRR
NEMF	MS	NM_025441.3	11_A	KAKEQAAAEAAEEQAAACRCGSQPVSLCQCQKIL
ANKRD45	MS	NM_028664.1	18_T	FGVNGARRNSRIGEFRKVTIFLTARV
RFC3	MS	NM_027009.2	10_A	LKSAPLQATTTLKLIPVMRGTATEL
CALR4	MS	NM_001033226.4	13_T	FFSLSLSFLHRWMDKTVGTI
MS255	MS			YFSCDKRCIKHYAGNKSLLTFSGY
MS927	MS			ICMSPPLLWATLQAPETTSAACKASYRPEGLYL
RBM	Mis-splicing			GRVIECDVVKGSCQDGEAVHWKSAPGGHRAGDPLTLRAVREGAGM
SLAIN2 (1-78)	Mis-splicing			IPRMQPQASANHCQLLKVMVA
SMC1A	Mis-splicing			TAIIGPNGSGCSGVYCHEEPQGEDSSV

CHAPTER 5

Safety Evaluation of a Frameshift Antigen Based Prophylactic Cancer Vaccine in Dog

5.1 Summary

5.1.1 Background

Neoantigens are important for immune checkpoint inhibitor treatment and many ongoing clinical trials are testing combinational therapy of ICI and point mutation based on a personalized neoantigen cancer vaccine [14, 15]. The genetic vaccine (Oncept) which was designed to against tyrosinase was approved to treat dog melanoma [132] [16]. However, here we propose that a broadly effective frameshift neoantigen based prophylactic cancer vaccine could be an alternative path for cancer therapy. This safety trial was conducted to evaluate the safety and immune response for a prophylactic cancer vaccine composed of 21 frameshift antigens.

5.1.2 Methods

Three healthy dogs were enrolled and they received two rounds of genetic immunization (300 μg antigen plasmid and 300 μg poly I:C) and a protein boost (300 μg antigen peptides and 300 μg poly I:C) intradermally. Blood was collected after each protein boost for antibody response assays, and extra blood was collected at week 0,2,4,6,9 for safety assessment. PBMCs were extracted from the blood for T cell assays.

5.1.3 Results

No adverse effects were detected from the vaccination. A strong antibody response was detected for all three dogs after the first protein boost but was then suppressed by third round of immunization. T cell responses were detected after the first round of immunization and then it was strongly boosted by the second immunization. No tumor has developed in these three dogs so far and they were still under monitoring.

5.1.4 Conclusion

The frameshift antigen based cancer vaccine had no adverse effect and they induced both B cell and T cell immune response successfully.

5.2 Introduction

Cancer vaccine development suffered decades of failures with the conventional self-antigen based therapeutic vaccine [134,135]. With the success of immune checkpoint inhibitor treatments, neo-antigens are found to be the real target of the immune system and the number of neo-antigens (mutation load) is strongly correlated to better outcome of the ICI treatment [136]. Therefore, the idea of combining a neo-antigen based cancer vaccine and ICI treatment looks straightforward and promising [137,138].

However, most genomic mutations are patient specific point mutations, therefore, effectively making it impossible to develop a broadly effective neoantigen vaccine based on these genomic mutations. The personalized cancer vaccine idea was developed under these circumstances in response to this realization. However, this approach would

involve tremendous cost and the long development cycle may scare away some patients from using this vaccine.

We came up with a new idea of searching for frameshift neoantigens at the mRNA level instead of the genomic level. This approach has the advantages of being more immunogenic, more error prone and much more abundant. Three sources of frameshift neoantigens are available: alternate splicing frameshift variants; gene fusion frameshift variants and microsatellite indel induced frameshift variants. Note that we use the term variants because these are not heritable mutations in the DNA. The feasibility of using such frameshift antigens was described in previous chapters.

In this study, we selected 21 dog frameshift antigens from miss-splicing, trans-splicing and microsatellite regions. The vaccine was delivered intradermally with genetic immunization and protein boosts. Blood was collected at multiple time points for immune response and safety assessments.

5.3 Results

5.3.1 Dog Prophylactic Cancer Vaccine Components

Frameshift antigens from trans-splicing and miss-splicing were predicted from human tumor EST libraries. Dog homologs of human frameshift antigens were used in the 'dog version' cancer vaccine. While microsatellite FS were directly predicted from dog homo-polymer regions in coding sequences which had at least 7 homo-nucleotide repeats. Only FS antigens with human homologs were selected to provide for a human cancer vaccine

trial that could be conducted with highly conserved vaccine components if the dog efficacy vaccine trial was proved to be successful. Primers for most FS antigens were designed and FS antigens were validated by PCR in the dog tumor cDNA samples.

A dog frameshift peptide array was made to screen for serum positive FS antigens in 7 dog cancer types as described in chapter 4. Some FS antigens were then tested in the mouse tumor breast cancer (4T1) and melanoma (B16F10) injection models and showed protection. A detailed selection protocol is described in Fig5-1.

21 frameshift antigens (Table5-1) were selected as the best vaccine candidates which met most criteria described in the protocol (Fig5-1). The vaccine included all three frameshift antigen sources and most frameshift candidates had a peptide length of 20 amino acids to 100 amino acids. Because of their length, frameshift antigens are presumed to include multiple good MHC I and MHC II epitopes. Unfortunately, MHC prediction was not used as criterion since there is not an MHC prediction algorithm for dogs.

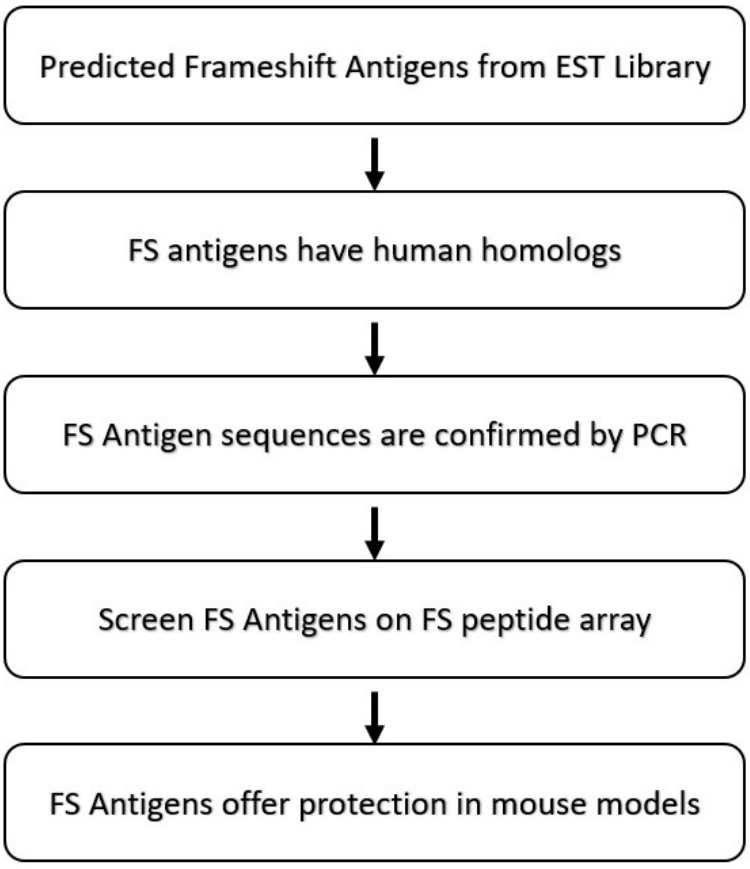


Figure 5-1. Protocol for selecting frameshift antigens.

Five criteria were used in selecting FS antigens for the dog cancer vaccine safety trial: these antigens were predicted from tumor EST libraries for mis-splicing antigens; the antigens had human homologs so the vaccine could be transformed into a human cancer vaccines; most frameshift antigens were confirmed by PCR; most FS antigens were positive in dog cancer serum samples from the frameshift peptide array screening; some FS antigens were tested in mouse models and they offered protection in injection tumor models.

The vaccine schedule included two rounds of genetic immunization and protein boost delivered intradermally (Fig5-2). A pre-bleed was taken before immunization, 2nd bleed and 3rd bleeds were taken after each protein boost, extra bleeds at week 2, 4, 9 were taken for safety measurements.

Table 5-1 Summary of dog prophylactic cancer vaccine candidates. Sequence confirmed: for miss-splicing and trans-splicing candidates, the predicted FS transcripts were detected in dog tumor cDNA. For microsatellite candidates, the predicted MS were confirmed in the coding region of the dog genes.

Name	FS type	FS peptide size	Sequence comfirmed	Positive on FS array	Protection in mouse model	Human Homolog	Dog Vaccine 1	Dog Vaccine 2
SMC1A	miss-splicing	27	✓		✓	✓	✓	✓
SLAIN2	miss-splicing	21		✓	✓	✓	✓	✓
RBM14-RBM4	trans-splicing	45			✓	✓	✓	✓
ELAC1-SMAD4	trans-splicing	44		✓		✓	✓	✓
CCDC13-HHATL	trans-splicing	83		✓		✓	✓	✓
C4H5orf22	microsatellite	30	✓		✓	✓	✓	✓
EFHC2	microsatellite	36	✓	✓		✓	✓	✓
OPHN1	microsatellite	80	✓	✓		✓	✓	✓
EIF2B3	microsatellite	56	✓			✓	✓	✓
SLC44A4	microsatellite	77				✓	✓	✓
ZDHHC17	miss-splicing	21		✓	✓	✓		✓
SLC35F5	microsatellite	23	✓		✓	✓		✓
CHD2	microsatellite	12	✓		✓	✓		✓
CEP290	microsatellite	25	✓	✓		✓		✓
ARV1	microsatellite	28	✓	✓		✓		✓
PLEKHM2	microsatellite	29	✓	✓		✓		✓
SWAP70 minus	microsatellite	30	✓	✓		✓		✓
ZC3HAV1	microsatellite	47	✓	✓		✓		✓
HNRNPH1	microsatellite	39	✓	✓		✓		✓
GAP43	microsatellite	32	✓	✓		✓		✓
SLITRK3	microsatellite	102	✓	✓				✓

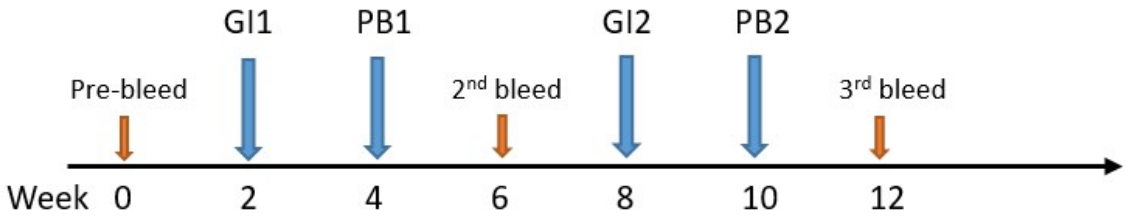


Figure 5-2. Immunization and blood collection schedule. GI: genetic immunization; PB: protein boost

5.3.2 There were no adverse events from this dog cancer vaccine

Blood tests were conducted at weeks 0, 2, 4, 6, 9 for assessing adverse effects of the cancer vaccine. Overall there were no adverse effects from the vaccination. No abnormal blood analyte concentration was detected at week 2 compared to week 0. At week 4, one dog showed only one percent higher red blood cells (56%) than normal range (40%~55%) and the percentage soon went back to normal level on week 6 and week 9 (Table S5-2). Similar results were found in neutrophil counts, monocyte counts and magnesium concentration. One dog had slightly higher or lower counts/concentration than the normal range in one time point. The counts/concentration then quickly went back to normal range (Table S5-1 to S5-3).

Overall, the FS antigen based prophylactic cancer vaccine was proved to be very safe.

Table 5-2 Vaccine-related abnormal blood analyte concentration in 3 dogs

Analyte	Week 2	Week 4	Week 6	Week 9
Hb	0	0	0	0
Hct	0	**1**	0	0
RBC	0	0	0	0
MCV	0	0	0	0
RDW	0	0	0	0
MCHC	0	0	0	0
CHCM	0	0	0	0

Platelets	0	0	0	0
MPV	0	0	0	0
WBC	0	0	0	0
Neutrophil	0	**1**	0	0
Lymph	0	0	0	0
Mono	0	0	0	**1**
Eo	0	0	0	0
Glucose	0	0	0	0
BUN	0	0	0	0
Creatinine	0	0	0	0
Phosphorous	0	0	0	0
Calcium	0	0	0	0
Magnesium	0	0	0	**1**
Total Protein	0	0	0	0
Albumin	0	0	0	0
Globulin	0	0	0	0
A/G Ratio	0	0	0	0
Cholesterol	0	0	0	0
CK	0	0	0	0
Bilirubin	0	0	0	0
ALP	0	0	0	0
ALT	0	0	0	0

AST	0	0	0	0
GGT	0	0	0	0
Sodium	0	0	0	0
Potassium	0	0	0	0
Chloride	0	0	0	0
Bicarbonate	0	0	0	0
Anion Gap	0	0	0	0
Calc Osmolality	0	0	0	0
PT	0	0	0	0
PTT	0	0	0	0
FDP	0	0	0	0

5.3.3 Vaccinated dogs showed positive Antibody Response

Blood was collected at week 0 (pre-bleed), week 6 (protein boost 1) and week 10 (protein boost 2) as shown in Figure 5-2 for antibody response assessment. Antibody response was measured for the 21-antigen pool and individual antigens in these three time points.

Antibody response against 21-antigen pool was significantly higher after protein boost 1 compared to pre-bleed (p value is 0.02 with one tail student t-test). However, the antibody response was then suppressed after protein boost 2 (Fig5-3A), which was possibly caused by over-immunization for certain candidates.

Antibody response against individual antigens was also assessed. Three categories of antigens were found from the antibody response pattern:

Ab response against 4 antigens (SLAIN2, GAP43, SLC44A4, SMC1A) was significantly increased from pre-bleed to PB1, but was then suppressed by PB2 (Fig5-3B, same pattern as the 21-antigen pool Ab response);

Ab response against another 4 antigens (CHD2, ZDHHC17, C4H5of22, ZC3HAV1) was not detected after PB1 but was then detected after PB2 (Fig5-3C), which indicated two rounds of immunization was necessary for these 4 antigens;

Ab response against the rest of the 13 antigens was not significant different between pre-bleed, PB1 and PB2, which indicated these antigens were outcompeted or they were not immunogenic at the antibody level.

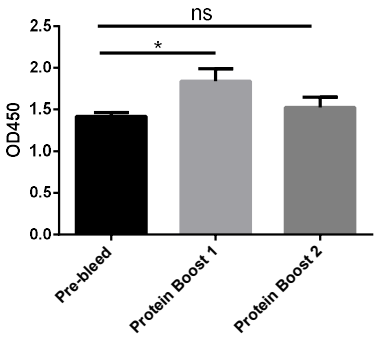
5.3.4 Each dog had a unique antibody response profile after each vaccination

The immunosignature assay was also conducted to measure the antibody response in these three dogs. Serum of the three dogs with 4-time points was applied on the CIM 330K non-natural random peptides array. Unsupervised hierarchal classification of the 12 samples showed that pre-bleed, PB1, PB2 and endpoint of each dog were classified as its own group and not mixed with other sample sets, which indicates each dog had its unique immune signature regardless of the immunization (Fig5-4A).

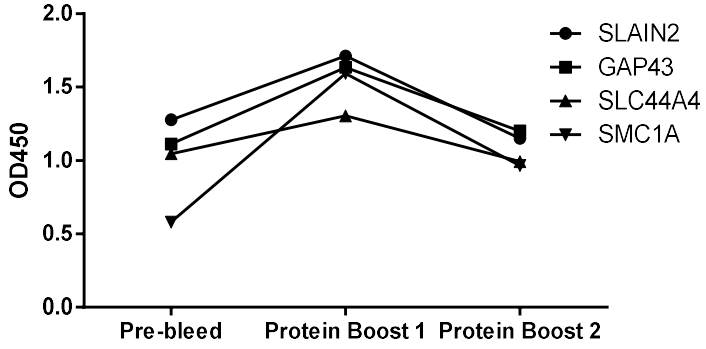
The analysis of 4-time points of each dog indicated that the antibody profile varied at each time point. Dog P had a large portion of the active antibody immune response after

PB1 (PVY3-2), but only a small percentage of the antibody response remained at high level after PB2 (PVY3-3). A totally different set of peptides showed positive response after PB2 (Fig5-4B). These data showed that the antibody profile of each dog was not very stable and it varied a lot with each immunization. The same phenomenon was found in dog Q and dog U as well (Fig5-4C, D). The results were consistent with the ELISA analysis in Fig 5-3, which also showed that different antigens were activated or suppressed with the two rounds of protein boost.

A



B



C

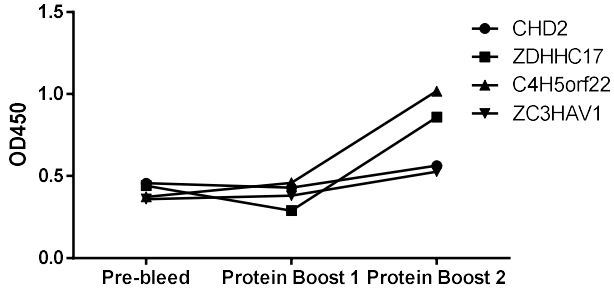
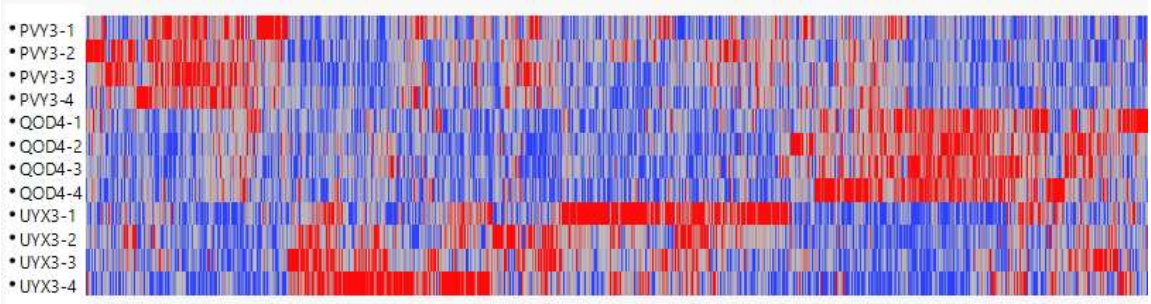
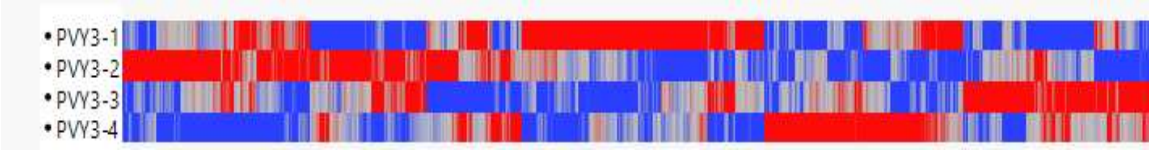


Figure 5-3. Antibody Response analysis of the vaccine. A. Antibody response against the 21 antigen pools in naïve samples, samples after protein boost 1 and protein boost 2. B. Antibody response against 4 individual antigens was detected in protein boost 1 but suppressed in protein boost 2. C. Antibody against 4 individual antigens was detected after protein boost 2.

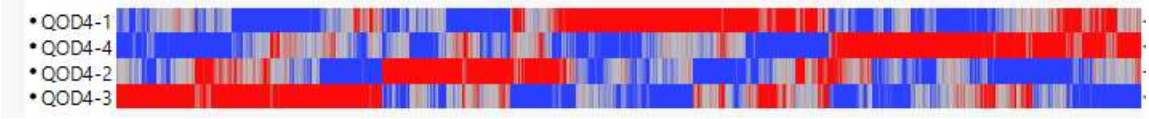
A



B



C



D

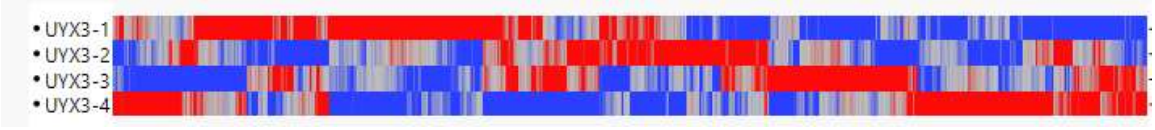


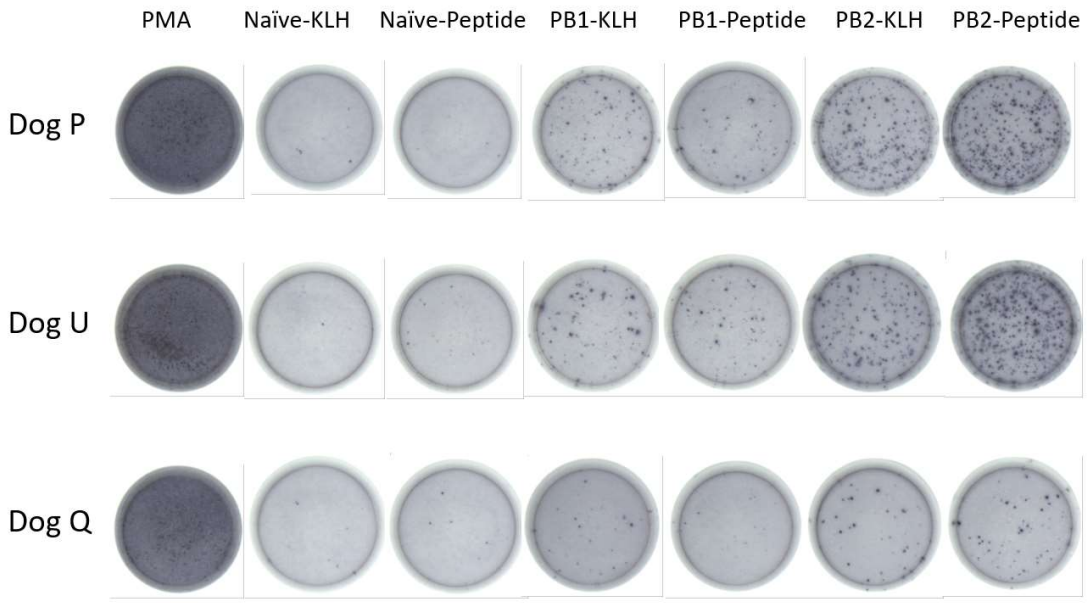
Figure 5-4. Immunosignature assay of the vaccine. A. Unsupervised hierarchal classification (Ward method) of the 3 dogs with 4-time points (1: pre-bleed; 2: PB1; 3: PB2; 4: End timepoint of the trial); B~D. Same analysis as in A was performed for individual Dog P, Dog Q and Dog U.

5.3.5 Vaccinated dogs showed positive T cell immune responses

PBMCs were extracted from blood collected at pre-bleed, PB1 and PB2. IFN γ ELISPOT assays and intracellular staining were used to measure the T cell response against the immunized antigens.

Pre-bleed PBMCs had very low T cell responses with 48hr KLH stimulation or the 21-antigen pool stimulation. IFN γ secreting T cells could be rarely detected (Fig5-5A). With two rounds of immunization, IFN γ secreting T cells increased dramatically from pre-bleed to PB1 and from PB1 to PB2. Two immunizations were necessary for eliciting high T cell responses. Quantification of the results are summarized in Fig5-5B. PBMCs from PB1 and PB2 had significantly more IFN γ secreting T cells compared to pre-bleed (p-value <0.01). Interestingly, KLH stimulation or the 21-antigen pool stimulation induced about the same amount of IFN γ secreting T cells after PB1, but the 21-antigen pool stimulation induced significantly more active T cells than KLH after PB2, which indicated T cells were directed to target our vaccinated antigens after PB2 and not KLH.

A



B

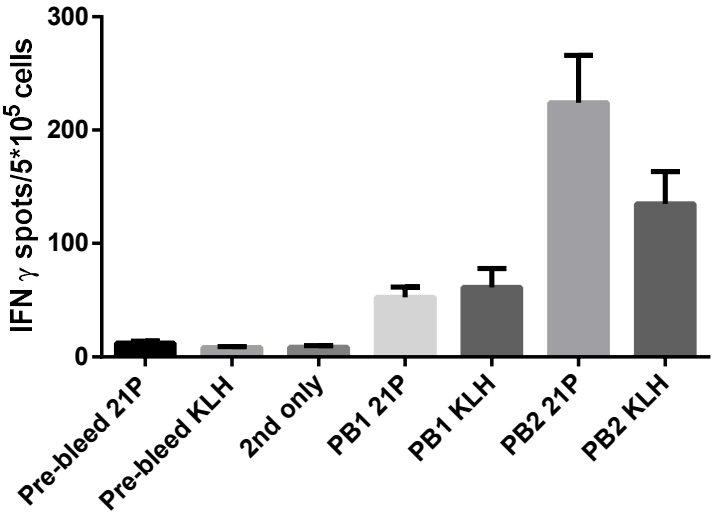
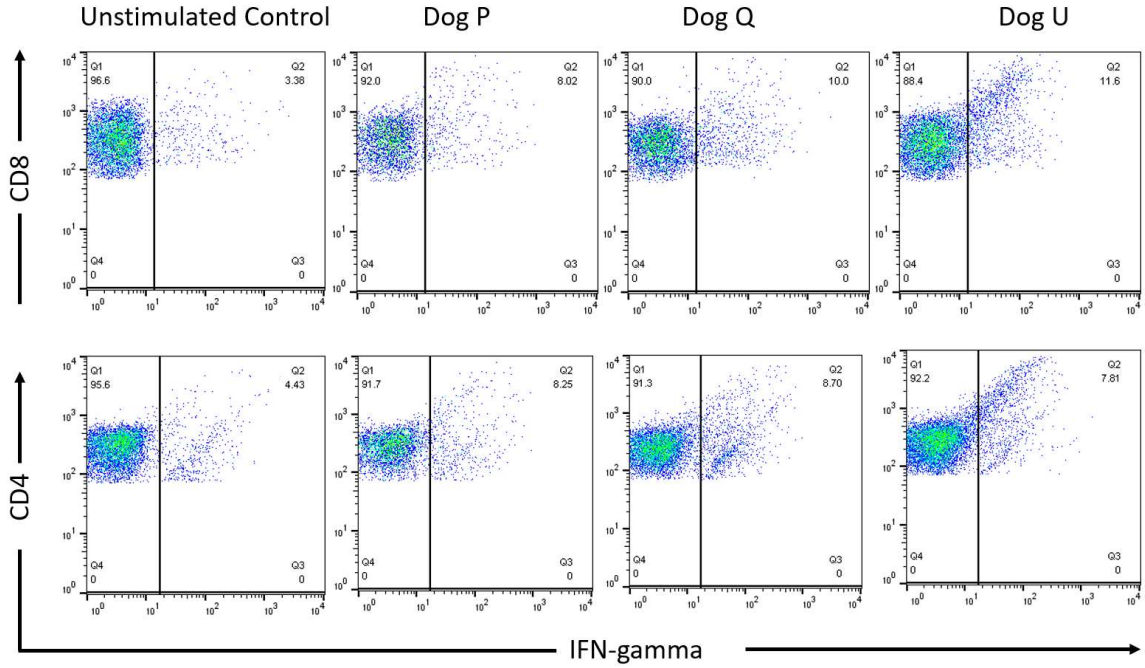


Figure 5-5. T-cell response analysis of the vaccine. A. IFN γ releasing assay graph of 3 dogs in the safety trial. B. Summary statistics of the IFN γ releasing spots.

A T cell intracellular staining assay was also conducted to measure the T cell response. PBMCs after PB2 were used for the assay. Since there was not a very clear lymphocytes population in the staining, all the PBMCs were used for the analysis (FigS5-1). As shown in Fig5-6A, all three dogs had more CD4+ IFN γ secreting T cells and more CD8+ IFN γ secreting T cells than unstimulated control when stimulated with the 21 antigen pools. Quantification of the results showed that there were slightly more active IFN γ secreting CD8+ cells (~10%) than CD4+ cells (7%~8%). KLH stimulation and 21 antigen pool stimulation had about the same percentage of active T cells, while positive control (PMA + Ionomycin) induced more active T cells.

A



B **C**

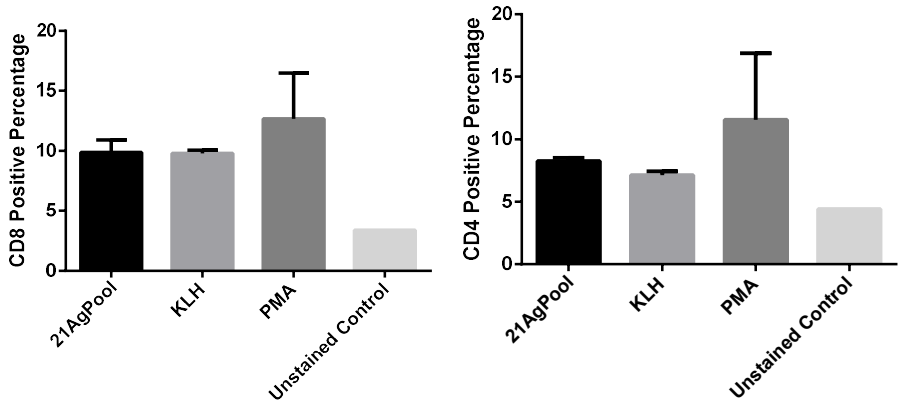


Figure 5-6. IFN γ secreting T cell subtype analysis. A. IFN γ releasing CD4 and CD8 T cells with the 21-antigen pool stimulation. Unstimulated T cells were used as control. B. Quantification of CD4 and CD8 IFN γ secreting T cells with stimulation of the 21-antigen pool, KLH and Positive control (PMA+Ionomycin).

Both the ELISPOT assay and intracellular T cell staining assay showed there was a positive T cell response after two rounds of immunization, which indicates that the vaccine was not only safe but also could induce strong B cell and T cell responses.

5.4 Discussion

This safety trial of the 21-frameshift antigen pool, formulated for a dog prophylactic cancer vaccine, demonstrated that the vaccine was safe with no adverse effects and it could induce strong B cell and T cell response.

The idea of using frameshift antigens as a broadly effective prophylactic cancer vaccine source could reduce the vaccine development cost significantly and may revolutionize the

cancer therapy field, but it has not been explored. Here we demonstrated that this approach was safe and applicable.

We recognize that intradermal injection was not the optimal delivery method for the genetic vaccine. Vaccination with gene gun could induce stronger T cell response [139]. Furthermore, with the development of virus-based vaccination [141], RNA liposome vaccination [140] and other techniques [142], effective immune responses from the vaccine could be enhanced dramatically.

In addition, the key step of developing a successful cancer vaccine was the antigen selection process. With our current strategy, antibody response against individual antigens in the 21 antigens pool showed that only a small portion of our antigens (8 out of 21) induced B cell response successfully after PB1 and PB2. This could be due to antigen competition and over immunization problems. There is still huge space for improvement of the vaccine components and vaccination protocol. A second version of the dog cancer vaccine was developed and discussed in Appendix B.

In summary, the data in this study suggested that a frameshift antigens pool based prophylactic cancer vaccine was very safe and it induced strong B cell and T cell response.

5.5 Methods and Materials

5.5.1 Dogs enrolled in the safety trial

Three dogs (Beagles) were purchased and used in the safety trial. There were named with Dog PVY3, QOD4, UYX3 in the paper.

5.5.2 Vaccination Schedule

GI: DNA fragments encoding each antigen were cloned into the plasmids (PCMVi-UB and PCMVi-LS-rCOM-TT) and the two plasmids were mixed equally as an antigen composition for the GI. Antigens were equally mixed according Table 5-1 for each candidate. One GI is composed by 300ug mixed antigens and 300ug Poly I:C in 500ul PBS. The vaccine was delivered intradermally.

PI: Overlapped peptides were synthesized for each antigen. Synthesized peptides for the same antigen were equally mixed as a pool and conjugated with KLH 1:1 by weight as an antigen composition for the PI. Each peptide conjugated KLH were equally mixed according the Table 5-1 for each candidate. One PI was composed by 300ug mixed peptide-KLH and 300ug Poly I:C in 500ul PBS. The vaccine was delivered intradermally.

5.5.3 Analysis of antibody response by ELISA

ELISA was used to detect specific antibody response against immunized peptides. Multiple ELISA experiments in the paper were performed with the same protocol as in previous chapters. Briefly, 96-well plates (Thermo Fisher Scientific, Waltham, MA) were coated with 50ul 10 μg/ml peptide or KLH overnight at 4 ºC, then it was blocked with 3% BSA/PBST for 1hr at 37 ºC. After incubation with sera for 1hr at 37 ºC, wells were probed with HRP Anti-dog IgG for 1hr at 37 ºC, then the plate was developed using TMB

substrate solution and terminated with 0.5N HCl. Plates were assessed using SpectraMax 190 Molecular Devices at OD 450nm.

5.5.4 Immunosignaturing Assay

The CIM 330K peptide array was used for immunosignature assays. A detailed protocol has been published previously [82]. Basically, microarrays were pre-washed in 10% acetonitrile, 1% BSA to remove unbound peptides. Then the slide was blocked with 1*PBS, 3%BSA, 0.05% Tween 20, 0.014% β-mercaptohexanol for 1hr at RT. After that the slide was loaded into multi-well array gasket, serum samples were diluted 1:5000 in 1*PBS, 3%BSA, 0.05% Tween 20 (PBST) dilution buffer. Serum samples were incubated for 1hr at RT with BioTek 405TS plate washer. Anti-dog IgG secondary antibody with conjugated dye was added to serum samples at final concentration of 5 nM. Unbound secondary antibody was washed away with PBST and water. Arrays were then centrifuged, dried and scanned. GenePix 8.0 (Molecular Devices, CA) was used for analyzing scanned picture. Each array was normalized to median and JMP Pro (SAS institute, CA) was used for non-supervised clustering.

5.5.5 IFN-gamma ELISPOT Assay

The ELISPOT assay was used to calculate frequency of cells releasing interferon gamma. A PVDF plate was pre-coated with canine anti-IFN γ monoclonal antibody overnight at 4 ºC, then the plate was blocked with blocking buffer (1%BSA, 5%Sucrose in PBS) for 2hr at room temperature. After that, 0.5 million dog PBMC cells were co-incubated with 20ug/ml FS peptides or 1ug/ml Ionomycin and 50ng/ml PMA for 48hrs in a humidified 37 ºC CO2 incubator. The wells were then aspirated and washed with wash

buffer (0.05%Tween-20 in PBS), detection antibody was added to each well and incubated overnight at 4 ºC. After washing wells 3 times with wash buffer, Streptavidin-AP was added to each well and incubated for 2hr at room temperature, then plate was developed with substrate solution (BCIP/NBT) in dark. The plate was scanned and then positive spots were counted by software.

5.5.6 T-cell in vitro staining Assay

Dog PBMCs were collected at weeks 0, 4 and 10 and stored in liquid nitrogen. Cells were recovered from liquid nitrogen and cultured in RMPI1640 complete media (1million cells/200µl) in 96 well plates. Dog vaccinated peptides pool (20 µg/ml) were added to the media and PMA plus Ionomycin was used as positive control. 4 µl of BD Golgistop was added for every 6ml of media at the same time. After 7hr incubation, cells were washed and stained with anti-CD4 and anti-CD8 antibodies for 30 mins at 4°C, then cells were fixed in Fixation/Permeabilization solution for 20 mins at 4°C, after that, cells were washed and stained with anti- IFN γ antibody for 30 mins in the dark at 4°C. Cells were then suspended in staining buffer and analyzed in flow cytometer.

Supplemental Data

Table S5-1. Blood work for dog PVY3

Dog PVY3							
Analyte	**Normal Range**	**Units**	**Week 0**	**Week 2**	**Week 4**	**Week 6**	**Week 9**
Hb	13.0 - 20.0	g/dL	17.7	17.1	17.8	17.5	17.4
Hct	40 - 55	%	52	49	51	51	51
RBC	5.5 - 8.5	10^6/uL	8.01	7.72	7.95	7.91	7.92
MCV	62 - 74	fL	65	64	64	65	65
RDW	12.0 - 15.0	%	12.6	12.1	12.1	12.5	12.4
MCHC	33 - 36	g/dL	34	35	35	34	34
CHCM	33 - 36	g/dL	34	35	**35**	34	33
Platelets	200 - 500	10^3/uL	366	352	382	366	340
MPV	7.5 - 14.6	fL	8.7	8.8	8.7	8	8.6
WBC	4.5 - 15.0	10^3/uL	11.9	10	14.8	10.7	9.2

Neutrophil	2.6 - 11	10^3/uL	6.8	6.4	**11.4**	7.5	6.1
Lymph	1 - 4.8	10^3/uL	**4.9**	3	3.1	2.5	2.3
Mono	0.2 - 1.0	10^3/uL	0.2	0.4	0.3	0.4	0.6
Eo	0.1 - 1.2	10^3/uL	0	0.2	0	0.3	0.3
Glucose	70 - 115	mg/dL	**51**	87	77	92	80
BUN	7.0 - 30	mg/dL	24	11	22	11	14
Creatinine	0.6 - 1.6	mg/dL	1	0.7	1.1	0.7	0.7
Phosphorous	2.5 - 6.0	mg/dL	5.2	3.5	4.9	3.7	4.3
Calcium	9.0 - 11.5	mg/dL	10.5	10.5	10.3	10.4	9.9
Magnesium	1.8 - 2.4	mg/dL	2	1.8	2	1.8	1.8
Total Protein	5.0 - 7.0	g/dL	6.1	6.3	6.1	6.2	6
Albumin	3.0 - 4.3	g/dL	3.8	3.9	3.7	3.7	3.8
Globulin	1.5 - 3.2	g/dL	2.3	2.4	2.4	2.5	2.2

A/G Ratio	0.9 - 2.4	RATIO	1.7	1.6	1.5	1.5	1.7
Cholesterol	130 - 300	mg/dL	186	177	189	168	198
CK	50 - 275	IU/L	324	189	260	176	239
Bilirubin	0.0 - 0.2	mg/dL	0	0.1	0	0.1	0.1
ALP	15 -140	IU/L	32	32	36	30	32
ALT	10.0 - 90	IU/L	37	37	31	33	32
AST	15 - 45	IU/L	33	29	30	31	30
GGT	0.0 - 9	IU/L	1	3	0	3	3
Sodium	142 - 152	mEq/L	147	147	146	148	148
Potassium	3.9 -5.4	mEq/L	5.25	4.06	4.58	4.19	4.06
Chloride	108 - 118	mEq/L	110	112	109.9	111.9	111.8
Bicarbonate	15 - 25	mEq/L	19.5	18.4	18.9	21	30.7
Anion Gap	12.0 - 23	nmol/L	23	21	22	19	20
Calc Osmolality	260 - 290	mOsm/kg	**295**	290	**292**	**292**	**292**

PT	7.1 - 9.1	Sec	7.7	7.5	6.9	7.8	7.5
PTT	9.4 - 15.0	Sec	10.3	11.1	10.8	10.2	10.9
FDP	0 - 4	ug/mL	NP	NP	Neg	Neg	Neg

Table S5-2. Blood work for dog QOD4

QDD4							
Analyte	**Normal Range**	**Units**	**Week 0**	**Week 2**	**Week 4**	**Week 6**	**Week 9**
Hb	13.0 - 20.0	g/dL	17.8	17.9	19.5	18.6	18.2
Hct	40 - 55	%	50	50	**56**	54	53
RBC	5.5 - 8.5	10^6/uL	7.18	7.24	7.99	8.07	7.62
MCV	62 - 74	fL	70	69	70	67	69
RDW	12.0 - 15.0	%	13	12.8	13.3	13.3	13.3
MCHC	33 - 36	g/dL	35	36	35	34	34
CHCM	33 - 36	g/dL	35	36	36	34	35
Platelets	200 - 500	10^3/uL	254	232	284	312	242

MPV	7.5 - 14.6	fL	8.9	8.6	10.1	8.8	8.8
WBC	4.5 - 15.0	10^3/uL	8.4	7.4	9.1	9	9.3
Neutrophil	2.6 - 11	10^3/uL	5.4	4.8	6.6	5.9	6.5
Lymph	1 - 4.8	10^3/uL	1.9	1.9	2	2.3	2.4
Mono	0.2 - 1.0	10^3/uL	**1.1**	0.6	0.5	0.5	0.3
Eo	0.1 - 1.2	10^3/uL	0	0.1	0	0.3	0.1
Glucose	70 - 115	mg/dL	98	85	95	84	92
BUN	7.0 - 30	mg/dL	23	14	21	15	20
Creatinine	0.6 - 1.6	mg/dL	1	0.9	1.1	0.8	0.9
Phosphorous	2.5 - 6.0	mg/dL	5.3	3.9	5.7	4.3	4.4
Calcium	9.0 - 11.5	mg/dL	10.1	10.3	10.1	10.5	9.6
Magnesium	1.8 - 2.4	mg/dL	1.9	1.9	2	2.1	**1.7**

Total Protein	5.0 - 7.0	g/dL	5.6	5.9	5.7	6	5.5
Albumin	3.0 - 4.3	g/dL	3.6	3.7	3.7	3.9	3.4
Globulin	1.5 - 3.2	g/dL	2	2.2	2	2.1	2.1
A/G Ratio	0.9 - 2.4	RATIO	1.8	1.7	1.9	1.9	1.6
Cholesterol	130 - 300	mg/dL	203	194	219	244	214
CK	50 - 275	IU/L	188	190	175	101	204
Bilirubin	0.0 - 0.2	mg/dL	0	0.1	0.1	0.1	0.1
ALP	15 -140	IU/L	39	34	42	27	44
ALT	10.0 - 90	IU/L	46	41	37	28	28
AST	15 - 45	IU/L	25	27	28	26	29
GGT	0.0 - 9	IU/L	0	0	0	3	1
Sodium	142 - 152	mEq/L	146	146	147	148	146
Potassium	3.9 -5.4	mEq/L	4.99	4.25	4.55	4.69	4.36
Chloride	108 - 118	mEq/L	112.6	112.7	108.6	111.5	111.5
Bicarbonate	15 - 25	mEq/L	20.8	17.2	20	21.7	19.1

Anion Gap	12.0 - 23	nmol/L	18	20	23	19	20
Calc Osmolality	260 - 290	mOsm/kg	**295**	289	**295**	**294**	**292**
PT	7.1 - 9.1	Sec	7.3	NA	7.4	7.9	NP
PTT	9.4 - 15.0	Sec	10.7	NA	10.2	10.6	NP
FDP	0 - 4	mg/dL	NP	NP	Neg	Neg	NP

Table S5-3. Blood work for dog UYX3

UYX3							
Analyte	**Normal Range**	**Units**	**Week 0**	**Week 2**	**Week 4**	**Week 6**	**Week 9**
Hb	13.0 - 20.0	g/dL	18.2	17.8	18.8	NP	18.6
Hct	40 - 55	%	52	51	64	NP	55
RBC	5.5 - 8.5	10^6/uL	7.77	7.62	8.11	NP	8.27
MCV	62 - 74	fL	67	67	66	NP	67
RDW	12.0 - 15.0	%	13.1	13	12.9	NP	13.3

MCHC	33 - 36	g/dL	35	35	35	NP	34
CHCM	33 - 36	g/dL	35	35	**36**	NP	34
Platelets	200 - 500	10^3/uL	361	318	32	NP	317
MPV	7.5 - 14.6	fL	9.5	9	9.9	NP	9.1
WBC	4.5 - 15.0	10^3/uL	9.5	7.9	10.1	NP	8.4
Neutrophil	2.6 - 11	10^3/uL	6.5	4.7	5.6	NP	5.8
Lymph	1 - 4.8	10^3/uL	2.7	2.1	3.9	NP	2.5
Mono	0.2 - 1.0	10^3/uL	0.2	0.9	0.6	NP	**0.1**
Eo	0.1 - 1.2	10^3/uL	0.2	0.2	0	NP	0
Glucose	70 - 115	mg/dL	72	90	82	NP	79
BUN	7.0 - 30	mg/dL	23	22	27	NP	17
Creatinine	0.6 - 1.6	mg/dL	1	0.9	1.1	NP	0.8
Phosphorous	2.5 - 6.0	mg/dL	4.7	5.6	5.6	NP	5

Calcium	9.0 - 11.5	mg/dL	10.4	10.4	10.3	NP	10
Magnesium	1.8 - 2.4	mg/dL	2.1	2.1	2	NP	2
Total Protein	5.0 - 7.0	g/dL	6.1	6	5.9	NP	5.9
Albumin	3.0 - 4.3	g/dL	3.9	3.8	3.8	NP	3.8
Globulin	1.5 - 3.2	g/dL	2.2	2.2	2.1	NP	2.1
A/G Ratio	0.9 - 2.4	RATIO	1.8	1.7	1.8	NP	1.8
Cholesterol	130 - 300	mg/dL	237	234	254	NP	227
CK	50 - 275	IU/L	130	114	122	NP	113
Bilirubin	0.0 - 0.2	mg/dL	0.1	0.1	0.1	NP	0.1
ALP	15 -140	IU/L	30	28	28	NP	28
ALT	10.0 - 90	IU/L	30	29	289	NP	27
AST	15 - 45	IU/L	28	26	23	NP	28
GGT	0.0 - 9	IU/L	2	2	0	NP	3
Sodium	142 - 152	mEq/L	147	145	147	NP	149
Potassium	3.9 -5.4	mEq/L	5.14	4.57	4.99	NP	4.79

Chloride	108 - 118	mEq/L	110.8	113.3	110.9	NP	112.9
Bicarbonate	15 - 25	mEq/L	19.8	15.7	19.4	NP	18.9
Anion Gap	12.0 - 23	nmol/L	22	21	22	NP	22
Calc Osmolality	260 - 290	mOsm/kg	**295**	**291**	**297**	NP	**297**
PT	7.1 - 9.1	Sec	7.2	7.8	7.8	NP	7.8
PTT	9.4 - 15.0	Sec	10.3	11	9.7	NP	10.5
FDP			NP	NP	Neg	NP	Neg

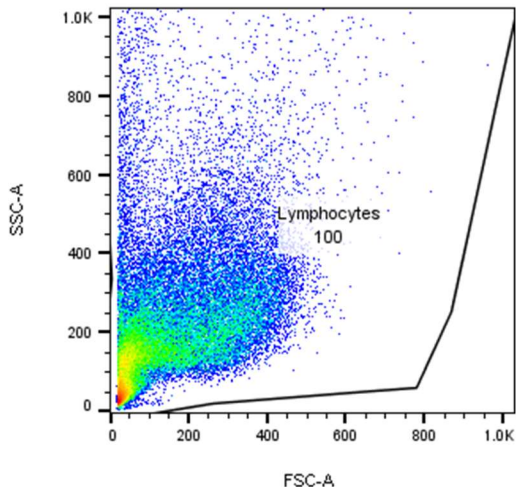


Figure S5-1. Gating of T cells.

Since there was not a clear lymphocytes population, all cells were used for analysis.

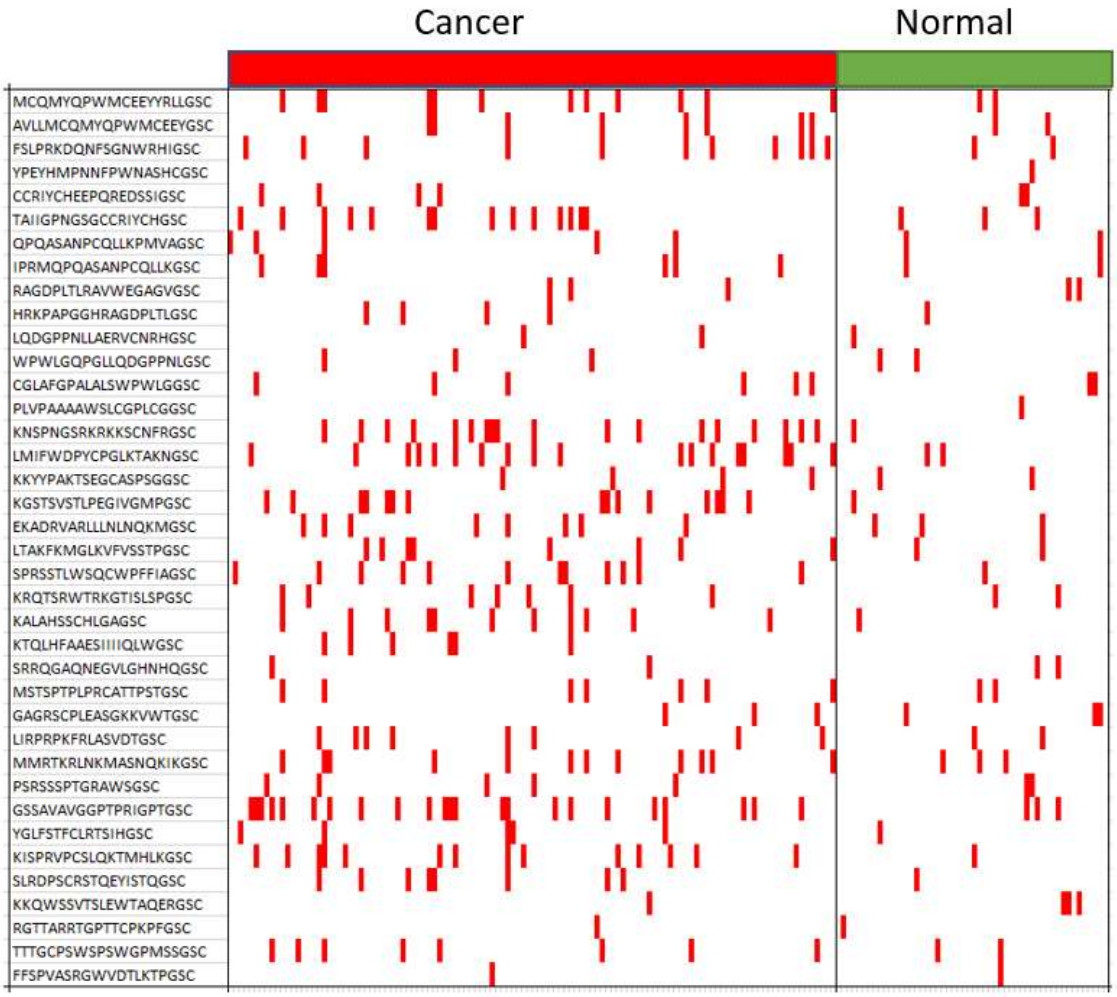


Figure S5-2. Positive Antibody Response coverage of the Dog Cancer Vaccine in Dog Cancer Samples.

Total coverage is 84.5% in all the 116 dog cancer samples.

Table S5-4 Dog Cancer Vaccine Antibody Response Coverage by cancer type

	Sample Numbers	Coverage
Carcinoma	13	92.31%
Fibrosarcoma	20	85.00%
Hemangiosarcoma	17	82.35%
Lymphoma	10	80.00%
Malignant histiocytosis	4	75.00%
Mast Cell Tumor	22	81.82%
Osteosarcoma	20	85.00%
Sarcoma	8	87.50%
Synovial cell sarcoma	2	100.00%

Chapter 6

Human Exon Junction Frameshift Peptide Database and its Applications

6.1 Abstract

Frameshift (FS) mutations from microsatellite regions and mis-splicing events were detected frequently in tumor mRNA. In addition, FS mutation load was also reported to be a better biomarker of beneficial immune checkpoint inhibitor treatment response than total mutation load [18]. However, FS peptides from mis-splicing have not been systematically investigated. Here I created a human an exon junction FS database which covered all possible FS peptides from human exon junctions (as well as MS FSs) and summarized the characteristics for the database. Furthermore, I provided an indirect method to find potential FS antigens from cancer immunosignature data with this FS database. In the meantime, a new platform to assay unbiasedly the anti-FS humoral immune response - the human 400K frameshift peptide array which covers all possible FS peptides was developed.

6.2 Introduction

Alternative splicing is a regulated process in mRNA maturation which uses different exons of the same gene to generate different functional proteins. It is reported that 95% of multi-exonic genes are alternatively spliced in humans [83]. Meanwhile, gene fusions are commonly detected in cancer patients due to chromosomal changes. Frameshift mutations within these two processes can create neo-antigens. Since the alternative splicing and chromosomal mutations are so frequent, it is likely that frameshift neo-

antigens from mis-splicing are common as well with the high mutation rate of RNA spliceosome. However, these potential FS peptides from mis-splicing have not been systematically investigated or used in cancer vaccine development. In this chapter, I built the database for all potential FS peptides from mis-splicing in exon junctions and provide potential applications for using this database.

One report in 2012 analyzed frequent frameshift mutations from mis-splicing events in human tumor EST libraries. 2,996 FS peptides from mis splicing within the same gene were found and 96 FS peptides were specific or highly enriched in tumor libraries. While 321 FS peptides from mis-splicing in gene fusions were found [81]. However, with the limited number of tumor libraries and sequencing data, only a very small percentage of all possible FS mutations were detected.

All possible FS peptides from human exon junctions were analyzed in this chapter and the basic characteristics are summarized. A large human breast cancer immunosignature dataset including different stages of breast cancer was used to predict putative FS antigens indirectly according to different patterns. An unbiased anti-FS screening platform is under development. The human 400K peptide array was designed to cover all possible microsatellite frameshift peptides and mis-splicing FS peptides.

6.3 Results

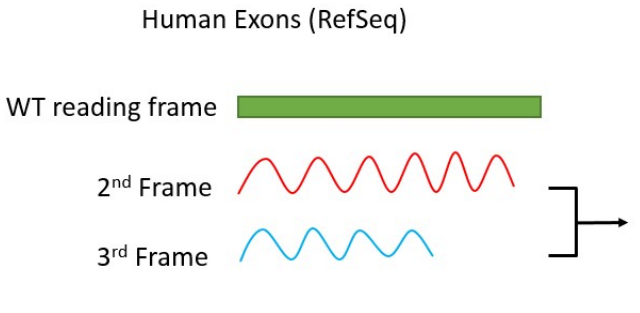
6.3.1 Characteristics of the human exon junction FS database

Frameshift peptides from microsatellite regions are not the sole source of potential frameshift antigens. To investigate unbiasedly potential frameshift antigens in human genes, all possible splicing frameshift peptides from alternate splicing within genes or

between genes (gene-fusions) need to be covered. To predict the frameshift peptide sequences from the second and third reading frames of every human exon, a total number of 41,930 mRNA sequences representing 19,270 genes were downloaded from NCBI RefSeq database [52] and coordinates of 421,337 exons were recorded. Then the second and third reading frames of each exon were translated into FS peptide sequences and stored in a human exon junction FS database. As shown in Figure 6-1, the ID of each frameshift candidate includes the mRNA accession number, exon number of this gene and its corresponding frame.

Summary statistics of the database are shown in Table 6-1. There are total 842,672 output frameshift sequences with a minimum length of 0 and maximum length of 1082 amino acids. The median FS length is 12 and the average length is 17 amino acids, and there are about 15 million FS amino acids in total.

The distribution of the frameshift peptides length is summarized in Figure 6-2. The number of frameshift candidates decreases exponentially as FS peptide length increases. The curve of actual FS length distribution is close to a random distribution (calculated by frequency of stop codons). The Pearson correlation between the two curves is 0.99. These data indicate frameshift peptide length is not under selection and they follow a random distribution.



Human Exon Junction Frameshift Database

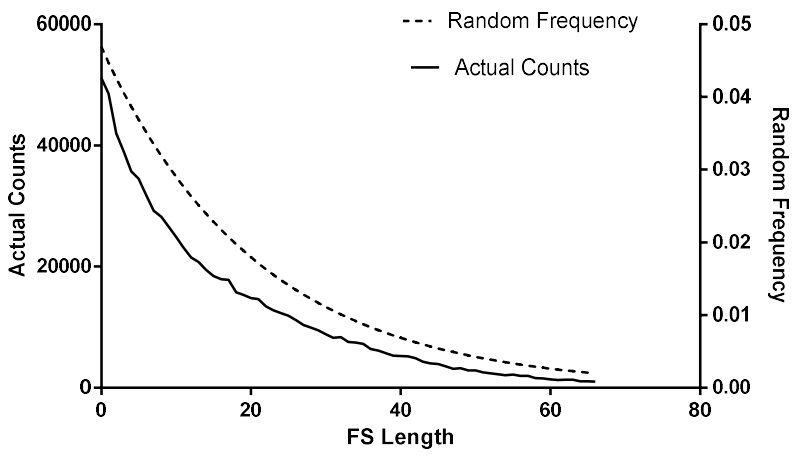
ID	Gene	Frameshift Peptide
NM_001204516.1_Exon1_2nd	KIFAP3	CKGRTPDTSK
NM_001243136.1_Exon2_2nd	PELI3	CTTSPRRSSP
NM_001301138.1_Exon1_2nd	VPS39	CTTLSSQCRS
NM_001302453.1_Exon1_2nd	CCDC151	CTLRLPPVRN
NM_001304441.1_Exon2_2nd	MMP8	CNKYLKRSQL
NM_001308192.1_Exon2_2nd	CPEB4	CTHWRVHSLT
NM_001308351.1_Exon4_2nd	ZNF433	CKKPSGTWPL
NM_014350.3_Exon1_2nd	TNFAIP8	CTPKQKNPRK
NM_139320.1_Exon3_2nd	CHRFAM7A	CKNIASTSIF

Figure 6-1. The procedure for building the Human Exon Junction Frameshift Database.

Table 6-1. Summary statistics of human exon junction FS database.

Mean	**17.66**
Standard Error	0.02
Median	12.00
Mode	0.00
Standard Deviation	20.47
Sample Variance	418.86
Kurtosis	74.21
Skewness	4.80
Range	1082.00
Minimum	0.00
Maximum	1082.00
Sum	14880663.00
Count	842672.00

A



B

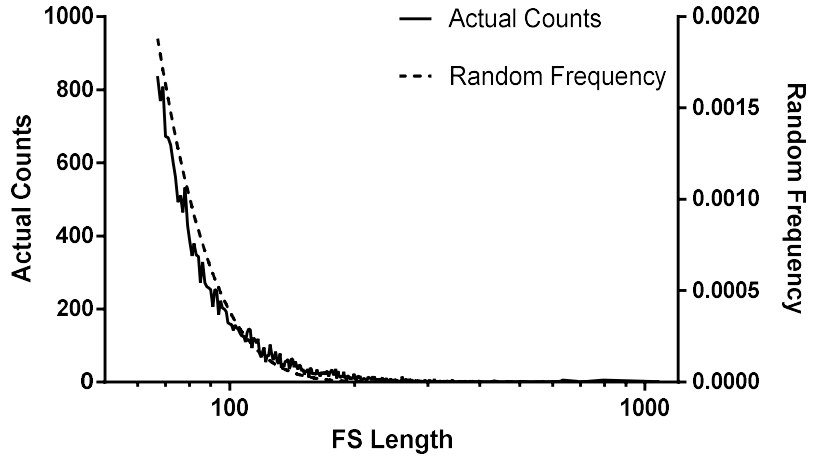


Figure 6-2. Random distribution of frameshift peptide length. A-B. Actual frameshift peptide counts of different frameshift peptide lengths (FS Length 0-80 was shown in A and above 80 was shown in B) and corresponding random frequency with different FS peptide length. Both distribution curves were similar to the random distribution and FS antigens with long FS length had higher similarity to the random distribution.

The characteristics of microsatellite frameshift peptides are summarized in previous chapters. Here it is also interesting to find out whether the amino acid composition of the splicing frameshift peptides from exon junctions is like the microsatellite frameshift peptides. These two databases are both generated from human coding genes but with totally different selection criteria. The MS FS database covers all MS regions in coding genes regardless of the microsatellite locations, while the exon junction FS database covers all frameshift peptides from the start location of every exon and it is 30 times bigger than the MS database.

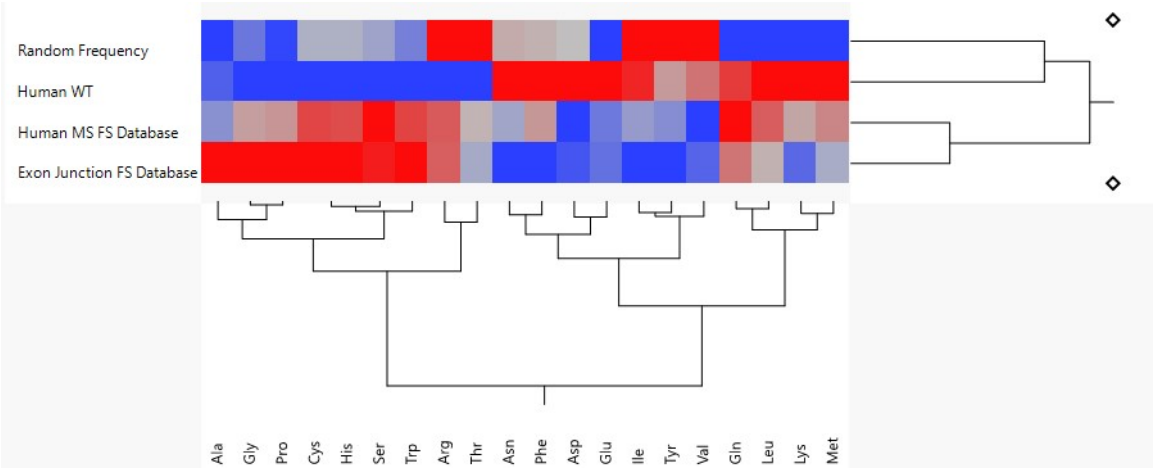
Detailed amino acid composition data is summarized in table 6-2 and a two-way clustering heatmap is shown in Figure 6-3. The exon junction FS database has almost identical amino acid composition as the MS FS database with a correlation of 0.98. These two databases both have high correlation with the random frequency of 0.88 ~ 0.9, which indicates that the amino acid composition of frameshift peptides is not under high selection pressure. Alanine, Glycine, Proline, Cysteine, Histidine, Serine and Tryptophan are enriched in frameshift sequences. It is unknown why some of these amino acids have higher percentage in frameshift sequences than the random frequency.

Table 6-2. Amino acid composition of the human exon junction FS database and its comparison to random frequency, WT sequences and the human microsatellite FS database.

	AA	Random Frequency	Human WT	Human MS FS Database	Exon Junction FS Database
Ala	A	6.25%	6.36%	6.53%	7.83%
Arg	R	9.38%	5.27%	8.91%	8.88%
Asn	N	3.13%	3.98%	2.72%	2.03%
Asp	D	3.13%	5.04%	2.08%	2.22%
Cys	C	3.13%	1.94%	4.04%	4.29%
Glu	E	3.13%	7.88%	3.68%	3.59%
Gln	Q	3.13%	5.00%	5.13%	4.82%
Gly	G	6.25%	5.85%	7.25%	8.30%
His	H	3.13%	2.54%	3.57%	3.71%
Ile	I	4.69%	4.60%	3.52%	2.80%
Leu	L	9.38%	9.55%	9.53%	9.49%
Lys	K	3.13%	6.81%	5.04%	3.53%
Met	M	1.56%	2.09%	1.95%	1.79%
Phe	F	3.13%	3.43%	3.18%	2.61%
Pro	P	6.25%	6.20%	8.06%	9.37%
Ser	S	9.38%	8.88%	10.01%	9.98%
Thr	T	6.25%	5.42%	5.85%	5.73%
Trp	W	1.56%	1.03%	3.04%	3.31%
Tyr	Y	3.13%	2.46%	1.83%	1.40%

Val	V	6.25%	5.67%	4.08%	4.31%

A



B

	Random Frequency	***Human WT***	***Human MS FS Database***	***Exon Junction FS Database***
Random Frequency	1.00			
Human WT	0.71	1.00		
Human MS FS Database	0.90	0.72	1.00	
Exon Junction FS Database	0.88	0.66	0.98	1.00

Figure 6-3. Compare amino acid composition of the human exon junction database to different databases A. Two-way clustering of the amino acid composition of the human exon junction database, human WT proteome, human MS FS database and random frequency. B. Correlation matrix of amino acid composition.

6.3.2 Exon Frameshift peptide sequences in WT human proteome

Frameshift peptide sequences of one gene can have partial identity to the wildtype protein sequence another gene-encoded protein, presumably by chance. Frameshift peptide sequences with identity to sequences of WT proteins are not ideal neo-antigens and need to be removed.

To compare the frameshift peptides with WT sequences, human RefSeq protein sequences were used as WT sequences and compared to human exon junction frameshift peptides. BLAST (Basic Local Alignment Search Tool) was used to search for sequence matches. Output frameshift peptide matches with a minimum of 15 identical amino acids and minimum 98% similarity to WT protein sequences were classified into a group of frameshift peptides which contain WT peptide sequences.

There is a total of 2,033 frameshift candidates with WT matches with these criteria, which represents less than 1% of the whole FS exon junction database. Since these matches indicate that WT protein sequences include partial matches of frameshift sequences, it is interesting to know whether these matches are enriched in a specific group of genes with certain motifs. Gene ontology enrichment analysis was used to analyze these 1,399 unique genes from the 2,033 FS genes with WT matches [84]. Results of both biological process and molecular function gene ontology analysis

indicated that these 1,399 unique genes were not highly enriched in one specific biological process or molecular function. The most enriched biological process was protein complex subunit organization with 1.54-fold enrichment and only 134 FS containing genes, and the most enriched molecular function was identical protein binding with 1.5-fold enrichment and only 147 FS containing genes.

Table 6-3 Biological process gene ontology analysis of FS genes with WT matches

GO biological process	FS genes with WT matches	Expected gene number	Fold Enrichment	P-value
protein complex subunit organization (GO:0071822)	134	86.93	1.54	6.22E-03
response to stress (GO:0006950)	285	221.15	1.29	3.06E-02
metabolic process (GO:0008152)	750	660.53	1.14	8.32E-03
biological_process (GO:0008150)	1243	1156.2	1.08	5.57E-07
cellular process (GO:0009987)	1071	997.19	1.07	4.34E-02

Table 6-4 Molecular function gene ontology analysis of FS genes with WT matches

GO molecular function	FS genes with WT matches	Expected Number	Fold Enrichment	P-value
identical protein binding (GO:0042802)	147	98.25	1.5	3.28 E-03
catalytic activity (GO:0003824)	473	401.01	1.18	4.90 E-02
protein binding (GO:0005515)	872	744.33	1.17	1.08 E-08
binding (GO:0005488)	1068	974.54	1.1	4.99 E-05
molecular_function (GO:0003674)	1239	1156.26	1.07	1.59 E-06

Besides gene ontology analysis, it is also important to understand how these frameshift peptide sequence of one transcript can be transformed into wildtype protein sequence of another transcript. Within these 2,033 WT matches of FS candidates, 1,682 of these WT matches (83%) and corresponding frameshift peptide sequences were from different protein variants of the same gene. One example of this category was shown in Figure 6-4. Zinc finger protein 1 had 8 different isoforms. Frameshift peptide sequence of exon 4 in isoform 7 existed in wildtype protein sequences of isoform 1. It turned out that during

alternate splicing, exon 4 of isoform 1 was spliced into the third reading frame of isoform 7, which changed the frameshift peptide sequence of isoform 7 into the wildtype sequences of isoform 1. Most the first category WT peptide matches were produced by the same mechanisms. While the rest of the 351 WT matches and corresponding FS peptides are from different genes with unknown mechanisms.

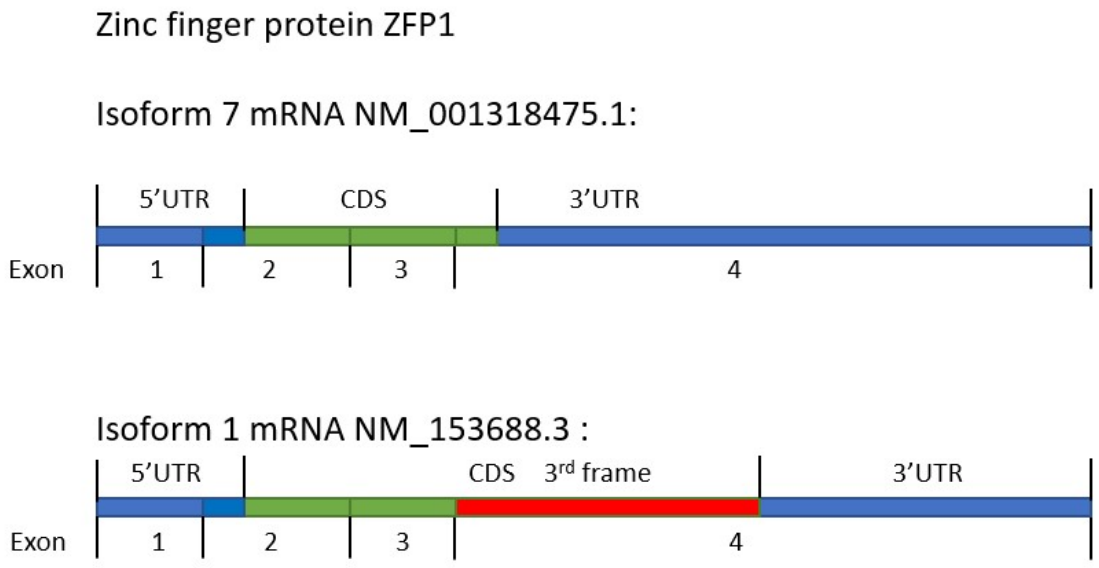


Figure 6-4. ZFP1 as an example of WT matches of frameshift peptide sequences. Blue color represents untranslated regions (5'UTR and 3'UTR), green and red color represents translated sequences, red color of isoform 1 in exon 4 is in the 3rd reading frame of isoform 7.

6.3.3 Frameshift Motifs in the120K Immunosignature Arrays

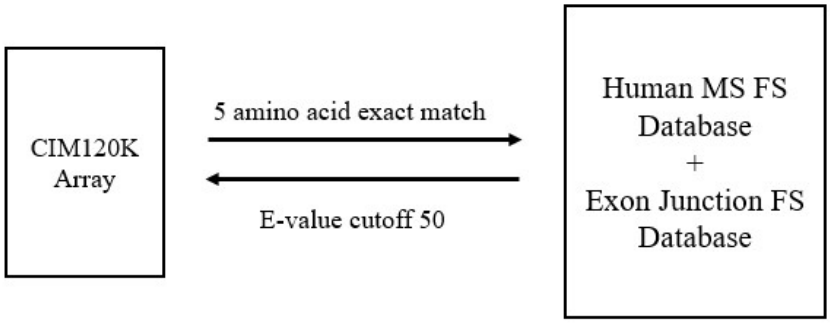
Immunosignature is a peptide microarray-based technology which measures the activity of the humoral immune system [85, 86]. There are many versions of immunosignature peptide arrays and the current version is the CIM120K peptide array. This peptide array includes ~120,000 12-mer random peptides and it is specifically designed to have high

coverage of random 5-mer peptide space for antibody recognition. Several studies indicated that the immunosignature technology could be used for cancer diagnosis in multiple cancer types [82], Alzheimer's disease diagnosis [87] and valley fever diagnosis [88] etc.

Besides the diagnosis potential of the immunosignature array, it is also interesting to 'decipher' the peptide sequences of disease associated antigens from the antibody binding activities to random peptide sequences. One of our assumptions behind the idea of using frameshift antigens for a prophylactic cancer vaccine is that frameshift transcripts from microsatellite regions or mis-splicing events are more abundant in cancer patients due to generally higher expression levels of all the transcripts and potential dysfunctional quality control systems (i.e. NMD, proteasome associated degradation) in cancer cells. Therefore, humoral immune activities of cancer patients can potentially be driven by antibody responses against frameshift antigens and frameshift antigen peptide sequences (motifs) may be in part responsible for the cancer immunosignature. If true, it would imply that specific FS motifs can be found in the immunosignature of cancer patients.

To map the frameshift peptide sequences to random peptide sequences of immunosignature arrays, the total 122,926 random 12-mer array peptides were Blasted against the human microsatellite FS database and exon junction FS database. FS-random peptide matches were recorded if they shared at least 5 identical amino acids. The e-value cutoff was set as 50 for this comparison.

A



B

Array Peptide	Array Motif	FS Antigen	FS peptide	# of identical AAs
LWQESESGVALS	LWQESESGVAL	NM_001321984.1_Exon17_3rd	LWQQSEKGVAL	8
HDYERKVGVLSG	ERKVGVLSG	NM_001099686.2_Exon4_2nd	ERKVGVLSG	9
LHRWDHLSVLSG	LHRWDHLSV	NM_001010917.2_Exon2_3rd	LQRWDHLSV	8
RQQWQEVGVASG	RQQWQEV	NM_001171713.1_Exon14_2nd	RQQWQEV	7
LQSRYFNKASAL	LQSRYFNKASAL	NM_001282704.1_Exon10_2nd	LQSRYFRKASLL	9
SAEDFRWKHDGA	EDFRWKH	NM_014680.3_Exon14_2nd	EDFRWKH	7
EWAVQPANVSLS	WAVQPANVSLS	NM_001038493.1_Exon1_3rd	WAAQPANVSFS	8
EYFFSGVGVASG	FFSGVGVAS	NM_182640.2_Exon1_2nd	FFSGVGVAS	9

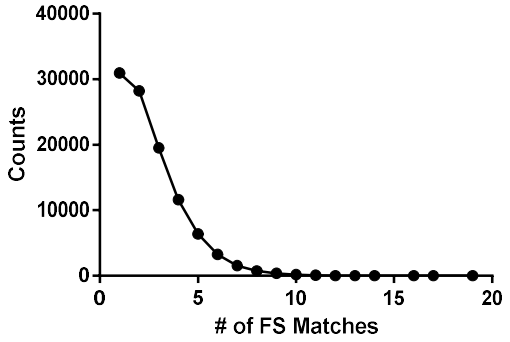
Figure 6-5. Procedure of mapping FS motifs in CIM120K immunosignature peptide array. A. Mapping diagram, Blastp-short procedure was used for comparison. B. Examples of mapping output of array peptides mapped to frameshift antigen sequences.

Out of these 122,926 12-mer CIM120K peptides, 102,916 random peptides (84%) have at least 1 FS match with the procedure in Fig6-5. There are total 267,448 FS matches from these random peptides, which indicates each array peptide has about 2.6 FS matches on average. Only 5.1% of these FS matches are from human MS FS database and the rest 94.9% FS matches are from human exon junction database. The ratio is about the same as the ratio between database sizes. The distribution of FS matches per array peptide is shown in Figure 6-6. 30% of the array peptides only have one FS match and 87.7% of these peptides have 1-4 FS matches, two array peptides have as many as 19 FS matches.

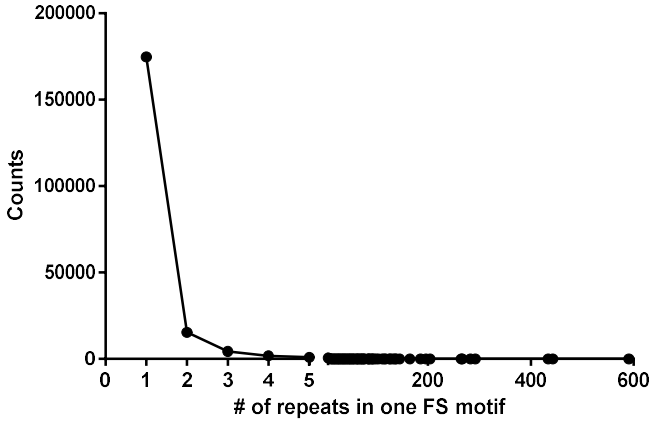
Most FS motifs (87%) appear only once in these array peptides and the number of FS motifs drops exponentially as the repeats number increase, some FS motifs have over 200 repeats and they are summarized in Figure 6-6.

With these FS motifs of array peptides available, it is possible to find the potential FS antigens behind the antibody-random peptide reactivity. The next step is to test human cancer patient serum samples on the CIM120K array and find any cancer specific array peptides and then decipher the antibody signals to the original FS antigens.

A



B



C

FS Motif	# of repeats
SGALSG	591
LGVLSG	443
LSGVAL	434
LSGALS	292
ASGVAL	283
ALGVAL	267
LSVALG	265
LSGASG	205
LGALSG	204

Figure 6-6. Distribution of FS matches of CIM120K array peptides and FS motifs. A. Distribution of FS matches per array peptide. B. Distribution of FS motif repeats. C. List of FS motifs which have over 200 repeats in Blast outputs.

6.3.4 Anti-Frameshift Motif immune response in Breast Cancer Patients

To characterize humoral immune activities of breast cancer patients, a cohort of 949 samples were tested on CIM120K peptide arrays by the Peptide Array Core in CIM. These samples included 100 DCIS (ductal carcinoma in situ) samples from Abcodia, 399 breast cancer Stage I samples from Abcodia and Duke University, 100 benign samples from Duke University and 350 healthy control samples from various source (Duke, Abcodia and ASU). Benign breast tumors (lumps) are very common among women and it is a non-cancerous condition. In this chapter both benign and healthy control samples are treated as control group. DCIS is usually considered as the earliest stage of breast cancer and it is non-invasive [89].

It is known that the host immune system can respond to a tumor and both adaptive and innate immune responses to tumor are detected [10]. The theory of immunoediting is proposed based on studies in mouse models. It includes three phases of host-tumor interactions: host immune response kills highly immunogenic tumor cells in the elimination phase; if not, the tumor then becomes dormant and poorly immunogenic tumor cells replace highly immunogenic cells in equilibrium phase; finally, the tumor will overcome the host immune system by creating an immunosuppression microenvironment and consistently removing immunogenic tumor cells in the escape phase [90]. However, these studies only uncover the relation between cell mediated

immune responses and the tumor. The interaction between the humoral immune system and the tumor has not been investigated. And it is also not known which clinical stages are corresponding to these three phases. Therefore, it is interesting to know how the antibody profiles of cancer patients change during cancer progression and if the frameshift antigens are involved in the antibody response.

Three stages of breast cancer were included in our dataset: Benign and healthy control; DCIS (Stage 0); Stage I. Antibody profiles against random peptides change over each stage. How these antibody signals change in each stage and the potential frameshift antigens behind these signals are crucial for both tumor immunology and cancer vaccine antigen selection. Here my analysis was focused on three patterns of antibody response during the three stages. Each pattern might uncover a unique process of tumor-host immune reaction:

1. Antibody response increased in each stage: peptide signal went higher from healthy control to DCIS by 1.5-fold and from DCIS to Stage I by another 1.5-fold;

2. Antibody response increased from healthy control to DCIS, and then decreased from DCIS to Stage I;

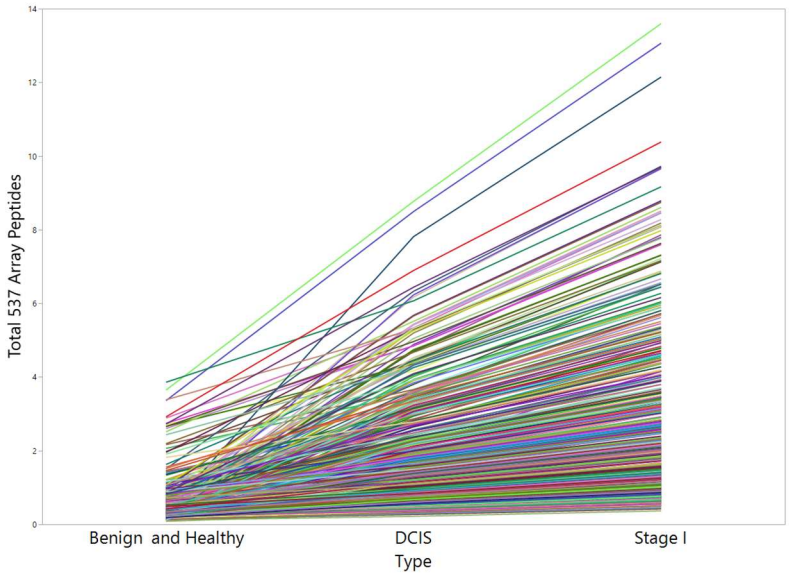
3. Antibody response decreased from healthy control to DCIS and Stage I.

Pattern 1: The first pattern was that antibody response against some array peptides increased during each stage, which indicated that the antibody response against a portion of antigens was boosted during each stage and it was not suppressed by the

immunoediting process. The average of normalized fluorescence intensity of all samples in each stage was calculated, and array peptides with an intensity which increased 1.5-fold from healthy control to DCIS and another 1.5-fold from DCIS to Stage I were selected. A total 537 array peptides were selected with this pattern (Figure 6-7). Unsupervised hierarchical clustering results showed that almost all healthy and benign samples had low reactivity to these peptides, while the majority of Stage I samples and 1/3 of the DCIS samples had high reactivity. Another 2/3 of the DCIS samples had low reactivity to these peptides. These results indicated that antibody response against these 537 peptides was boosted in a portion of DCIS patients (30%) but it was further boosted in most of Stage I patients (80%). 20% of the Stage I patients did not have an antibody response against these peptides, which meant that 20% of the population might not present the antigens behind these 537 peptides and they would not benefit from the vaccine formulated by the FS antigens predicted with these 537 peptides.

Even though the FS antigens behind these 537 peptides might not cover the whole population, it would be interesting to know which FS antigens could be related to these 537 peptides. A portion of the related FS antigens with best E-value (high homology between array peptides and FS antigens) were summarized in Table 6-5.

A


Total 537 Array Peptides
14
12
10
8
6
4
2
0
Benign and Healthy
DCIS
Stage I
Type

B

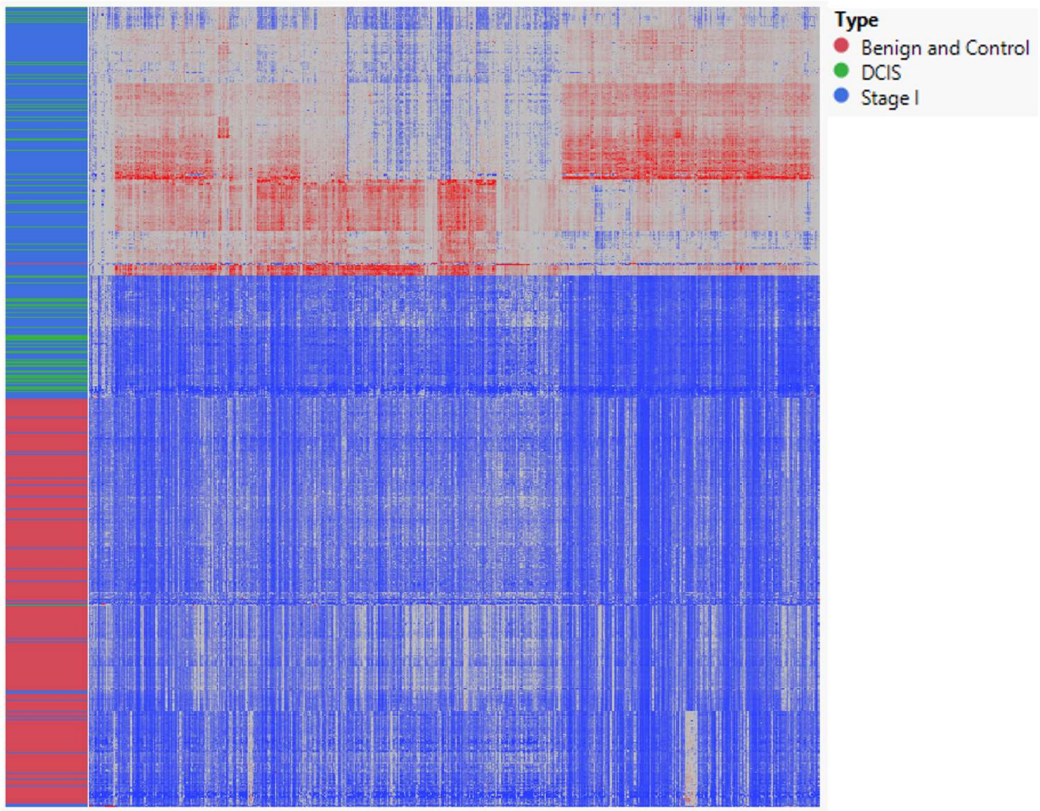

Type
Benign and Control
DCIS
Stage I

Figure 6-7. The first pattern of antibody profile changes in breast cancer stages: antibody response increased 1.5-fold from healthy control to DCIS and 1.5-fold from DCIS to Stage I. A. The average of peptide signal in each stage was plotted against each stage. B. Hierarchical clustering with 537 selected peptides.

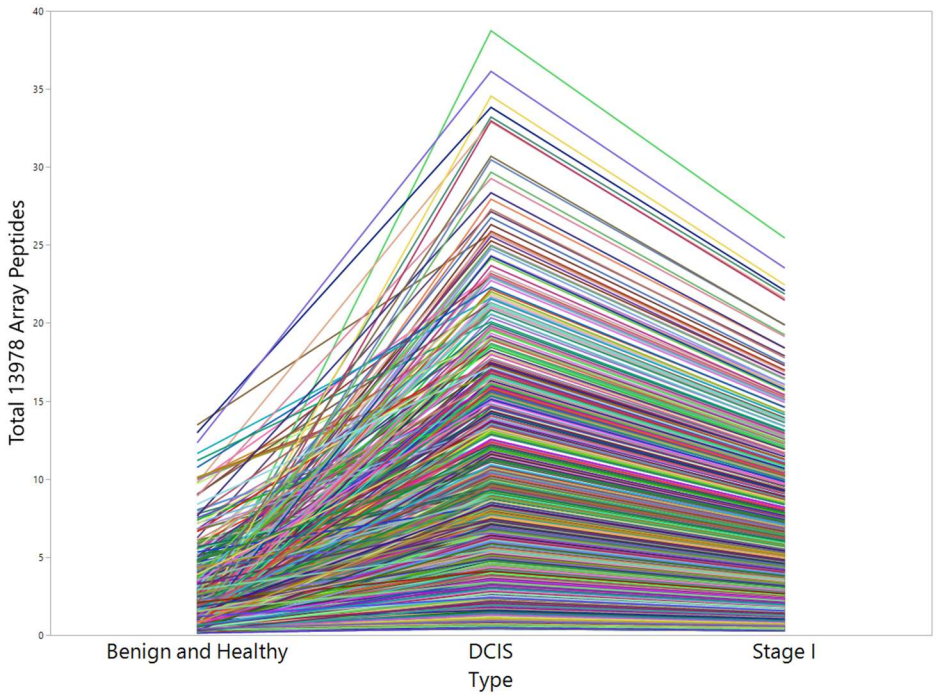
Table 6-5 Best FS antigens from Pattern 1 by E-value

Array Peptide	Array Motif	FS Antigen	Gene	FS Motif	E-Value
KEWNEQRWVLLS	WNEQRW	NM_003041.3_Exon14_3rd	SLC5A2	WNEQRW	0.027
QSREREALFVAL	REREALF	NM_001098787.1_Exon4_3rd	BET1L	REREALF	0.15
DYEQWHVALGAL	EQWHVAL	NM_001300815.1_Exon7_3rd	CIRBP	EEWHVAL	0.25
SNDKRPVLVALG	RPVLVALG	NM_004386.2_Exon8_3rd	NCAN	RPVLVPLG	0.43
DGFQEYAQYVLG	GFQEYAQYVL	NM_001286721.1_Exon5_2nd	ANKRD10	GFQECAQFLL	0.66
WVAKQEFKVLLG	QEFK--VLLG	NM_001042583.2_Exon6_3rd	CD1E	QEFKTSVLLG	0.72
DKEPPVLVALSG	PVLVALSG	NM_001282471.1_Exon3_3rd	PRORY	PVLIALSG	0.75
SEWPQRYHVLVL	EWPQR-YH	NM_133264.4_Exon6_3rd	WIPF2	EWPQRFYH	0.82
DKVWLHVLGVAS	WLHVLGVAS	NM_001276254.2_Exon5_2nd	IFNL4	WLHTLGLAS	0.94
DPSSRNHDVLVL	RNHDVLVL	NM_001290047.1_Exon17_2nd	CECR2	RNHQVLVL	1

Pattern 2: The second pattern referred to an antibody response which increased from healthy control to DCIS, but then decreased from DCIS to Stage I. This pattern might indicate that the antigens which induced this antibody response were selected against in Stage I and tumor cells presenting these antigens were removed by the immunoediting process. If this assumption is true, the potential FS antigens behind these array peptides are highly immunogenic and perfect vaccine antigens for early or prophylactic cancer vaccine development. Peptides which had 1.5-fold higher signal in DCIS group than healthy control group and Stage I group were selected for analysis. A total number of 13,979 peptides passed the criteria and the peptide number increased 26-fold compared to pattern 1 which only had 537 peptides (Figure 6-8), which indicated a lot more antigens

might have gone through immunoediting from DCIS to Stage I. The result was consistent with other immunosignature studies with multiple cancer stages (unpublished data) – the humoral immune system was mostly suppressed in the early stages of breast cancer. The hierarchical clustering result showed that all DCIS samples had medium to high reactivity to these 13,979 peptides, a small portion of Stage I samples were mixed with DCIS samples and had high reactivity, while most of Stage I samples were suppressed and only had medium reactivity to these peptides. Healthy control samples and benign samples had much lower reactivity to these peptides than the other two groups. A group of Stage I samples had very low reactivity within Stage I samples (Figure 6-8B) and they were separated from other samples. 80% of these low reactivity samples were from Abcodia and the cause might be related with sample qualities or other unknown reasons.

A



B

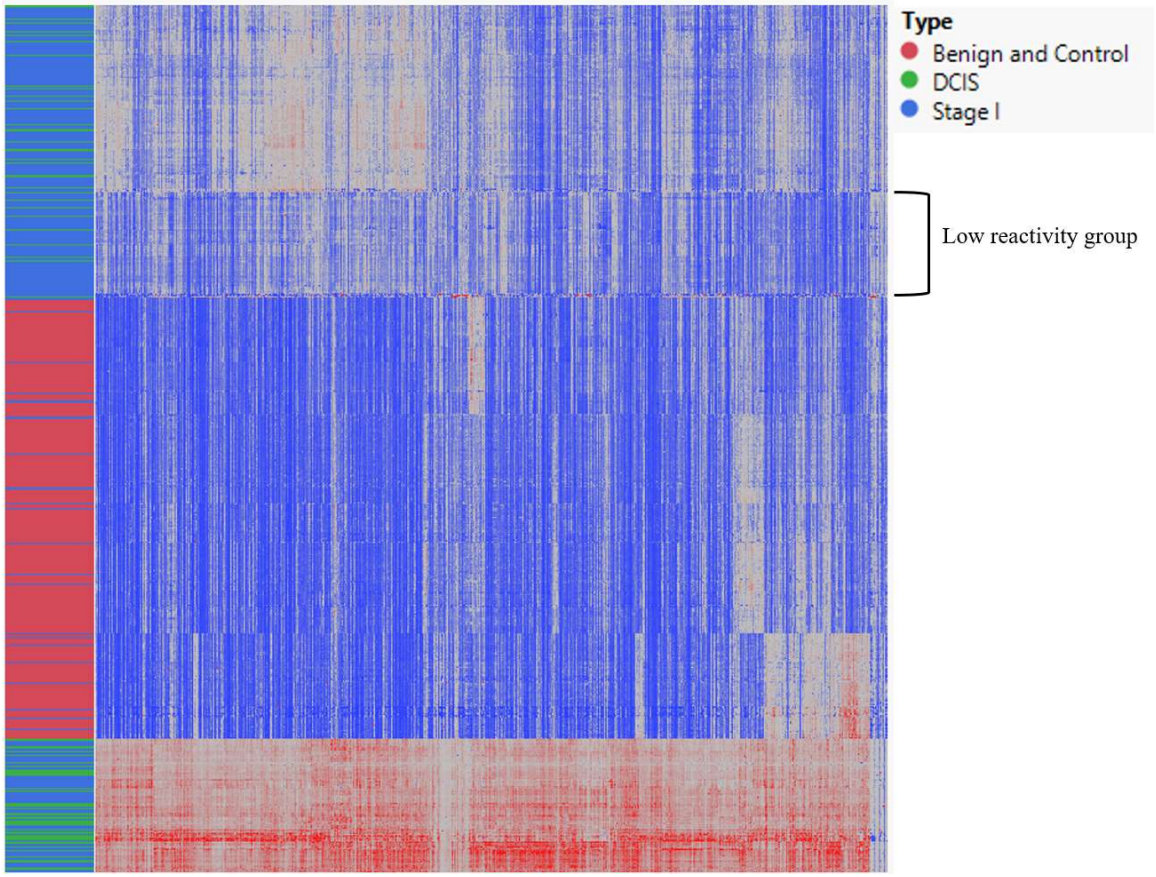


Figure 6-8. The second pattern of antibody profile changes in breast cancer stages: antibody response increased 1.5-fold from healthy control to DCIS and then decreased 1.5-fold from DCIS to Stage I. A. The average of peptide signal in each stage was plotted against each stage. B. Hierarchical clustering with 13,979 selected peptides.

Potential FS antigens from the second pattern might be good vaccine antigens as mentioned previously. Since the number of peptides was increased 26-fold compared to pattern 1, more corresponding frameshift antigens with conserved sequences could be found. Table 6-6 summarized a small portion of the potential FS antigens sorted with best

e-values. The top FS antigens could have 8 to 9 exact amino acids matches to 12-mer array peptides.

Table 6-6 Best FS antigens from Pattern 2 by E-value

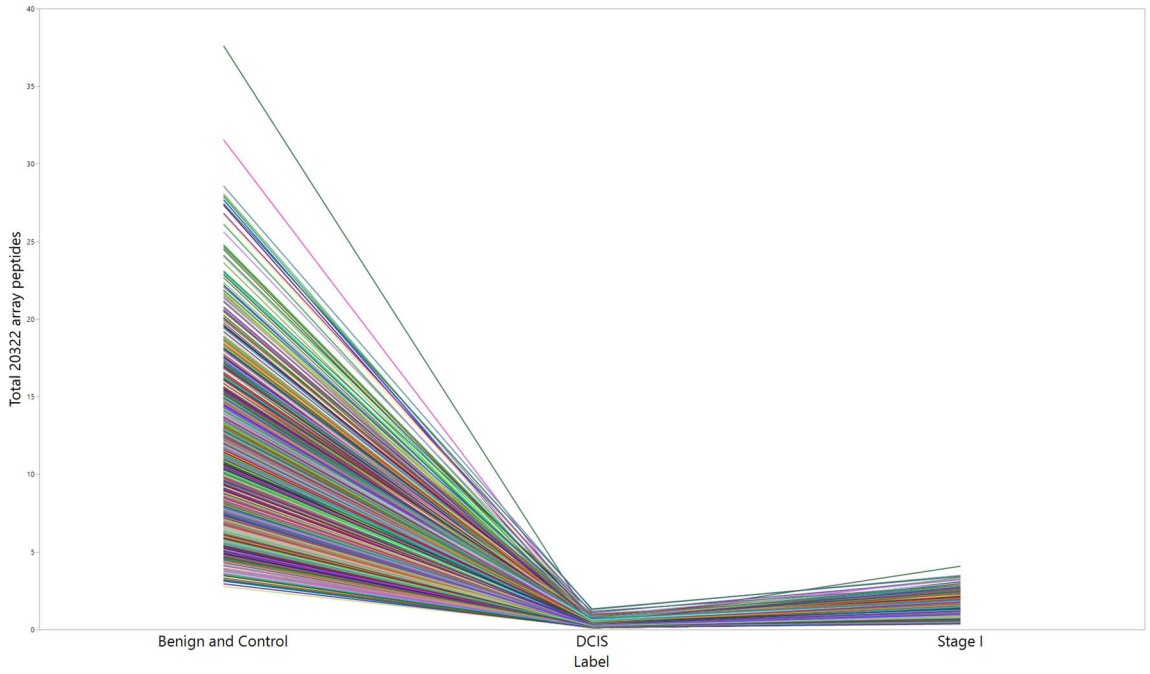
Array Peptide	Array Motif	FS Antigen	FS Motif	Gene	E-value
LQSRYFNKASAL	LQSRYFNKASAL	NM_001282704.1_Exon10_2nd	LQSRYFRKASLL	PDIA6	0.007
YNQNRVLWALAG	NQNRVLWAL	NM_001135729.1_Exon6_2nd	NQDRILWAL	TOM1	0.018
GDGHLALSVALG	GHLALSVALG	NM_001303427.1_Exon1_2nd	GHLGLSVALG	LOC110117498-PIK3R3	0.019
VHDSVWDSGVLS	SVWDSGVL	NM_001852.3_Exon32_2nd	SVWDSGVL	COL9A2	0.021
QWKGRSYDVFSV	QWKGRSYD	CCDS6315.1|Hs104|chr8	QWKGRRYD	CSMD3	0.026
PVWLVSFPNHDA	VWLVSFP	NM_001136.4_Exon3_2nd	VWLVSFP	AGER	0.035
KVSRREPWAYDS	VS---RREPWAYDS	NM_001080976.1_Exon2_3rd	VSLHHRREPRSYDS	DSE	0.04
NAAYDAFYWLSL	DAFYWLS	NM_003049.3_Exon3_3rd	DAFYWLS	SLC10A1	0.042
EEYVYENASVSG	VYENASVSG	NM_000712.3_Exon7_3rd	VYENDSVSG	BLVRA	0.05
SKQHFWALVLSG	HFWALVL	NM_001320373.1_Exon17_2nd	HFWALVL	VWA5B2	0.052
KAHKHSALSVLS	HSALSVLS	NM_001146032.1_Exon25_2nd	HSALSVLS	FCHO2	0.061
PASFNLFLQENL	PA----SFNLFLQ	NM_001197216.2_Exon8_2nd	PARSHLSFNLFLQ	ASGR1	0.062
RHFAFVGKASLG	FVGKASLG	NM_020817.1_Exon10_3rd	FVGKASLG	CCDC191	0.062
LWQKKSGVALSG	WQKKSGVALSG	NM_207361.5_Exon24_3rd	WQKKCGVPFSG	FREM2	0.068
ALNHPNQFASGS	HPNQFASG	NM_018190.3_Exon7_2nd	HPNQYASG	BBS7	0.083
KNAPFRALSALS	NAPFRALSAL	NM_001127715.2_Exon1_3rd	NAPVRALSAL	STXBP5	0.091
AVGFLEASGALG	VGFLEASG	NM_001206615.1_Exon3_2nd	VGFLEASG	EHF	0.092
NQHHYDALVLSG	HHYDALVL	NM_015496.4_Exon23_2nd	HHYDPLVL	VIRMA	0.093
NVLNWVFRNASG	NVLNWVF	NM_018656.2_Exon5_2nd	NVLNWVF	SLC35E3	0.095
SRQRWLNSVLSG	SRQR---WLNSVL	NM_015033.2_Exon17_3rd	SRQRPPPWLGSVL	FNBP1	0.098
LWRNYALSVLSG	LWRNYALSVLSG	NM_001167917.1_Exon4_2nd	LWRSY-LCVLSG	VEPH1	0.1
NRHVLLFPAVLG	LLFPAVLG	NM_001001670.2_Exon4_2nd	LLFPAVLG	SPATA31D1	0.1
ANSVDALSVALG	SVDALSVAL	NM_017509.3_Exon2_2nd	SVDALTVAL	KLK15	0.11
SRHRVFRAGVSG	SRHRVFRAGVS	NM_001145415.1_Exon15_2nd	SRHRT-RAGVS	SETDB1	0.12
QEPDVHLPWLFS	QEPDVHLP	NM_001142646.2_Exon9_2nd	EEPDVHLP	TPRA1	0.12
GKVWLKDSGVLG	WLKDSGVLG	NM_001280547.1_Exon5_3rd	WLRDAGVLG	PAX5	0.12
QFLLWLSREKGV	FLLWLSREKG	NM_181501.1_Exon17_3rd	FFLWNSREKG	ITGA1	0.12
AKFRLWAGVALG	RLWAGVAL	NM_001271765.1_Exon7_2nd	RLWAGIAL	SLC16A5	0.13
FPAQVFVPWDLS	QVFVPWDL	NM_001161580.1_Exon10_3rd	QGFVPWDL	POC1A	0.13
AKQKQYDASVAS	KQKQYDASV	NM_005422.2_Exon9_3rd	KQKQYDSGV	TECTA	0.13
PEARNGEYHAVL	EARNGEY	NM_017794.4_Exon22_3rd	EARNGEY	FOCAD	0.13
FAEQPYFWASVL	EQPYFW	NM_014289.3_Exon8_3rd	EQPYFW	CAPN6	0.14
KDLLGRAENASL	LLGRAENA	NM_015323.4_Exon14_2nd	LLGRAENA	UFL1	0.14

Pattern 3: the third pattern was the dominant pattern and it referred to certain antibody responses which decreased dramatically with tumor initiation (decreased from healthy control to DCIS and Stage I). There were many potential causes for this phenomenon and two which I had thought of: 1. The immune system was suppressed at very early stages of tumor development (before DCIS) and this dominant pattern was due to non-specific immunosuppression; 2. Healthy people which had immune responses against these specific antigens would not develop a tumor because this specific immune response had strong anti-tumor effects. If it was caused by non-specific immunosuppression, then the array peptides from this pattern would not help with cancer vaccine development; 3. This pattern could be caused by the maturation process of anti-FS antibodies, the early forms of these antibodies could have low specificity but later they matured into highly specific antibodies targeting other FS peptides shown in pattern 1 or pattern 2. However, if it was caused by strong anti-tumor effect in healthy people, then the potential FS antigens from these peptides might be good source for a prophylactic cancer vaccine.

Since a large portion of the whole 120K peptides were suppressed from healthy people to cancer patients, stricter selection criteria were applied in the analysis. Peptides with signals which had dropped 3-fold from healthy control to DCIS and Stage I were selected for further analysis. A total number of 20,322 peptides passed the criteria, the signals for these peptides dropped significantly from benign and control samples to DCIS, and then increased slightly from DCIS to Stage I which still had much lower signals than the control group (Figure 6-9). Potential FS antigens behind these peptides were listed in Table 6-6 which were sorted by e-values.

Frameshift antigens from the three patterns shared between each other due to repeat motifs in different array peptides. A Venn diagram in Figure 6-10 showed that more than 60% of predicted FS antigens in pattern 1 were shared in pattern 2 or 3, which was mainly caused by certain repetitive FS motifs in many array peptides as shown in Figure 6-6.

A



B

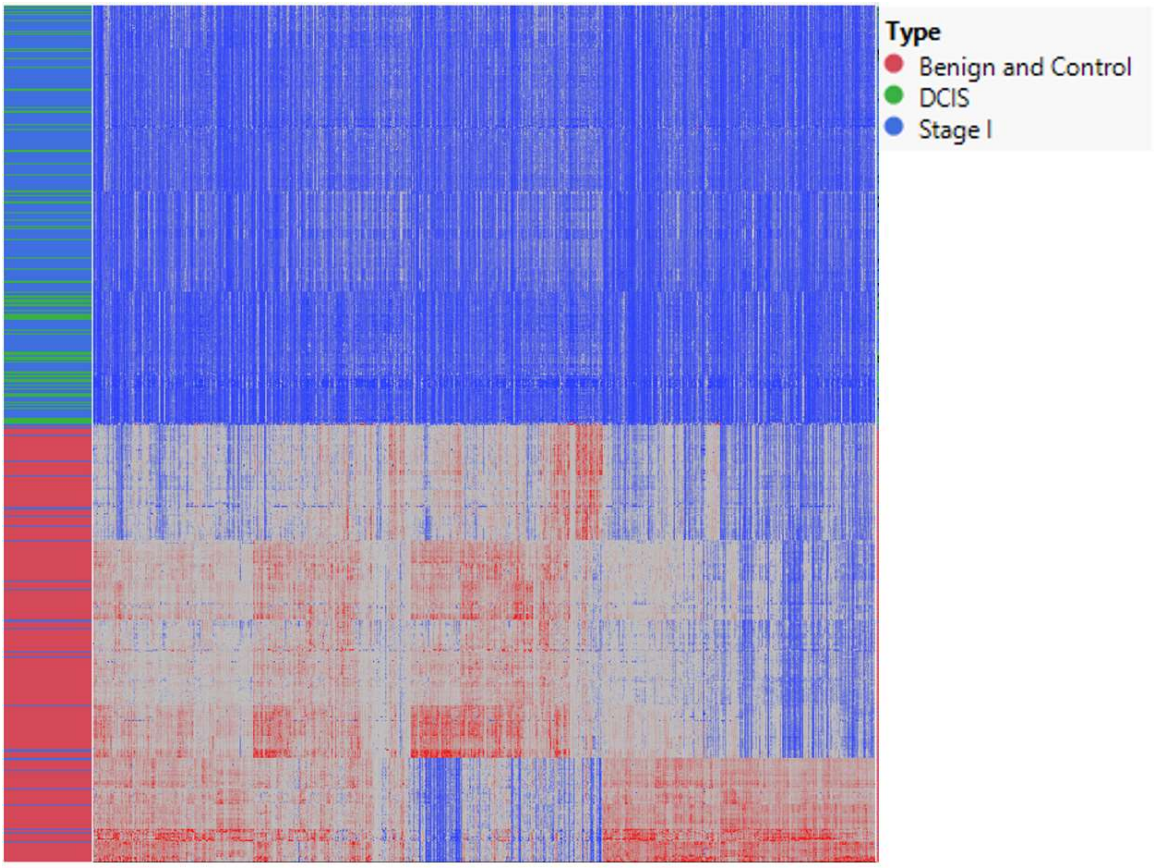


Figure 6-9. The third pattern of antibody profile changes in breast cancer stages: antibody response decreased 3-fold from healthy control to DCIS and Stage I. A. The average of peptide signal in each stage was plotted against each stage. B. Hierarchical clustering with 20,322 selected peptides.

Table 6-6 Best FS antigens from Pattern 3 by E-value

Array Peptide	Array Motif	FS Antigen	FS Motif	Gene	E-value
LWQESESGVALS	LWQESESGVAL	NM_001321984.1_Exon17_3rd	LWQQSEKGVAL	ENTPD5	0.002
DAPSELHLSGSG	PSELHLSGS	NM_182584.2_Exon5_3rd	PSELHLSGS	C20orf203	0.014
WLRAEHALEHSV	WLRAEHALEHS	NM_001261442.1_Exon4_2nd	WHRAEHALQRS	EXTL2	0.028
AQYDPAPNHVAL	AQYDPA-PNHVAL	NM_001286643.1_Exon9_3rd	AQYDPAKPNYISL	MTERF3	0.031
WNYRHAPYKGVL	WNYRHAP	NM_001320458.1_Exon5_2nd	WDYRHAP	WNT9B	0.038
RSLQKWHLGVSG	RSL-QKWHLGV	CCDS6160.1|Hs104|chr8	RNLKQKWHLGI	RP1	0.044
NSGLEARKHLSS	LEARKHLSS	NM_001105069.1_Exon5_2nd	LEARKHLPS	ACSM2B	0.048
YFPGHQPHDSAL	GHQPHDSA	NM_001317924.1_Exon30_3rd	GHQPHDSA	WDR19	0.051
FEWRDVLHGQKD	EWRDVLHGQK	NM_001160002.1_Exon5_3rd	EWRGVLHGER	NRG1	0.052
EQQGEKGVALSG	QQGEKGVAL	NM_001321984.1_Exon17_3rd	QQSEKGVAL	ENTPD5	0.06
RDAYRPLWHLVA	RPLWHLV	NM_001036.4_Exon36_2nd	RPLWHLV	RYR3	0.065
RYANFRHSNDLG	NFRHSND	NM_001004477.1_Exon1_3rd	NFRHSND	OR10X1	0.066
FGAARPVALGLS	GAARPVALGL	NM_001321831.1_Exon12_3rd	GAARPVLLGL	NFKBID	0.074
EFWNGFRKEDGS	FWNGFR	NM_001281765.2_Exon5_2nd	FWNGFR	EPHA5	0.09
FSLGFYGQNKHV	FSLGFYG	NM_031925.2_Exon12_3rd	FSLGFYG	TMEM120A	0.092
NLENGDALSRNA	GDALSRNA	NM_000350.2_Exon3_3rd	GDALSRNA	ABCA4	0.098
EDAPSKRWNSGS	KRWNSGS	NM_022470.3_Exon6_2nd	KRWNSGS	ZMAT3	0.1
DGFFPSQRDSAS	FPSQRDS	NM_005422.2_Exon13_3rd	FPSQRDS	TECTA	0.1
DKHQEHVRVASS	DKHQEHVRVASS	NM_012392.3_Exon2_2nd	DKHQEPLRVATT	PEF1	0.11
YRPSVEPYNDSG	YRPSVEPYNDSG	NM_023935.2_Exon2_2nd	YRPSVEP---SG	DDRGK1	0.11
LSPGEYAGFPHD	LSPGEYAGF	NM_024927.4_Exon1_2nd	LSPGDYGGF	PLEKHH3	0.11
YFHQRWKSNVSG	YFHQRW	NM_001277269.1_Exon25_3rd	YFHQRW	OTOG	0.11
FWFRYGESQEKV	FRY-GESQEKV	NM_001089.2_Exon33_3rd	FRYSGESQGKV	ABCA3	0.11
KAPHKVSEKSGS	KAPHKVSEKSGS	NM_001199692.1_Exon7_2nd	KVPHRVAEKGGS	SLC4A2	0.11
WGRVNLQLSGAS	WGRVNLQLSG	NM_001164473.2_Exon17_3rd	WGRV-LELSG	FNBP1L	0.12

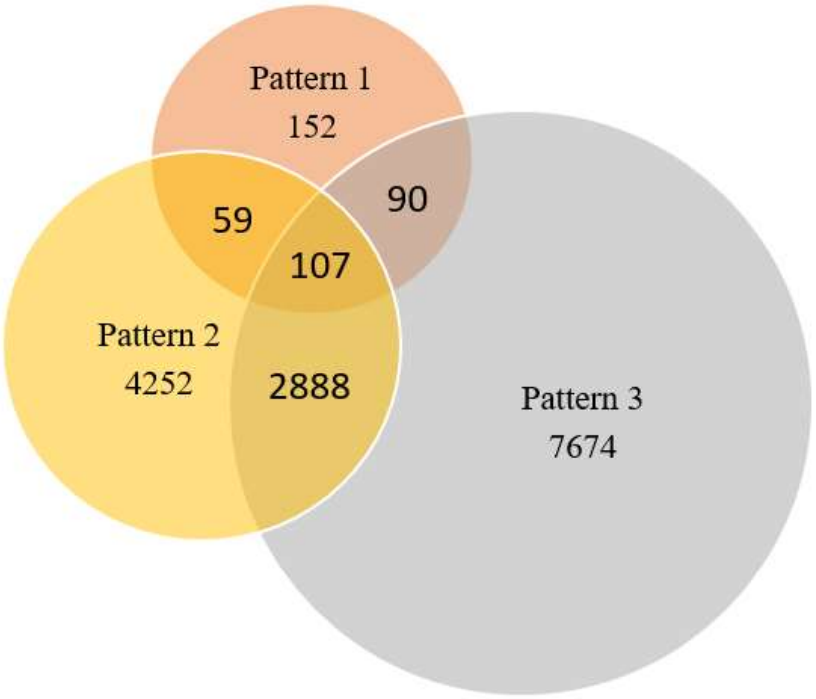


Figure 6-10. Venn Diagram of Predicted FS antigens from Pattern 1, 2 and 3.

6.3.5 Human 400K Frameshift Peptide Array

The method of finding FS motifs from immunosignature data was promising in predicting potential FS antigens potentially eliciting the antibody response. However, this method used only partial information (5-9 mer motifs) of the 12-mer random peptides in predicting the FS antigens and this prediction might not be accurate. Further validation experiments need to be conducted to prove the anti-FS response. To measure direct anti-FS humoral immune response in cancer patients, an unbiased frameshift peptide array which covers all possible microsatellite and exon junction FS peptides is needed.

It is possible to synthesize millions of peptides with in situ photolithographic synthesis technology. The 12-plex format slide has the capacity of including 392,000 15-mer peptides per array, which is sufficient to cover the whole human MS and exon junction FS peptides.

The design of the human 400K frameshift peptide array includes all 14,588 human MS FS peptides, and all human exon junction frameshift peptides with peptide length ranges from 7-100aa. FS candidates with 100 amino acids or more are excluded from the array because they may be functional proteins.

The design of human 400K frameshift peptide is completed and this array is currently under synthesis and analysis

6.4 Discussion

In this chapter, I created the human exon junction frameshift database which included all possible frameshift peptides from every human exon. The analysis of FS peptide length and amino acid composition indicated that exon junction FS peptides were not under high selection pressure and they were close to random distribution. A group of FS peptides had identical sequences in wildtype protein variants of the same gene which was caused by alternate splicing of the last exon. In addition, I established a method of searching for potential FS antigens from immunosignature arrays. An immunosignature data set with different stages of breast cancer samples was used to find potential FS antigens against

three different patterns. Furthermore, a human 400K FS peptide array which covers all possible FS peptides from MS and exon junctions is under development.

Frameshift antigens were rare in cancer patients at the genomic level, but recently we found microsatellite frameshift mutations and mis-splicing frameshift mutations from EST tumor libraries were very common in the mRNA of tumor tissues and tumor cell lines. It was difficult to detect frameshift peptides in tumor tissues by mass spectrometry, but the human humoral immune system could amplify signals of frameshift antigens if these FS antigens were expressed and presented. Then we could detect antibodies against these FS antigens. The development of the human 400K FS peptide array was based on this idea and with this platform we can screen anti-FS immune responses across cancer patients unbiasedly. Highly immunogenic FS antigens which have not yet been discovered might be revealed in this platform.

The Immunosignature technology maximizes the capability of recognizing all kinds of antibodies by increasing the 5-mer diversity and this platform contains a huge number of motifs. It could successfully separate different kinds of cancer types and the tumor specific immune response might be caused by tumor specific FS neo-antigens. It was possible to transform the random peptide array data to anti-FS data since there were a large number of FS motifs in these random peptides. However, this method only used a portion of the 12-mer peptide sequences for the comparison. The antibody might not recognize the FS motif per se when it is bound to the corresponding 12-mer peptide. So, there were a considerable number of false positives in the analysis and further validation experiments are needed.

In summary, the human exon junction FS peptides are potential source for cancer vaccine development and this human 400K FS array can be a great platform for cancer diagnosis, vaccine discovery

6.5 Methods

6.5.1 Human Exon Junction Frameshift Database

Human mRNA sequences and exon coordinates were downloaded from NCBI Refseq Database [52]. mRNA sequences of each exon were calculated by exon coordinates, then the second frame and third frame of these mRNA sequences were translated into frameshift peptides.

6.5.2 Database Characteristics Analysis

Frameshift peptide length and amino acid compositions of the database were calculated by the FS peptides. Reference amino acid composition was calculated by WT human RefSeq proteome. Unsupervised two-way hierarchical clustering was conducted with JMP Pro 13 (Cary, NC).

6.5.3 Human Breast Cancer Immunosignature Dataset

A total of 949 samples with healthy control, benign breast tumor, DCIS and Stage I breast samples were collected from Duke, Abcodia and ASU. Samples were tested on the CIM120K array and standard immunosignature assay protocols were used [82]. Data was normalized to the median of each array.

6.4.4 Find FS motifs in CIM120K Array and Breast Cancer IMS data analysis

120K 12-mer random array peptides were compared to the FS database via BLAST [55]. The E-value cutoff was set as 50 and FS matches with 5 amino acids exact match or more were included. Samples in the breast cancer dataset were separated into three groups: healthy control/benign; DCIS; Stage I. JMP Pro 13 was used for two-way clustering.

CHAPTER 7

Early Treatment of Breast Cancer with Checkpoint Inhibitors in the neuN Transgenic Mouse Model

7.1 Abstract

Immune checkpoint blockade was approved initially for treating late stage cancer patients when other treatments failed. Adverse events were common in these treatments and only a proportion of patients would benefit. Here we proposed the idea of early immune checkpoint inhibitor (ICI) treatment at early cancer diagnosis by immunosignatures. We tested this idea in the mouse FVB/N neuN transgenic mammary tumor model. Results demonstrated that very early ICI treatments delayed tumor initiation significantly. Furthermore, timing was shown to be an important factor in ICI treatment as early treatment (8 weeks after the earliest detection time point) gave the best protection in this model. Additionally, immunosignatures were used to separate responders from non-responders in the ICI treatments. Remarkably, non-treatment mice presenting a responder signature also had delayed tumor initiation which indicates a pre-determined anti-tumor immune status could predict the tumor progression. With these results, a model for early immunotherapy was summarized to predict best ICI treatment timepoint.

7.2 Introduction

Immune checkpoint blockade (ICB) has gained great success in treating late stage cancer patients in multiple cancer types [91-93]. At least 6 monoclonal antibodies targeting

PD1/PDL-1/CTLA-4 have been approved by the FDA. However, large doses of immune checkpoint inhibitors cause a broad spectrum of toxicities in many patients and autoimmune diseases in a portion of patients. Persistent immune related AEs and deaths have also been reported [94-96]. In the meantime, ICB treatment was not effective in all late stage patients. A durable objective response to immune checkpoint blockade ranges from 20% to 50% in different cancer types [93, 97-103]. To improve the current situation, we proposed the idea of treating cancer patients much earlier with less immune checkpoint inhibitors. Less ICI could be used with less tumor burden at much earlier stages. This approach could be more effective because immunosuppression was not so overwhelming in early cancer patients as in late stage patients. This approach was only feasible with the Immunosignature technology which could diagnose cancer much earlier than current diagnostics. The NeuN transgenic mouse breast cancer model was used to test this idea in this chapter.

Immunosignature technology uses random peptide arrays to profile the humoral immune system of individuals. Details of this technology were reviewed in Chapter 6. Immunosignatures were also used for early cancer diagnosis in the FVB/N neuN transgenic mouse model. The technology could diagnose early tumorigenesis as far as 12 weeks before the first palpable tumor [104].

In the meantime, since only a proportion of patients respond to ICI treatment and the treatment cost is very high it would be beneficial to have a biomarker that could distinguish who would respond. Biomarkers to predict responders vs. non-responders are important for reducing the cost and improving the efficiency. There are three available

biomarkers: PDL1 expression in tumor tissues; tumor mutational load and CD8+ T cells tumor infiltration [105, 106]. These three biomarkers are functionally related to each other, and they cannot separate responders perfectly from non-responders. For example, the response rate was 48% in PDL1+ patients and 15% in PDL1- patients [107]. The difference is significant but not sufficient to be used as a biomarker in the clinic. The immunosignature platform may have advantages for biomarker discovery with 120,000 peptide features that unbiasedly bind antibodies. We are testing this platform in predicting responders vs. non-responders in this the FVB/N neuN mouse model.

In this chapter, I compared three different treatment timepoints with the combination treatment of anti-PDL1 and anti-CTLA4 mAbs: very early treatment (16 weeks); early treatment (24-26 weeks); late treatment (first palpable tumor). Furthermore, immunosignatures were used for predicting responders vs. non-responders to ICI with serum samples prior to the treatment.

7.3 Results

7.3.1 FVB/N neuN Transgenic Mouse Breast Cancer Model and Breast Cancer Early detection

The FVB/N neuN transgenic mouse model was established in 1992 by William Muller. The ERBB2 (also known as HER2) gene was overexpressed under the mouse mammary tumor virus promotor (MMTV) [108, 109]. Overexpression of HER2 gene induces mammary tumors and metastasis after a long latency. The tumor free curve within a group of 28 transgenic mice is shown in Figure 7-1. The median age for the first palpable

tumor in this mouse model was 31.7 weeks, with 90% of these mice having the first palpable tumor before 40 weeks.

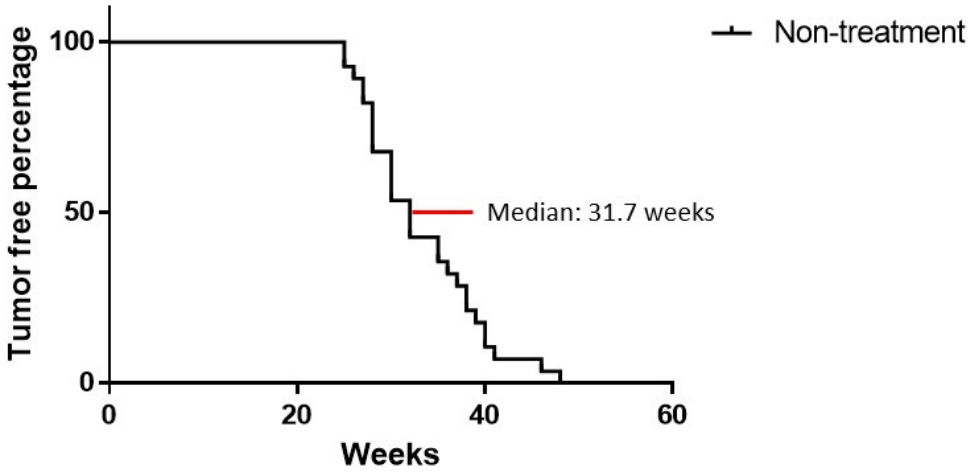
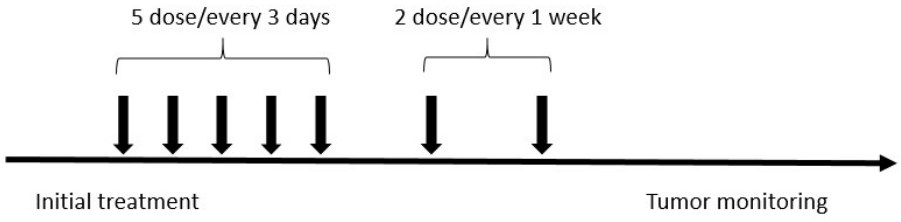


Figure 7-1. Tumor free curve and median age for first palpable tumor in the FVB/N neuN transgenic mouse model.

Though the first palpable tumor was detected at 31.7 weeks, the humoral immune system of these transgenic mice might have changed way before the palpable tumor was evident. Immunosignature arrays (CIM 10K version 2) were used previously for early detection of breast tumors in this model. This report showed that transgenic mice could not be separated from the wildtype mice at 12 weeks, but these transgenic mice had significantly different immunosignatures compared to wildtype mice at 16 weeks, which indicated that the earliest detectable time point was between 12 to 16 weeks [104]. This time point was 15 weeks on average before the first palpable tumor.

7.3.2 Very early ICI treatment delayed the first palpable tumor while early treatment boosted the tumor protection

The earliest time point for cancer detection was at 16 weeks. To compare early treatment versus the current strategy (treatment at late stage for cancer patients), four groups of mice were included in the experiment: a very early treatment group in which mice were treated at the earliest detection time point (16 weeks); an early treatment group at 24 to 26 weeks which allowed the early tumor to grow for several weeks; a late treatment group which were treated after the 1st palpable tumor and a non-treatment group with no treatments. The initial treatment date was different between groups but the treatment schedule was the same as shown in Figure 7-2.

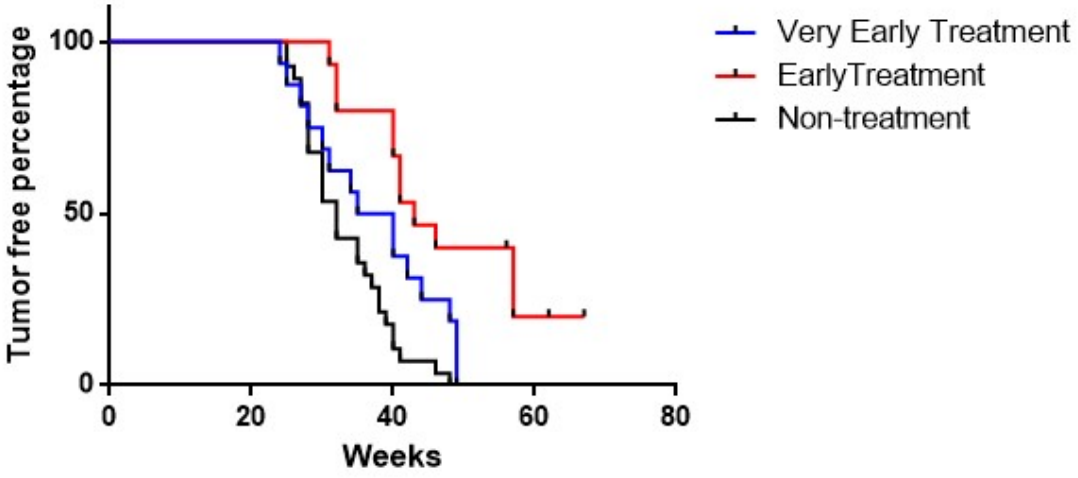


Groups	Initial Treatment
Non-treatment Group	No treatment
Very Early Treatment Group	Week 16
Early Treatment Group	Week 24-26
Late Treatment Group	Treat after first palpable tumor, around week 33

Figure 7-2. Three different ICI treatment regimens in the FVB/N neuN mouse model, 7 doses of ICI treatments were injected via IP. Each dose of ICI treatment included 100 μg anti-CTLA4 mAb and 200 μg anti-PDL1 mAb.

Tumor development was monitored weekly and tumor initiation age was recorded for each mouse. Treatment at the Very Early stage delayed the tumor initiation significantly compared to non-treatment group (p-value=0.032). The median age for first palpable tumor in the Very Early treatment group was 38 weeks compared to 32 weeks in non-treatment group. Treatment at the Early stage, to my surprise, further boosted the protection compared to Very Early treatment (p-value=0.029). The median age for the first palpable tumor was 41 weeks and 4 mice were tumor-free till now. The median age will be at least 41 weeks over time. These data indicate that the Early treatment regimen provided the best protection in delaying tumor initiation, even more than Very Early treatment.

A



B

Group Comparison	P-value	Significance
Very Early Treatment vs. Early Treatment	0.029	*
Very Early Treatment vs. Non-treatment	0.0164	**
Early Treatment vs. Non-treatment	0.0001	****

Figure 7-3. Tumor initiation was significantly delayed by Very Early treatment and Early. Tumor free curve for Very Early treatment group (16 mice), Early treatment group (15 mice) and non-treatment group (28 mice).

Tumor volumes for each mouse were recorded weekly after the first palpable tumor was detected. The summary of the tumor volume data is shown in Figure 7-4. Tumor growth was not suppressed by any of these three regimens once it started. ICI treatment after the 1st palpable tumor did not suppress the tumor growth compared to the non-treatment group. This is consistent with other reports in which therapeutic ICI treatments did not protect mice from tumor challenge alone [110]. Average tumor volume decreased at some time points since the number of mice was reduced in these points because some mice were euthanized when the tumor volume reached the 2,000 mm^3. The Very Early treatment and the Early treatment groups had slightly slower tumor growth but there was no significant difference. These data support the idea that ICI is more effective at early stage with less tumor cells and less suppressive tumor microenvironment. Surprisingly, treatment at the Early stage was more effective than the Very Early stage.

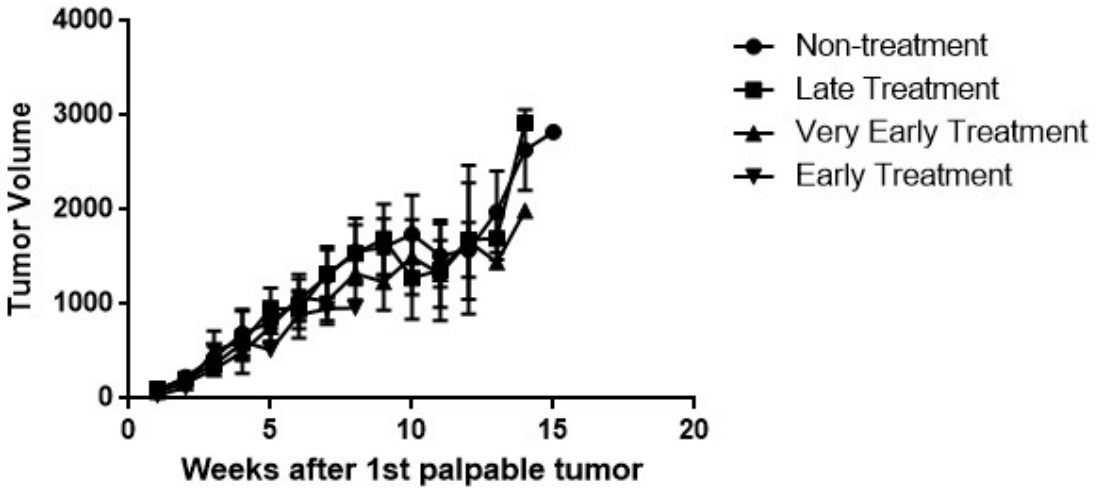


Figure 7-4. Tumor growth curves between the 4 groups were not significant different. Tumor volumes were recorded weekly after the first palpable tumor was detected.

7.3.3 Immunosignature technology can separate responders vs. non-responders prior to the treatment

In the Very Early treatment group, 6 mice (37.5%, responder) developed the first palpable tumor after 41.5 weeks, while the tumor initiation age for another 7 mice (44%, non-responder) was less than 33.5 weeks which was close to the non-treatment group (Figure 7-5A). Similar percentages of responders and non-responders were seen in human ICI clinical trials. PDL1 expression level and mutational load are currently being investigated as biomarkers to separate responders from non-responders. Here the immunosignature technology was used to attempt to classify the responders and non-responders.

Serum samples were collected before the treatment at week 16 and tested with the CIM120K immunosignature array. 2,700 peptides were selected by comparing the two groups with a student's t-test (p-value <0.01). Unsupervised two-way hierarchical

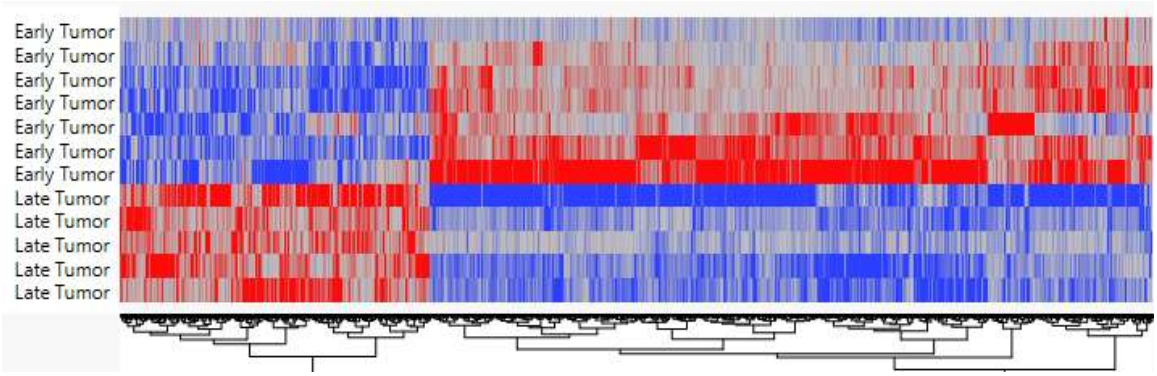
clustering separated the responders and non-responders with 100% accuracy (Figure 7-5B). PCA analysis also separated these two groups with 100% accuracy. It was interesting that the antibody response against most of these 2,700 peptides was lower in responders than non-responders, while 30% of these peptides had much higher signals in responders.

Besides using the immunosignature platform, the human frameshift peptide array which included 300 predicted frameshift antigens (reviewed in chapter 3) was also used for diagnosing responders with serum samples prior to the treatment. A similar approach was applied as with the immunosignatures and 33 peptides were selected by t-test (p-value <0.05). Responders and non-responders were separated with 100% accuracy (Figure 7-6). These data confirmed that it is possible to use immunosignatures or even frameshift peptide array to separate responders vs. non-responders before the treatment.

A

Mouse	Tumor Initiation Age (weeks)	Type
ET5-6	23.7	Non-responder
ET5-3	24.7	Non-responder
ET3-2	27.14	Non-responder
ET2-2	27.57	Non-responder
ET3-3	30.14	Non-responder
ET1-3t	30.86	Non-responder
ET3-6	33.57	Non-responder
ET2-8	41.57	Responder
ET2-10	47.57	Responder
ET3-8	48.57	Responder
ET1-5	48.86	Responder
ET1-8	48.86	Responder

B



C

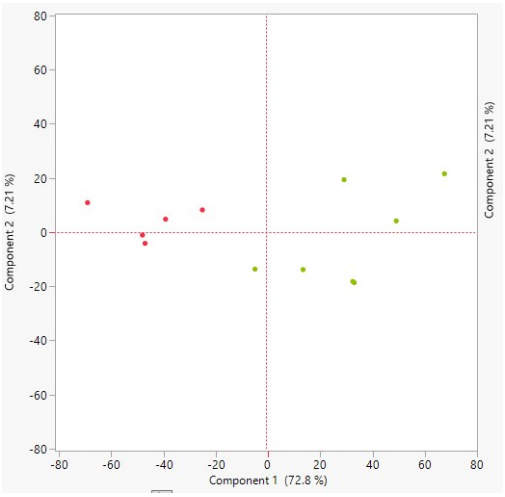


Figure 7-5. Immunosignatures can separate responders from non-responders in the ICI Very Early treatment group. A. Tumor initiation ages for each mouse. B. Early tumor samples: non-responders; late tumor: responders. 2,700 peptides were selected by T-test. C. PCA analysis for these 13 mice, green dots were non-responders and red dots were responders.

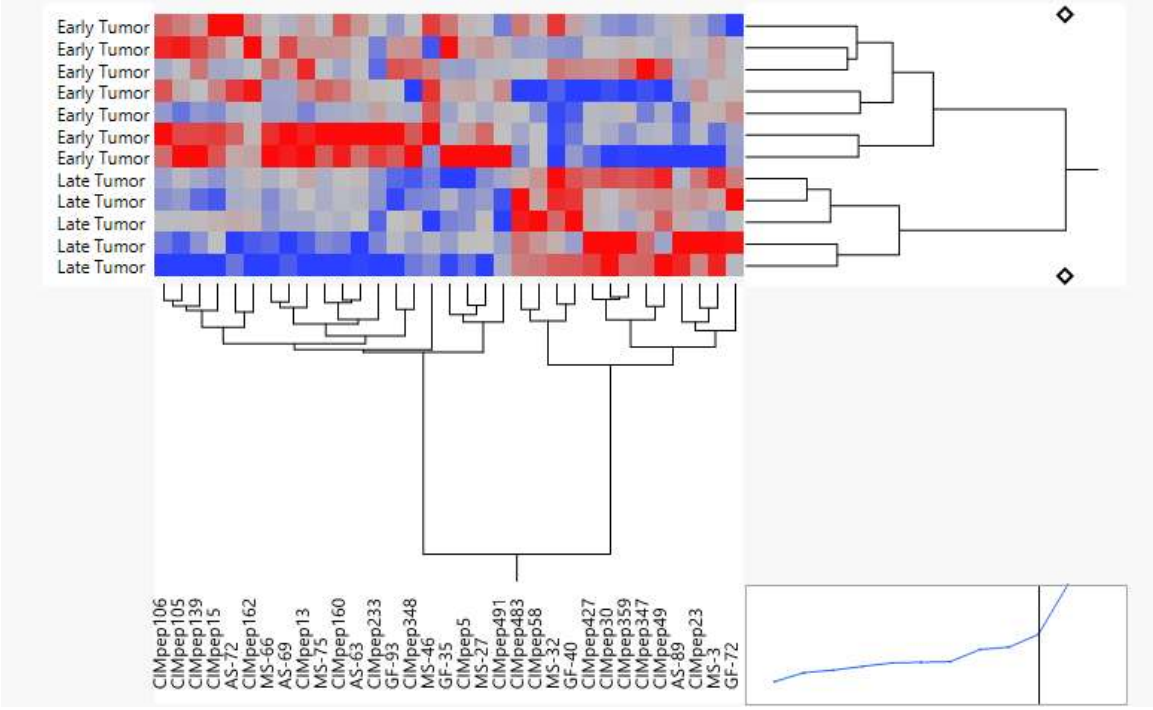


Figure 7-6. The Human Frameshift Peptide Array can separate responders from non-responders in ICI early treatment group.

Each column represented one frameshift peptide from the human frameshift peptide array, each row represented one mouse sample from responder group (late tumor) and non-responder group (early tumor).

7.3.4 Transgenic mice with the responder immunosignature develop tumor later even without ICI treatment

Whether patients would benefit from ICI treatment has been related to the conditions of patients before the treatment. Three major predictors were tumor PDL1 expression, mutational load and the proportion of active CD8+ infiltrating lymphocytes. In summary, a pre-existing active immune response against tumor would boost the efficacy of ICI treatment. Under this hypothesis, patients with pre-existing active immune response against the tumor would be predicted to have better survival even without ICI treatment.

To test this idea, mice from non-treatment group were separated into a late tumor initiation group and an early tumor initiation group, and the responder immunosignature from the ICI treatment (the same 2,700 peptides as in Figure 7-5) was used to classify later tumor initiation events and early tumor initiation events.

A total of 11 mice (Figure 7-7A) were included in the test. 6 mice developed the tumor very early (before 27.6 weeks) and 5 mice had late tumor initiation (after 39.6 weeks). Serum samples were also collected at week 16 and the CIM120K arrays were used.

These 2,700 peptides were selected with student's t-test by comparing responder vs. non-responders in ICI Very Early treatment group. 10 mice were correctly predicted (91% accuracy), and there was only 1 mouse (NT3-3, marked with black arrow) in the late tumor initiation group was misclassified into the early tumor initiation group. The immunosignature pattern of these non-treatment mice was close to the responder vs. non-responder mice in early ICI treatment group with weaker differences.

One mouse was misclassified using the responder immunosignature. To investigate whether the misclassification was due to the responder immunosignature or not, late tumor events and early tumor events in non-treatment group were compared directly. 1,509 peptides were selected with t-test (p-value<0.05). The same mouse (NT3-3) was still misclassified using these 1,509 peptides (Figure 7-8), which indicates that the responder immunosignature had the same classification power in separating early tumor events and late tumor events.

Since the responder immunosignature could predict which mice would grow tumor later without treatment, it was interesting to know whether the prediction could work in the reverse direction (use the late tumor immunosignature against the non-treatment group to predict which mice would be responders to ICI treatment). These same 1,509 peptides were used for classifying responders vs. non-responders in Very Early treatment group. These peptides could not separate responders from non-responders and samples were mixed together. A major difference between these 1,509 peptides (late tumor signature) and the previous 2,700 peptides (responder signature) was that most of these 1,509 peptides had higher signals in late tumor samples which was the opposite in the responder signature.

The responder signature (33 peptides in Figure 7-6) from the frameshift peptide array was tested with the same method. However, these 33 peptides could not separate early tumor events from late tumor events in the non-treatment group. Early tumor samples were totally mixed with late tumor samples and there was not a clear signature (Figure 7-9).

These results demonstrated that mice which could potentially benefit from ICI treatment (with responder immunosignature) would develop tumor later than mice with non-responder immunosignature without any treatments. In other words, this data indicates that the hypothesis that mice with pre-existing active immune response against the tumor would develop tumor later with or without ICI treatments, and the immunosignature technology could be used for identifying this responder signature. The frameshift peptide array, using much fewer features, failed in predicting responders.

A

Mouse	Tumor Initiation Age (weeks)	Type
NT3-7	24.57	Early Tumor
NT1-2	25.14	Early Tumor
NT4-6	26.14	Early Tumor
TT5-4	27.14	Early Tumor
TT5-6	27.14	Early Tumor
TT3-5	27.57	Early Tumor
NT3-3	39.57	Late Tumor
TT4-3	39.57	Late Tumor
NT1-Tat	40.71	Late Tumor
TT2-2	45.86	Late Tumor
TT2-6	47.86	Late Tumor

B

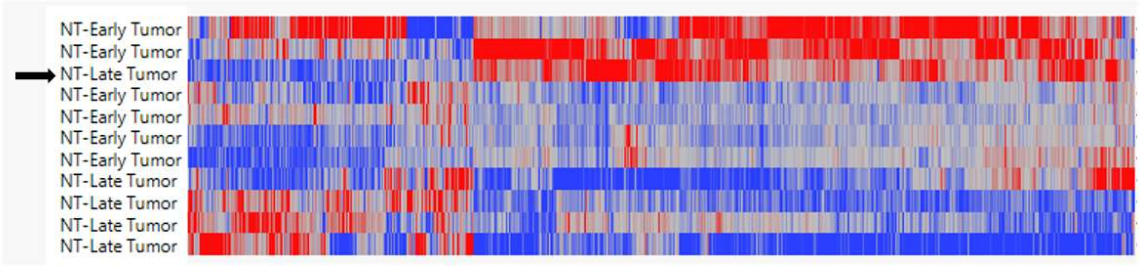
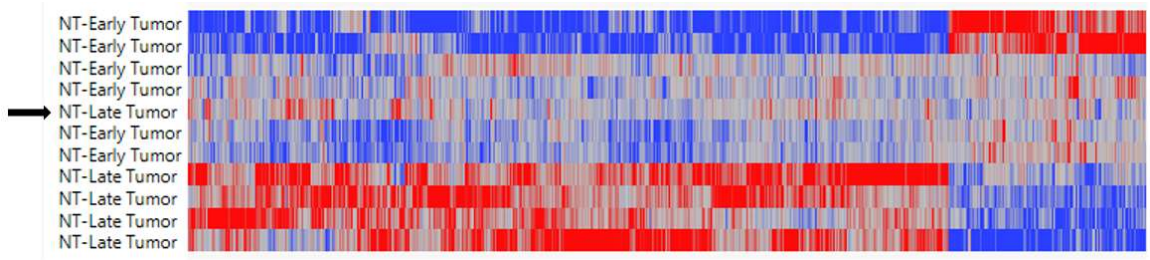


Figure 7-7. Responder immunosignature from ICI treatment could separate late tumor initiation mice from early tumor initiation mice without ICI treatment. A. List of 11 mice in non-treatment group. B. Unsupervised two-way hierarchical clustering of 11 mice in A, black arrow indicated one misclassified sample.

A



B

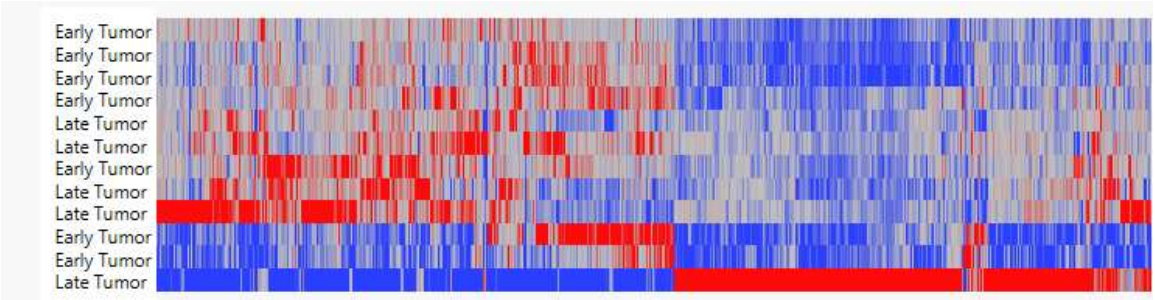


Figure 7-8. Using 1,509 peptides selected by comparing early tumor events and late tumor events for classification (p-value <0.05). A. same mice (NT3-3) was misclassified with these 1,509 peptides. B. These 1,509 peptides couldn't separate responders vs. non-responders in early treatment group.

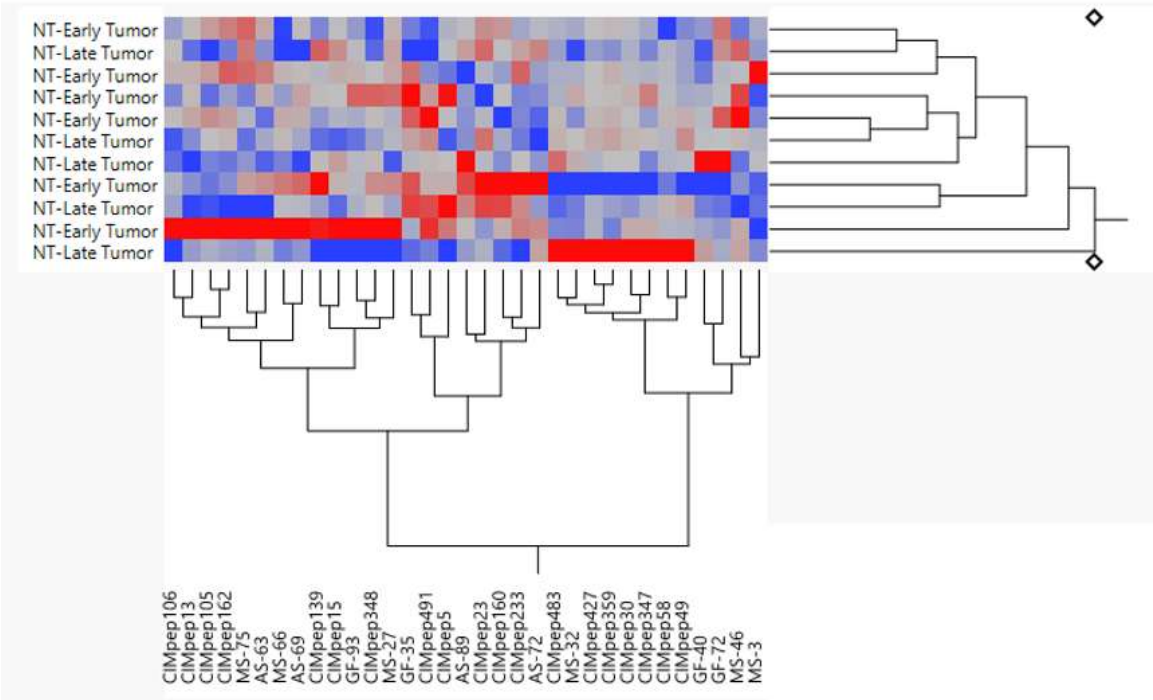


Figure 7-9. Responder signature from human frameshift peptide array could not separate early tumor events from late tumor events in the non-treatment group.

Each column represented one frameshift peptide from the human frameshift peptide array, each row represented one mouse sample from late tumor events and early tumor events in the non-treatment group.

7.3.5 A model for early immunotherapy cancer treatment

Based on the results from these experiments, I propose a model to explain the response to checkpoint inhibitors treatment in cancer patients (Figure 7-10). It is known that three important factors are important in determining the treatment time: tumor burden, number of tumor neo-antigens and immunosuppression. When tumor burden is too high and there are billions of tumor cells, lymphocytes can be outnumbered and it is difficult for ICI treatment to be effective. Meanwhile, the number of tumor neoantigens grows as the tumor burden increases and more tumor neoantigens are related to a strong anti-tumor

immune response with ICI treatment. Immunosuppression is the third factor which increases as the tumor grows and this makes ICI treatment less effective. Here we use the capacity of the immune system as the measure in the y-axis of the figure, which is the opposite of immunosuppression. Time A (week 16) represents the Very Early treatment in our previous mouse experiment. There is no immunosuppression at Time A, but there are too few tumor neo-antigens with few tumor cells and ICI treatments are relatively ineffective with not enough neoantigens. Time B (week 24-26) represents the Early treatment, which is the perfect timepoint for ICI treatment with enough tumor neoantigens and low immunosuppression at the same time. Time C (treat at palpable tumor) does not work at all with high tumor burden and high immunosuppression, the immune system cannot deal with too many tumor cells with the high level of immunosuppression at the same time. Although the number of the tumor neoantigens is large they are not effectively seen by the immune system due to suppression.

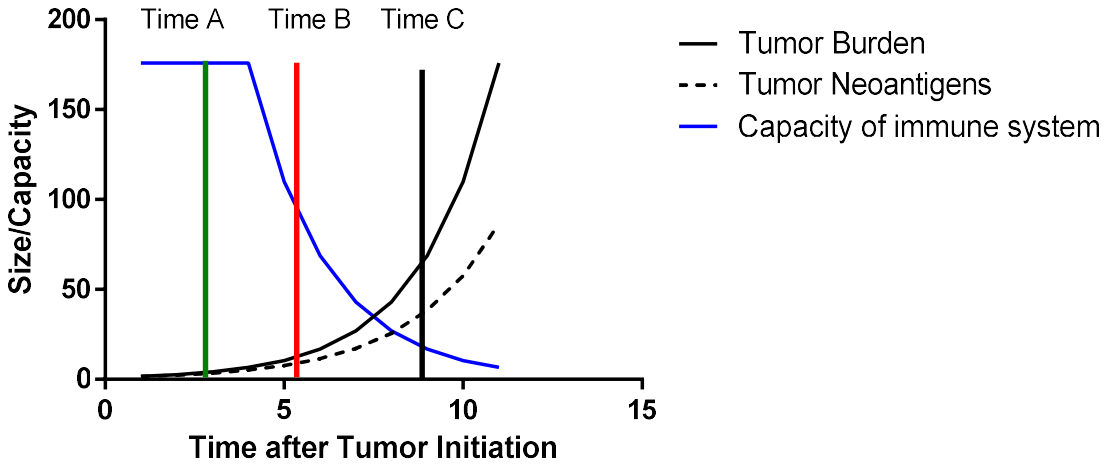


Figure 7-10. A simple Model for early ICI treatment in cancer patients.

Tumor burden and tumor neoantigens increase exponentially as tumor grows, while the capacity of immune system does not change in the earliest time of tumor initiation, but then drops exponentially as tumor grows and immunosuppression increases. Time A represents the Very Early treatment (week 16) in our mouse model. There is no immunosuppression but not enough tumor neoantigens, which makes the treatment relatively ineffective. Time B is the best treatment time (week 24-26) with relatively low immunosuppression, low tumor burden and relatively high tumor neoantigens. Time C has the advantages of more tumor neoantigens, but with high tumor burden and high immunosuppression, the treatment could not work.

7.4 Discussion

In this study, we compared the idea of 'early cancer detection, early ICI treatment' and conventional therapeutic ICI treatment in the FVB/N neuN transgenic mouse model. Conventional therapeutic ICI treatment did not suppress the tumor growth at all in this model, while Very Early treatment and Early treatment delayed the first palpable tumor from 31.7 weeks to 38 weeks and 41 weeks correspondingly. Furthermore, the immunosignature technology was successfully used to separate responders from non-responders and the responder immunosignature could predict late tumor initiation events even without ICI treatment.

Therapeutic ICI treatment did not suppress tumor growth in this FVB/N neuN mouse model, which is consistent with some reports using ICI treatments [143]. Tumor cells from this mouse model had much lower PDL1 expression compared to other models [110], which made therapeutic ICI treatment less effective in this model. However, Very Early and Early ICI treatment could delay the tumor initiation significantly, which could be related to less tumor cells and less immune suppression in the early stages. More interestingly, Early treatment had significantly better protection than Very Early

treatment. ICI activates the immune system by turning off the brakes, but cancer cells have to present enough neo-antigens for the immune system to produce CD8+ killing T cells. Higher number of mutations and immunogenic neo-antigens correlated to better response for ICI treatment [18]. Probably 16 weeks was too early for the ICI treatment in this mouse model even though the immunosignature could distinguish the difference compared to WT mice. These results indicate that timing may be crucial for effective ICI treatment.

Mutational load and PDL1 expression are commonly used for predicting responders in ICI treatment. Immunosignatures have the advantage of being a simple and fast test. However, immunosignatures assess the humoral immune system which is usually considered not important and has received less attention in cancer research. Our results show that immunosignature arrays can predict responders for ICI treatment accurately prior to the treatment in this mouse model, and this responder immunosignature could further predict which mice would have late tumor initiation without ICI treatment. The immune system is important in battling against cancer. Even for mice with the same breeding, different mice had different immune responses against the tumor initiation and mice with responder the immunosignature had more active anti-tumor immune responses. These mice developed tumor late naturally without treatment and could also be boosted by ICI treatment.

We noticed that the responder immunosignature could predict late tumor initation events in the non-treatment group, but the late tumor immunosignature could not predict responders in ICI treatment group. The responder immunosignature was mostly anti-

tumor immune response because the anti-tumor response was boosted by ICI treatment. While the late tumor immunosignature in the non-treatment group was not only composed of anti-tumor immune response, but also included other unknown components which might related to late tumor initiation. Since the anti-tumor immune response was the key for predicting responders or late tumor events, it is easy to understand why the reverse direction wouldn't work.

There were only 800 frameshift peptides (of 200,000 possible for mis-splicing and mis-transcription through MSs) in the human FS array. This platform could separate responders vs. non-responders from the ICI treatment but could not be used for predicting late tumor initiation events. This is presumably primarily due to the fewer number of features (800 vs 120,000) compared to the immunosignature arrays.

In summary, our results demonstrated that the idea of early diagnosis and early treatment was feasible with the immunosignature technology, and this approach delayed tumor initiation significantly in the FVB/N neuN model while conventional therapeutic ICI treatment did not suppress tumor growth at all. Immunosignature technology could also be used as biomarker to predict responders vs. non-responders in ICI treatments. Other early treatments include chemotherapy and vaccination are possible with this technology and remain to be investigated.

7.5 Methods

7.5.1 Mice and sera collection

Transgenic FVB/N neuN mice were kindly gifted by Dr. Chella David at Mayo, Rochester. Blood samples were collected from week 16 and every 4 weeks after that.

Serum was separated with serum separating Microcontainer (BD Biosciences, CA) and stored in -20C.

7.5.2 Binding serum samples to CIM120K peptide array and human frameshift peptide array

Serum samples were collected at week 16 prior to treatment and tested on the CIM120K peptide array and human frameshift peptide array. CIM120K array was reviewed in Chapter 6 and the protocol for running samples on the CIM120K array was summarized in the Chapter 5 Methods.

The human frameshift peptide array included 800 FS peptides from human coding genes (predicted microsatellite FS peptides and mis-splicing FS peptides from tumor EST libraries). These 800 peptides were printed on the NSB9 surface. The protocol for running samples on the human FS arrays was identical to the immunosignature assay.

Fluorescence intensities were normalized to median per array and analyzed with JMP Pro 13.

7.5.3 Immune checkpoint inhibitor treatments

Mice in treatment groups were treated with 7 doses of ICI as shown in Figure 7-2. Each dose of ICI treatment included 200 μg anti-PDL1 mAb (10F.9G2) from BioXcell and 100 μg anti-CTLA4 mAb (UC10-4F10-11) from BioXcell as well. ICI treatment was administered via IP.

CHAPTER 8

CONCLUSION

Tumor FS neo-antigens at the mRNA level are abundant resources for cancer diagnosis and both preventive and therapeutic cancer vaccine development. However, this area is fairly neglected by the scientific world. In this thesis, I demonstrated the feasibility of using FS neo-antigens for cancer vaccine development. Moreover, I pushed the idea of "early detection early treatment" forward with the Immunosignature technology as a diagnostic and immune checkpoint blockade as treatment.

I first built the microsatellite frameshift database for human, dog and mouse. Besides being as resources for vaccine development, these microsatellites and frameshift peptides from these regions showed interesting patterns and positive/negative selection pressures were seen in different subgroups. Very long FS peptides were enriched in the microsatellite regions and they had very unique amino acid composition and codon usage patterns which were not similar to the short FS or the WT proteome. The functions of these very long FS peptides were still unknown. Clearly MS FS candidates are better tumor antigens with high immunogenicity, high abundance and low chance to cause autoimmune diseases compared to antigens from genomic point mutations. The presence of microsatellite frameshift mutations was detected in both tumor cDNA samples and cell lines. Surprisingly, the mutation rate was very high in an 11A homopolymer region and mutation rates correlated with repeat length. A pool of 9 MS candidates showed

protection in mouse breast tumor model and they were simply selected based on informatics (repeat length, peptide length and MHC prediction).

Next, I designed the human and dog frameshift peptide array with 800 FS peptides representing 300 FS antigens from ~200K possible FS antigens. We screened hundreds of dog tumor samples on our FS array and selected positive and negative FS peptides in the cancer group from the array. It turned out positive FS peptides could offer protection in both mouse breast tumor and melanoma models while negative FS peptides could not. This phenomenon was further validated by another independent experiment with a similar approach from our group (unpublished data). These results showed that frameshift peptide arrays could be a good platform for FS antigen selection. It was puzzling that this group of negative FS peptides had 0 positive sample in the cancer group while there were positive samples in the normal control group, I had a preliminary explanation in the discussion but it could not explain this phenomenon very well and further investigation is needed.

The first version of dog cancer vaccine was developed with dog frameshift array and safety trial was conducted to evaluate the vaccine. The results demonstrated that this FS antigens based vaccine was safe with no adverse effect and it could induce strong B/T cell immune responses.

Meanwhile, I built the human exon junction FS database and the dog version database to cover all possible mis-splicing FS antigens. I provided a method to extract anti-FS immune from large immunosignature datasets. The results showed three interesting patterns of antibody response in different stages of breast cancer which might

provide insights in cancer development and vaccine design. Besides, our group designed the human/dog 400K frameshift peptide arrays which included all possible microsatellite FS antigens and mis-splicing antigens. It is currently under production.

And last, I tested the idea of "early cancer detection, early treatment" in mouse transgenic breast cancer model with current immune checkpoint blockade treatments. Immunosignature technology could detect cancer at very early stage in this mouse model. Results showed that the timing of immune checkpoint inhibitor treatment was crucial to maximize the treatment benefits. Treating at Early stage but not Very Early provided the best outcome which might indicate a minimum number of tumor antigens was required for ICIs to be effective. In the same time, treating at late stage (palpable tumor) did not offer any protection in this mouse model which meant that early treatment was necessary.

Based our hypothesis, frameshift neo-antigens are generated due to the abnormal transcriptional system and dysfunctional quality control system. The future of my research includes three areas: 1) explore the possibility of using human/dog 400K FS array for cancer diagnostics, antigen selection; 2) explore the anti-FS immune response in other chronic/acute diseases and the potential of using FS based vaccines to treat other diseases including aging; 3) explore the idea of "early detection early treatment" with new cancer treatments and combinational cancer treatments.

REFERENCES

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer Statistics, 2017.* CA Cancer J Clin, 2017. **67**(1): p. 7-30.

2. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

3. Tomasetti, C., L. Li, and B. Vogelstein, *Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention.* Science, 2017. **355**(6331): p. 1330-1334.

4. Kantoff, P.W., et al., *Sipuleucel-T immunotherapy for castration-resistant prostate cancer.* N Engl J Med, 2010. **363**(5): p. 411-22.

5. Kalos, M., et al., *T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia.* Sci Transl Med, 2011. **3**(95): p. 95ra73.

6. Morgan, R.A., et al., *Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy.* J Immunother, 2013. **36**(2): p. 133-51.

7. *FDA Approved Cancer Drugs*. Available from: http://wayback.archive-it.org/7993/20170111064250/http://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm279174.htm.

8. Mitri, Z., T. Constantine, and R. O'Regan, *The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy.* Chemother Res Pract, 2012. **2012**: p. 743193.

9. *Critics condemn bowel cancer drug rejection*. Available from: http://www.bbc.com/news/health-11060968.

10. Dunn, G.P., L.J. Old, and R.D. Schreiber, *The three Es of cancer immunoediting.* Annu Rev Immunol, 2004. **22**: p. 329-60.

11. Mittal, D., et al., *New insights into cancer immunoediting and its three component phases--elimination, equilibrium and escape.* Curr Opin Immunol, 2014. **27**: p. 16-25.

12. Topalian, S.L., et al., *Safety, activity, and immune correlates of anti-PD-1 antibody in cancer.* N Engl J Med, 2012. **366**(26): p. 2443-54.

13. Brahmer, J.R., et al., *Safety and activity of anti-PD-L1 antibody in patients with advanced cancer.* N Engl J Med, 2012. **366**(26): p. 2455-65.

14. Marquez-Rodas, I., et al., *Immune checkpoint inhibitors: therapeutic advances in melanoma.* Ann Transl Med, 2015. **3**(18): p. 267.

15. Fukuhara, H., Y. Ino, and T. Todo, *Oncolytic virus therapy: A new era of cancer treatment at dawn.* Cancer Sci, 2016. **107**(10): p. 1373-1379.

16. Snyder, A., et al., *Genetic basis for clinical response to CTLA-4 blockade in melanoma.* N Engl J Med, 2014. **371**(23): p. 2189-99.

17. van Rooij, N., et al., *Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma.* J Clin Oncol, 2013. **31**(32): p. e439-42.

18. Turajlic, S., et al., *Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis.* Lancet Oncol, 2017. **18**(8): p. 1009-1021.

19. Chang, M.T., et al., *Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity.* Nat Biotechnol, 2016. **34**(2): p. 155-63.

20. Vonderheide, R.H. and K.L. Nathanson, *Immunotherapy at large: the road to personalized cancer vaccines.* Nat Med, 2013. **19**(9): p. 1098-100.

21. Robert Petit, M.F.P., *The Potential of ADXS-NEO: Leveraging Listeria Monocytogenes as a Vector for Personalized Cancer Immunotherapy.* 2016.

22. *SynCon® immunotherapies: powerful antigen-specific immune responses*. Available from: http://www.inovio.com/technology/synthetic-vaccines/.

23. *PsiOxus Technology*. Available from: http://psioxus.com/technology-products/.

24. *Caperna Webpage*. Available from: https://www.modernatx.com/ecosystem/moderna-ventures.

25. Ott, P.A., et al., *An immunogenic personal neoantigen vaccine for patients with melanoma.* Nature, 2017. **547**(7662): p. 217-221.

26. Sahin, U., et al., *Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer.* Nature, 2017. **547**(7662): p. 222-226.

27. Mullard, A., *The cancer vaccine resurgence.* Nat Rev Drug Discov, 2016. **15**(10): p. 663-5.

28. Yadav, M., et al., *Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing.* Nature, 2014. **515**(7528): p. 572-6.

29. Moura, R.F., D.A. Dawson, and D.M. Nogueira, *The use of microsatellite markers in Neotropical studies of wild birds: a literature review.* An Acad Bras Cienc, 2017: p. 0.

30. Amos, W., *Mutation biases and mutation rate variation around very short human microsatellites revealed by human-chimpanzee-orangutan genomic sequence alignments.* J Mol Evol, 2010. **71**(3): p. 192-201.

31. Brouwer, J.R., R. Willemsen, and B.A. Oostra, *Microsatellite repeat instability and neurological disease.* Bioessays, 2009. **31**(1): p. 71-83.

32. Iyer, R.R., et al., *DNA mismatch repair: functions and mechanisms.* Chem Rev, 2006. **106**(2): p. 302-23.

33. Vilar, E. and S.B. Gruber, *Microsatellite instability in colorectal cancer-the stable evidence.* Nat Rev Clin Oncol, 2010. **7**(3): p. 153-62.

34. Kim, T.M., P.W. Laird, and P.J. Park, *The landscape of microsatellite instability in colorectal and endometrial cancer genomes.* Cell, 2013. **155**(4): p. 858-68.

35. Hause, R.J., et al., *Classification and characterization of microsatellite instability across 18 cancer types.* Nat Med, 2016. **22**(11): p. 1342-1350.

36. Le, D.T., et al., *PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.* N Engl J Med, 2015. **372**(26): p. 2509-20.

37. Spandidos, D., et al., *Microsatellite instability in patients with chronic obstructive pulmonary disease.* Oncol Rep, 1996. **3**(3): p. 489-91.

38. Samara, K., et al., *Microsatellite DNA instability in benign lung diseases.* Respir Med, 2006. **100**(2): p. 202-11.

39. Lynch, M., *Evolution of the mutation rate.* Trends Genet, 2010. **26**(8): p. 345-52.

40. Gout, J.F., et al., *Large-scale detection of in vivo transcription errors.* Proc Natl Acad Sci U S A, 2013. **110**(46): p. 18584-9.

41. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.

42. Rizvi, N.A., et al., *Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer.* Science, 2015. **348**(6230): p. 124-8.

43. Gubin, M.M., et al., *Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens.* Nature, 2014. **515**(7528): p. 577-81.

44. Schumacher, T.N. and R.D. Schreiber, *Neoantigens in cancer immunotherapy.* Science, 2015. **348**(6230): p. 69-74.

45. Brinkmann, B., et al., *Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat.* Am J Hum Genet, 1998. **62**(6): p. 1408-15.

46. Zhu, L., et al., *Microsatellite instability and survival in gastric cancer: A systematic review and meta-analysis.* Mol Clin Oncol, 2015. **3**(3): p. 699-705.

47. Woerner, S.M., et al., *Microsatellite instability of selective target genes in HNPCC-associated colon adenomas.* Oncogene, 2005. **24**(15): p. 2525-35.

48. Vaish, M., et al., *Microsatellite instability as prognostic marker in bladder tumors: a clinical significance.* BMC Urol, 2005. **5**: p. 2.

49. Metzgar, D., J. Bytof, and C. Wills, *Selection against frameshift mutations limits microsatellite expansion in coding DNA.* Genome Res, 2000. **10**(1): p. 72-80.

50. Ellegren, H., *Microsatellites: simple sequences with complex evolution.* Nat Rev Genet, 2004. **5**(6): p. 435-45.

51. Pruitt, K.D., et al., *The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes.* Genome Res, 2009. **19**(7): p. 1316-23.

52. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D733-45.

53. Greenman, C., et al., *Patterns of somatic mutation in human cancer genomes.* Nature, 2007. **446**(7132): p. 153-8.

54. Nakamura, Y., T. Gojobori, and T. Ikemura, *Codon usage tabulated from international DNA sequence databases: status for the year 2000.* Nucleic Acids Res, 2000. **28**(1): p. 292.

55. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.

56. Greenbaum, J., et al., *Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes.* Immunogenetics, 2011. **63**(6): p. 325-35.

57. Weiskopf, D., et al., *Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells.* Proc Natl Acad Sci U S A, 2013. **110**(22): p. E2046-53.

58. Wang, P., et al., *Peptide binding predictions for HLA DR, DP and DQ molecules.* BMC Bioinformatics, 2010. **11**: p. 568.

59. Andreatta, M. and M. Nielsen, *Gapped sequence alignment using artificial neural networks: application to the MHC class I system.* Bioinformatics, 2016. **32**(4): p. 511-7.

60. Peters, B. and A. Sette, *Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method.* BMC Bioinformatics, 2005. **6**: p. 132.

61. Sinicrope, F.A. and D.J. Sargent, *Molecular pathways: microsatellite instability in colorectal cancer: prognostic, predictive, and therapeutic implications.* Clin Cancer Res, 2012. **18**(6): p. 1506-12.

62. Kloor, M. and M. von Knebel Doeberitz, *The Immune Biology of Microsatellite-Unstable Cancer.* Trends in Cancer, 2016. **2**(3): p. 121-133.

63. Wang, Y., X. Liu, and Y. Li, *Target genes of microsatellite sequences in head and neck squamous cell carcinoma: mononucleotide repeats are not detected.* Gene, 2012. **506**(1): p. 195-201.

64. Benatti, P., et al., *Microsatellite instability and colorectal cancer prognosis.* Clin Cancer Res, 2005. **11**(23): p. 8332-40.

65. Bae, S., et al., *Microsatellite instability status is critical to analysis of survival in stage II colon cancer.* J Clin Oncol, 2012. **30**(6): p. 675-6; author reply 676-7.

66. Xiao, Y. and G.J. Freeman, *The microsatellite instable subset of colorectal cancer is a particularly good candidate for checkpoint blockade immunotherapy.* Cancer Discov, 2015. **5**(1): p. 16-8.

67. Dudley, J.C., et al., *Microsatellite Instability as a Biomarker for PD-1 Blockade.* Clin Cancer Res, 2016. **22**(4): p. 813-20.

68. Rosenbaum, M.W., et al., *PD-L1 expression in colorectal cancer is associated with microsatellite instability, BRAF mutation, medullary morphology and cytotoxic tumor-infiltrating lymphocytes.* Mod Pathol, 2016. **29**(9): p. 1104-12.

69. Bauer, K., et al., *T cell responses against microsatellite instability-induced frameshift peptides and influence of regulatory T cells in colorectal cancer.* Cancer Immunol Immunother, 2013. **62**(1): p. 27-37.

70. Reuschenbach, M., et al., *Serum antibodies against frameshift peptides in microsatellite unstable colorectal cancer patients with Lynch syndrome.* Fam Cancer, 2010. **9**(2): p. 173-9.

71. Maby, P., et al., *Correlation between Density of CD8+ T-cell Infiltrate in Microsatellite Unstable Colorectal Cancers and Frameshift Mutations: A Rationale for Personalized Immunotherapy.* Cancer Res, 2015. **75**(17): p. 3446-55.

72. You, K.T., et al., *Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors.* PLoS Biol, 2007. **5**(5): p. e109.

73. Pulaski, B.A. and S. Ostrand-Rosenberg, *Mouse 4T1 breast tumor model.* Curr Protoc Immunol, 2001. **Chapter 20**: p. Unit 20 2.

74. Kreiter, S., et al., *Mutant MHC class II epitopes drive therapeutic immune responses to cancer.* Nature, 2015. **520**(7549): p. 692-6.

75. Svarovsky, S., A. Borovkov, and K. Sykes, *Cationic gold microparticles for biolistic delivery of nucleic acids.* Biotechniques, 2008. **45**(5): p. 535-40.

76. Chambers, R.S. and S.A. Johnston, *High-level generation of polyclonal antibodies by genetic immunization.* Nat Biotechnol, 2003. **21**(9): p. 1088-92.

77. Sykes, K.F. and S.A. Johnston, *Genetic live vaccines mimic the antigenicity but not pathogenicity of live viruses.* DNA Cell Biol, 1999. **18**(7): p. 521-31.

78. Tang, D.C., M. DeVit, and S.A. Johnston, *Genetic immunization is a simple method for eliciting an immune response.* Nature, 1992. **356**(6365): p. 152-4.

79. Svarovsky, S.A., et al., *Self-assembled micronanoplexes for improved biolistic delivery of nucleic acids.* Mol Pharm, 2009. **6**(6): p. 1927-33.

80. Shen, L., *Investigation of Tumor Frame Shift Antigens for Prophylactic Cancer Vaccine, Cancer Detection and Tumorigenicity.* Arizona State University, 2012.

81. Lee, H., *Identification of Neo-antigens for a Cancer Vaccine by Transcriptome Analysis.* Arizona State University, 2012.

82. Stafford, P., et al., *Immunosignature system for diagnosis of cancer.* Proc Natl Acad Sci U S A, 2014. **111**(30): p. E3072-80.

83. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nat Genet, 2008. **40**(12): p. 1413-5.

84. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

85. Sykes, K.F., J.B. Legutki, and P. Stafford, *Immunosignaturing: a critical review.* Trends Biotechnol, 2013. **31**(1): p. 45-51.

86. Halperin, R.F., P. Stafford, and S.A. Johnston, *Exploring antibody recognition of sequence space through random-sequence peptide microarrays.* Mol Cell Proteomics, 2011. **10**(3): p. M110 000786.

87. Restrepo, L., et al., *Application of immunosignatures to the assessment of Alzheimer's disease.* Ann Neurol, 2011. **70**(2): p. 286-95.

88. Navalkar, K.A., et al., *Application of immunosignatures for diagnosis of valley fever.* Clin Vaccine Immunol, 2014. **21**(8): p. 1169-77.

89. Virnig, B.A., et al., *Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes.* J Natl Cancer Inst, 2010. **102**(3): p. 170-8.

90. Schreiber, R.D., L.J. Old, and M.J. Smyth, *Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion.* Science, 2011. **331**(6024): p. 1565-70.

91. Lian, B. and J. Guo, *Checkpoint inhibitors in treatment of metastatic mucosal melanoma.* Chin Clin Oncol, 2014. **3**(3): p. 37.

92. Postow, M.A., M.K. Callahan, and J.D. Wolchok, *Immune Checkpoint Blockade in Cancer Therapy.* J Clin Oncol, 2015. **33**(17): p. 1974-82.

93. Riess, J.W., P.N. Lara, Jr., and D.R. Gandara, *Theory Meets Practice for Immune Checkpoint Blockade in Small-Cell Lung Cancer.* J Clin Oncol, 2016.

94. Chen, T.W., et al., *A systematic review of immune-related adverse event reporting in clinical trials of immune checkpoint inhibitors.* Ann Oncol, 2015. **26**(9): p. 1824-9.

95. Abdel-Wahab, N., M. Shah, and M.E. Suarez-Almazor, *Adverse Events Associated with Immune Checkpoint Blockade in Patients with Cancer: A Systematic Review of Case Reports.* PLoS One, 2016. **11**(7): p. e0160221.

96. Michot, J.M., et al., *Immune-related adverse events with immune checkpoint blockade: a comprehensive review.* Eur J Cancer, 2016. **54**: p. 139-48.

97. Schadendorf, D., et al., *Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma.* J Clin Oncol, 2015. **33**(17): p. 1889-94.

98. Motzer, R.J., et al., *Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma.* N Engl J Med, 2015. **373**(19): p. 1803-13.

99. Motzer, R.J., et al., *Nivolumab for Metastatic Renal Cell Carcinoma: Results of a Randomized Phase II Trial.* J Clin Oncol, 2015. **33**(13): p. 1430-7.

100. Garon, E.B., et al., *Pembrolizumab for the treatment of non-small-cell lung cancer.* N Engl J Med, 2015. **372**(21): p. 2018-28.

101. Larkin, J., et al., *Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma.* N Engl J Med, 2015. **373**(1): p. 23-34.

102. Robert, C., et al., *Pembrolizumab versus Ipilimumab in Advanced Melanoma.* N Engl J Med, 2015. **372**(26): p. 2521-32.

103. Hamid, O., et al., *Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma.* N Engl J Med, 2013. **369**(2): p. 134-44.

104. Duan, H., *Early Detection and Treatment of Breast Cancer by Random Peptide Array in neuN Transgenic Mouse Model.* Arizona State University.

105. Topalian, S.L., et al., *Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy.* Nat Rev Cancer, 2016. **16**(5): p. 275-87.

106. Snyder, A., et al., *Genetic basis for clinical response to CTLA-4 blockade in melanoma.* N Engl J Med, 2014. **371**(23): p. 2189-2199.

107. Sunshine, J. and J.M. Taube, *PD-1/PD-L1 inhibitors.* Curr Opin Pharmacol, 2015. **23**: p. 32-8.

108. Guy, C.T., et al., *Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease.* Proc Natl Acad Sci U S A, 1992. **89**(22): p. 10578-82.

109. Muller, W.J., et al., *Synergistic interaction of the Neu proto-oncogene product and transforming growth factor alpha in the mammary epithelium of transgenic mice.* Mol Cell Biol, 1996. **16**(10): p. 5726-36.

110. Nolan, E., et al., *Combined immune checkpoint blockade as a therapeutic strategy for BRCA1-mutated breast cancer.* Sci Transl Med, 2017. **9**(393).

111. Forbes, S.A., et al., *COSMIC: High-Resolution Cancer Genetics Using the Catalogue of Somatic Mutations in Cancer.* Curr Protoc Hum Genet, 2016. **91**: p. 10 11 1-10 11 37.

112. Futreal, P.A., et al., *A census of human cancer genes.* Nat Rev Cancer, 2004. **4**(3): p. 177-83.

113. Krupp, M., et al., *RNA-Seq Atlas--a reference database for gene expression profiling in normal tissue by next-generation sequencing.* Bioinformatics, 2012. **28**(8): p. 1184-5.

114. Scott, A.M., J.P. Allison, and J.D. Wolchok, Monoclonal antibodies in cancer therapy. Cancer Immun, 2012. 12: p. 14.

115. Hanna, K.S., A Review of Immune Checkpoint Inhibitors for the Management of Locally Advanced or Metastatic Urothelial Carcinoma. Pharmacotherapy, 2017. 37(11): p. 1391-1405.

116. Bergman, P.J., et al., Long-term survival of dogs with advanced malignant melanoma after DNA vaccination with xenogeneic human tyrosinase: a phase I trial. Clin Cancer Res, 2003. 9(4): p. 1284-90.

117. Yadav, M., et al., Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature, 2014. 515(7528): p. 572-6.

118. Brouwer, J.R., R. Willemsen, and B.A. Oostra, Microsatellite repeat instability and neurological disease. Bioessays, 2009. 31(1): p. 71-83.

119. Sakharkar, M.K., V.T. Chow, and P. Kangueane, Distributions of exons and introns in the human genome. In Silico Biol, 2004. 4(4): p. 387-93.

120. Pan, Q., et al., Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet, 2008. 40(12): p. 1413-5.

121. Lopez-Bigas, N., et al., Are splicing mutations the most frequent cause of hereditary disease? FEBS Lett, 2005. 579(9): p. 1900-3.

122. Ward, A.J. and T.A. Cooper, The pathobiology of splicing. J Pathol, 2010. 220(2): p. 152-63.

123. He, C., et al., A global view of cancer-specific transcript variants by subtractive transcriptome-wide analysis. PLoS One, 2009. 4(3): p. e4732.

124. Sveen, A., et al., Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. Oncogene, 2016. 35(19): p. 2413-27.

125. Skotheim, R.I. and M. Nees, Alternative splicing in cancer: noise, functional, or systematic? Int J Biochem Cell Biol, 2007. 39(7-8): p. 1432-49.

126. Lee, H., Identification of Neo-antigens for a Cancer Vaccine by Transcriptome Analysis. Arizona State University, 2012.

127. Shen, L., Investigation of Tumor Frame Shift Antigens for Prophylactic Cancer Vaccine, Cancer Detection and Tumorigenicity. Arizona State University, 2012.

128. van Rooij, N., et al., Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. J Clin Oncol, 2013. 31(32): p. e439-42.

129. Khodadoust, M.S., et al., Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. Nature, 2017. 543(7647): p. 723-727.

130. Vonderheide, R.H. and K.L. Nathanson, Immunotherapy at large: the road to personalized cancer vaccines. Nat Med, 2013. 19(9): p. 1098-100.

131. Sahin, U., et al., Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature, 2017. 547(7662): p. 222-226.

132. Ott, P.A., et al., An immunogenic personal neoantigen vaccine for patients with melanoma. Nature, 2017. 547(7662): p. 217-221.

133. Stafford, P., et al., Physical characterization of the "immunosignaturing effect". Mol Cell Proteomics, 2012. 11(4): p. M111 011593.

134. Tan, A.C., A. Goubier, and H.E. Kohrt, A quantitative analysis of therapeutic cancer vaccines in phase 2 or phase 3 trial. J Immunother Cancer, 2015. 3: p. 48.

135. Guo, C., et al., Therapeutic cancer vaccines: past, present, and future. Adv Cancer Res, 2013. 119: p. 421-75.

136. Snyder, A., et al., Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med, 2014. 371(23): p. 2189-99.

137. Lai, X. and A. Friedman, Combination therapy of cancer with cancer vaccine and immune checkpoint inhibitors: A mathematical model. PLoS One, 2017. 12(5): p. e0178479.

138. Kleponis, J., R. Skelton, and L. Zheng, Fueling the engine and releasing the break: combinational therapy of cancer vaccines and immune checkpoint inhibitors. Cancer Biol Med, 2015. 12(3): p. 201-8.

139. Tang, D.C., M. DeVit, and S.A. Johnston, *Genetic immunization is a simple method for eliciting an immune response.* Nature, 1992. **356**(6365): p. 152-4.

140. Kranz, L.M., et al., *Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy.* Nature, 2016. **534**(7607): p. 396-401.

141. Larocca, C. and J. Schlom, *Viral vector-based therapeutic cancer vaccines.* Cancer J, 2011. **17**(5): p. 359-71.

142. Lambricht, L., et al., *Clinical potential of electroporation for gene therapy and DNA vaccine delivery.* Expert Opin Drug Deliv, 2016. **13**(2): p. 295-310.

143. Nolan, E., et al., Combined immune checkpoint blockade as a therapeutic strategy for BRCA1-mutated breast cancer. Sci Transl Med, 2017. 9(393).

APPENDIX A

GENOMIC MUTATIONS FROM MICROSATELLITE REGIONS

Table A.1 Top 5 frequent Genomic Mutations for 40 cancer driver genes with Microsatellite regions

Gene Name	Top 1 Frequent Mutation	2nd	3rd	4th	5th	Mutation Type #1	#2	#3	#4	#5
NPM1	51.72%	36.89%	6.06%	1.12%	0.71%	Insertion - Frameshift	Unknown	Insertion - Frameshift	Unknown	Insertion - Frameshift
ASXL1	19.92%	18.31%	5.46%	4.02%	1.00%	**Insertion - Frameshift**	Unknown	**Deletion - Frameshift**	Nonsense	Frameshift
DNAJB1	10.98%	3.66%	3.66%	2.44%	2.44%	Insertion - Frameshift	coding silent	Missense	Missense	Missense
CTCF	7.90%	3.81%	2.72%	2.45%	1.91%	**Insertion - Frameshift**	Unknown	Nonsense	**Deletion - Frameshift**	Missense
PRCC	6.96%	5.22%	2.61%	2.61%	2.61%	**Insertion - Frameshift**	**Deletion - Frameshift**	Missense	Missense	Missense
HNF1A	6.27%	6.27%	5.35%	4.28%	3.82%	**Insertion - Frameshift**	**Deletion - Frameshift**	Unknown	Missense	Missense
CLTCL1	5.81%	3.28%	3.03%	2.53%	1.26%	Insertion - Frameshift	Missense	Unknown	coding silent	coding silent
APC	5.13%	4.18%	3.54%	2.66%	1.87%	**Insertion - Frameshift**	Nonsense	Unknown	**Deletion - Frameshift**	Nonsense
CDK12	1.43%	1.19%	1.19%	0.71%	0.71%	**Insertion - Frameshift**	Unknown	Missense	**Deletion - Frameshift**	Deletion - Frameshift
RPL22	54.48%	4.83%	4.14%	2.07%	2.07%	**Deletion - Frameshift**	**Insertion - Frameshift**	Unknown	Deletion - Frameshift	Deletion - Frameshift
CALR	41.40%	26.41%	12.52%	7.08%	2.61%	**Deletion - Frameshift**	**Insertion - Frameshift**	Unknown	Unknown	Deletion - Frameshift
NFKBIE	36.77%	4.04%	3.59%	3.14%	2.24%	Deletion - Frameshift	Missense	Missense	Unknown	Missense
ACVR2A	28.28%	6.67%	2.07%	1.84%	1.61%	**Deletion - Frameshift**	Missense	Unknown	**Insertion - Frameshift**	Deletion - Frameshift
RNF43	21.28%	3.72%	1.52%	1.52%	1.18%	Deletion - Frameshift	Deletion - Frameshift	Nonsense	Deletion - Frameshift	Nonsense
QKI	15.57%	6.56%	2.46%	2.46%	1.64%	Deletion - Frameshift	Unknown	Missense	Missense	Missense
NOTCH1	13.03%	4.02%	2.05%	1.89%	1.65%	Deletion - Frameshift	Unknown	Deletion - Frameshift	Missense	Missense
CRTC1	12.38%	6.44%	3.47%	2.48%	1.49%	**Deletion - Frameshift**	Unknown	Missense	**Insertion - Frameshift**	Missense
B2M	11.81%	9.28%	3.80%	3.38%	2.95%	Deletion - Frameshift	Unknown	Missense	Complex - frameshift	Nonsense
MSH6	10.70%	6.84%	3.68%	1.05%	0.88%	**Deletion - Frameshift**	**Insertion - Frameshift**	Unknown	Missense	Nonsense
TGFBR2	10.59%	9.04%	7.49%	2.58%	2.33%	**Deletion - Frameshift**	Missense	Unknown	**Insertion - Frameshift**	Nonsense
AXIN2	9.11%	3.54%	3.29%	3.29%	2.28%	**Deletion - Frameshift**	coding silent	Unknown	Deletion - Frameshift	**Insertion - Frameshift**
CDX2	8.43%	3.61%	2.41%	2.41%	2.41%	Deletion - Frameshift	coding silent	Unknown	coding silent	coding silent
BCL10	8.33%	5.95%	3.57%	3.57%	2.38%	**Deletion - Frameshift**	Missense	**Insertion - Frameshift**	coding silent	Complex
CD79A	8.33%	7.14%	3.57%	2.38%	2.38%	Deletion - Frameshift	Deletion - In frame	Unknown	coding silent	Missense
TCF7L2	7.93%	3.45%	1.38%	1.38%	1.38%	Deletion - Frameshift	Unknown	Missense	Missense	Missense
UBR5	7.39%	4.26%	1.14%	0.85%	0.71%	**Deletion - Frameshift**	Unknown	**Insertion - Frameshift**	Unknown	Missense
BLM	6.73%	3.53%	1.92%	1.92%	1.28%	Deletion - Frameshift	Unknown	Missense	Missense	Deletion - Frameshift
BRD3	5.83%	2.43%	2.43%	2.43%	1.94%	Deletion - Frameshift	Unknown	Missense	Missense	Missense
TFE3	5.80%	4.35%	3.62%	2.90%	2.17%	**Deletion - Frameshift**	**Insertion - Frameshift**	Unknown	Missense	Missense
ZNF198	5.79%	3.72%	2.48%	1.24%	1.24%	Deletion - Frameshift	Unknown	Deletion - Frameshift	Missense	Missense
BIRC3	5.66%	5.28%	4.53%	1.89%	1.89%	Deletion - Frameshift	Unknown	Whole gene deletion	Missense	Deletion - Frameshift
MLLT1	4.84%	3.23%	2.42%	2.42%	1.61%	Deletion - Frameshift	Deletion - In frame	Unknown	Missense	Missense
BCORL1	4.60%	4.29%	3.65%	1.75%	0.95%	**Deletion - Frameshift**	Missense	Missense	Missense	**Insertion - Frameshift**
MNX1	4.23%	2.82%	2.82%	2.82%	2.82%	**Deletion - Frameshift**	**Insertion - In frame**	coding silent	Missense	Missense
KDM5A	4.16%	3.18%	0.98%	0.98%	0.98%	Deletion - Frameshift	Unknown	Missense	Missense	coding silent
NRG1	3.61%	1.97%	1.31%	1.31%	1.31%	Deletion - Frameshift	Unknown	Missense	Missense	Missense
MECOM	3.41%	2.69%	1.62%	1.62%	1.44%	Deletion - Frameshift	Unknown	Missense	coding silent	coding silent
BCL9L	3.23%	3.23%	2.76%	1.84%	1.15%	**Deletion - Frameshift**	Deletion - Frameshift	Deletion - Frameshift	**Insertion - Frameshift**	coding silent
SALL4	3.08%	1.18%	1.18%	1.18%	1.18%	Deletion - Frameshift	Missense	Missense	Missense	coding silent
PMS1	1.99%	1.49%	1.49%	1.49%	1.49%	Deletion - Frameshift	Unknown	coding silent	Missense	Missense

Mutation#1	#2	#3	#4	#5
p.W288fs*12	p.?	p.?fs	p.W288fs	p.W290fs*10
p.G646fs*12	p.?	p.E635fs*15	p.R693*	p.G646fs*?
p.A85fs*20	p.D4D	p.E18K	p.G211R	p.G251D
p.T204fs*26	p.?	p.R448*	p.T204fs*18	p.R377H
p.Q380fs*12	p.Q380fs*47	p.A487S	p.P213L	p.R482C
p.G292fs*25	p.P291fs*51	p.?	p.S487N	p.I27L
p.G1202fs*18	p.G656C	p.?	p.E743E	p.A1124A
p.T1556fs*3	p.R1450*	p.?	p.E1309fs*4	p.R876*
p.T1463fs*>29	p.?	p.R890H	p.G1271fs*23	p.N474fs*8
p.K15fs*5	p.K16fs*9	p.?	p.K15fs*9	p.K89fs*3
p.L367fs*46	p.K385fs*47	p.?	p.?fs*?	p.E364fs*46
p.Y254fs*13	p.G34E	p.G34R	p.?	p.G131E
p.K437fs*5	p.K315N	p.?	p.R438fs*19	p.D96fs*54
p.G659fs*41	p.R117fs*41	p.R132*	p.S661fs*39	p.R145*
p.K134fs*14	p.?	p.R319Q	p.R324S	p.D131Y
p.P2514fs*4	p.?	p.P2514fs*?	p.L1600P	p.L1678P
p.S572fs*6	p.?	p.T328A	p.E427fs*24	p.G524A
p.L15fs*41	p.?	p.M1V	p.S16fs*27	p.K95*
p.F1088fs*2	p.F1088fs*5	p.?	p.V509A	p.E1322*
p.K128fs*35	p.R528H	p.?	p.P129fs*3	p.R497*
p.G665fs*24	p.L688L	p.?	p.E19fs*19	p.N666fs*41
p.V306fs*2	p.F31F	p.?	p.G62G	p.T212T
p.I46fs*24	p.G213E	p.I46fs*4	p.L147L	p.A43>?
p.R131fs*61	p.191_206del	p.?	p.F19F	p.R107Q
p.K462fs*23	p.?	p.P325L	p.R465C	p.Y400C
p.E2121fs*28	p.?	p.E2121fs*13	p.E2702fs*13	p.R23G
p.N515fs*16	p.?	p.P868L	p.V1321I	p.D757fs*4
p.P24fs*24	p.?	p.A373T	p.H74Y	p.P214S
p.G482fs*44	p.P483fs*>94	p.?	p.A148S	p.E572K
p.K1044fs*33	p.?	p.N1250fs*3	p.E64K	p.E682K
p.Q547fs*21	p.?	p.0?	p.E169K	p.E429fs*7
p.Q461fs*46	p.S323delS	p.?	p.R206H	p.A17T
p.P1681fs*20	p.V1660G	p.F111L	p.L1132P	p.G1682fs*4
p.P18fs*204	p.A134_G135insA	p.A330A	p.A35S	p.E357K
p.G1200fs*9	p.?	p.P1237L	p.R696W	p.S993S
p.K91fs*24	p.?	p.A547T	p.L287F	p.L397P
p.G614fs*30	p.?	p.R750Q	p.S824S	p.C204C
p.P1127fs*21	p.P449fs*14	p.Q452fs*11	p.P1128fs*42	p.Q1041Q
p.V995fs*14	p.E689K	p.G505S	p.R586H	p.T620T
p.K164fs*6	p.?	p.G918G	p.R202K	p.S399F

Mutation information for these 40 cancer driver genes was acquired from the COSMIC database [111]. Mutation frequency, mutation type and detailed mutation are summarized for the top 5 most frequent mutations of each gene. It was interesting that frameshift mutations were the most frequent mutation type in these 40 genes and both insertions and deletions were found in most of these. Mutation frequency represents the frequency of this specific mutation in all the patients who carried any kinds of mutations in this gene.

Table A.2 Distribution of mutation types in genomic mutations of cancer driver genes

Mutation Type	Total Mutations	% of this Mutation Type
Substitution - Missense	352619	67.26%
Unknown	41887	7.99%
Substitution - coding silent	35475	6.77%
Substitution - Nonsense	22433	4.28%
Deletion - Frameshift	18026	3.44%
Deletion - In frame	17050	3.25%
Insertion - In frame	16021	3.06%
Insertion - Frameshift	13036	2.49%
Whole gene deletion	2303	0.44%
Complex - deletion inframe	2257	0.43%
Frameshift	1604	0.31%
Complex - frameshift	772	0.15%
Complex - insertion inframe	429	0.08%
Complex - compound substitution	165	0.03%
Complex	102	0.02%
Nonstop extension	98	0.02%

Genomic mutations for all human cancer driver genes (total 591) were downloaded from COSMIC. Overall distribution of mutation types was summarized in the table, missense point mutations were the most dominant mutation type with 67% of all genomic mutations, while frameshift INDELs only accounted for 6% of all mutations.

APPENDIX B

THE NEXT VERSION OF DOG CANCER VACCINE

Introduction

The first dog cancer vaccine was described in Chapter 5 and the selection criteria did not include gene type information. It is known that if a gene is a cancer driver gene, the gene is more frequently mutated in cancer patients. And if a gene has high expression level in the earliest tumor initiation stage, it is going to make more errors at the transcription and RNA splicing level and thus more frameshift antigens. In addition, a frameshift antigen will be more difficult for cancer cells to get rid of if it is from an essential gene.

Therefore, in the second version of dog cancer vaccine, I added three more criteria: cancer driver genes, essential genes and highly expressed genes.

A list of cancer driver genes was adapted from the COSMIC (catalogue of somatic mutations in cancer) cancer consensus genes [112]. The list of essential genes was acquired from 5 studies of human essential genes and genes were included if they were found in 3 studies or more. The list of highly expressed genes was acquired from the RNA-Seq Atlas [113]. Gene expression from all the normal tissues was averaged and standard deviations were calculated. The gene expression values of normal tissues were used since cancer initiates from normal tissues and the earliest form of tumor should have similar expression values as normal tissues.

Three tables were summarized for the second version dog cancer vaccine. Table B.1 summarizes microsatellite candidates from genes that meet all three criteria: cancer driver gene, essential gene and highly expressed gene. Table B.2 summarizes the list of microsatellite candidates which have long MS regions and meet any one criteria of the

three. Table B.3 summarizes the list of MS candidates which have high antibody responses in 116 dog cancer samples from 9 types of dog cancers as discussed in Chapter 5.

The second version of dog cancer vaccine included 21 candidates with microsatellite candidates from these three tables.

Table B.1 List of Microsatellite Candidates from genes which meets all three criteria: cancer driver genes, essential genes and high expression genes

GENE NAME	PROTEIN NAME	MS Type	MS Length	INS/DEL	FS LENGTH	Essential	Driver	High Exp
SPEN	msx2-interacting protein	CCCCCCCCC	9	Del	246	Y	Y	Y
ATRX	transcriptional regulator ATRX isoform X4	AAAAAAAA	8	Ins	18	Y	Y	Y
CHD4	chromodomain-helicase-DNA-binding protein 4 isoform X7	AAAAAAA	8	Del	52	Y	Y	Y
ATRX	transcriptional regulator ATRX isoform X4	AAAAAAA	7	Ins	22	Y	Y	Y
ATRX	transcriptional regulator ATRX isoform X4	AAAAAAA	7	Del	29	Y	Y	Y
ATRX	transcriptional regulator ATRX isoform X4	AAAAAAA	7	Del	27	Y	Y	Y
CTCF	transcriptional repressor CTCF isoform X3	AAAAAAA	7	Ins	23	Y	Y	Y
SF3B1	splicing factor 3B subunit 1 isoform X2	TTTTTTT	7	Del	18	Y	Y	Y
ARID1A	AT-rich interactive domain-containing protein 1A isoform X5	CCCCCCC	7	Ins	17	Y	Y	Y
CHD4	chromodomain-helicase-DNA-binding protein 4 isoform X6	AAAAAAA	7	Ins	26	Y	Y	Y
CHD4	chromodomain-helicase-DNA-binding protein 4 isoform X7	AAAAAAA	7	Ins	33	Y	Y	Y
SPEN	msx2-interacting protein	GGGGGGG	7	Del	30	Y	Y	Y
USP8	ubiquitin carboxyl-terminal hydrolase 8	AAAAAAA	7	Ins	24	Y	Y	Y
USP8	ubiquitin carboxyl-terminal hydrolase 8	AAAAAAA	7	Del	17	Y	Y	Y
SMARCB1	natrix-associated actin-dependent regulator of chromatin subfamily B m	AAAAAAA	7	Del	17	Y	Y	Y
MYH9	myosin-9	CCCCCCC	7	Ins	26	Y	Y	Y
STAG2	cohesin subunit SA-2 isoform X1	TTTTTTT	7	Del	26	Y	Y	Y
ATRX	transcriptional regulator ATRX isoform X4	AAAAAAA	7	Del	40	Y	Y	Y
ATRX	transcriptional regulator ATRX isoform X4	AAAAAAA	7	Del	42	Y	Y	Y
CTCF	transcriptional repressor CTCF isoform X3	AAAAAAA	7	Del	42	Y	Y	Y
GMPS	GMP synthase [glutamine-hydrolyzing	AAAAAAA	7	Del	36	Y	Y	Y
RANBP2	E3 SUMO-protein ligase RanBP2 isoform X3	TTTTTTT	7	Del	35	Y	Y	Y
SETD2	histone-lysine N-methyltransferase SETD2 isoform X2	AAAAAAA	7	Del	34	Y	Y	Y
SETD2	histone-lysine N-methyltransferase SETD2 isoform X2	AAAAAAA	7	Del	35	Y	Y	Y
STK11	serine/threonine-protein kinase STK11 isoform X1	CCCCCCC	7	Del	50	Y	Y	Y
STK11	serine/threonine-protein kinase STK11 isoform X2	CCCCCCC	7	Ins	50	Y	Y	Y
NUMA1	nuclear mitotic apparatus protein 1 isoform X1	AAAAAAA	7	Ins	51	Y	Y	Y
TPR	nucleoprotein TPR	TTTTTTT	7	Del	64	Y	Y	Y
ARID1A	AT-rich interactive domain-containing protein 1A isoform X5	GGGGGGG	7	Del	73	Y	Y	Y
ARID1A	AT-rich interactive domain-containing protein 1A isoform X4	CCCCCCC	7	Ins	173	Y	Y	Y
CHD4	chromodomain-helicase-DNA-binding protein 4 isoform X7	AAAAAAA	7	Del	83	Y	Y	Y
CHD4	chromodomain-helicase-DNA-binding protein 4 isoform X6	AAAAAAA	7	Del	76	Y	Y	Y
SPEN	msx2-interacting protein	GGGGGGG	7	Ins	94	Y	Y	Y
SPEN	msx2-interacting protein	CCCCCCC	7	Del	61	Y	Y	Y
USP8	ubiquitin carboxyl-terminal hydrolase 8	AAAAAAA	7	Del	82	Y	Y	Y
EWSR1	RNA-binding protein EWS isoform X2	CCCCCCC	7	Del	52	Y	Y	Y
MDM4	protein Mdm4 isoform X1	AAAAAAA	7	Del	53	Y	Y	Y

Expression Average	Expression Standard deviation	mRNA Position
10.47	3.92	4.78%
19.11	8.34	62.67%
12.43	2.94	88.13%
19.11	8.34	88.89%
19.11	8.34	32.57%
19.11	8.34	88.89%
9.62	2.61	27.66%
41.29	12.53	57.55%
9.47	4.03	34.63%
12.43	2.94	3.72%
12.43	2.94	3.72%
10.47	3.92	0.13%
7.30	3.03	6.57%
7.30	3.03	15.46%
7.90	3.55	17.96%
22.60	14.74	7.46%
7.87	4.01	36.62%
19.11	8.34	51.54%
19.11	8.34	41.18%
9.62	2.61	27.66%
12.48	9.62	17.79%
9.06	3.24	84.64%
17.34	6.76	56.87%
17.34	6.76	12.10%
7.62	3.07	68.14%
7.62	3.07	68.43%
8.66	2.65	3.72%
13.80	6.10	90.14%
9.47	4.03	64.07%
9.47	4.03	9.82%
12.43	2.94	3.72%
12.43	2.94	3.72%
10.47	3.92	0.13%
10.47	3.92	34.88%
7.30	3.03	6.57%
24.01	6.94	36.58%
5.53	2.24	77.42%

Table B.2 List of MS candidates with long microsatellite regions and meets one criteria

GENE NAME	PROTEIN NAME	MS Type	MS Length	INS/DEL	FS LENGTH	Essential	Driver	High Exp
MORF4L1	mortality factor 4-like protein 1	TTTTTTTTTTTTTTTTTTTTTTTT	25	Del	21			Y
MED13	mediator of RNA polymerase II transcription subunit 13	TTTTTTTTTTTTTTT	16	Ins	31			Y
GOSR2	Golgi SNAP receptor complex member 2 isoform X3	CCCCCCCCCCCCCCCC	16	Del	23	Y		
PLD1	phospholipase D1	GGGGGGGGGGGGGGG	15	Del	36			Y
PLD1	phospholipase D1	GGGGGGGGGGGGGGG	15	Ins	77			Y
IRF4	interferon regulatory factor 4	CCCCCCCCCCCCCCC	15	Ins	46		Y	
IRF4	interferon regulatory factor 4	CCCCCCCCCCCCCCC	15	Del	64		Y	
INTS3	integrator complex subunit 3	CCCCCCCCCCCCCC	14	Del	56	Y		Y
PWWP2A	PWWP domain-containing protein 2A	GGGGGGGGGGGGGG	14	Ins	21		Y	
NSMCE4A	non-structural maintenance of chromosomes element 4 homolog A	CCCCCCCCCCCCCC	14	Del	48	Y		
UBR1	E3 ubiquitin-protein ligase UBR1 isoform X4	AAAAAAAAAAAAA	13	Ins	30			Y
RBP1	retinol-binding protein 1	CCCCCCCCCCCCC	13	Del	44			Y
RBM10	RNA-binding protein 10	CCCCCCCCCCCCC	13	Ins	38	Y	Y	
RBM10	RNA-binding protein 10	CCCCCCCCCCCCC	13	Del	58	Y	Y	
NSMCE4A	non-structural maintenance of chromosomes element 4 homolog A	TTTTTTTTTTTT	13	Ins	19	Y		
XRN1	5'-3' exoribonuclease 1 isoform X5	AAAAAAAAAAAA	12	Ins	17	Y		Y
XRN1	5'-3' exoribonuclease 1 isoform X6	AAAAAAAAAAAA	12	Ins	19	Y		Y
EIF4G2	eukaryotic translation initiation factor 4 gamma 2	GGGGGGGGGGGG	12	Del	169	Y		Y
SBNO1	protein strawberry notch homolog 1 isoform X3	AAAAAAAAAAA	11	Ins	23	Y		Y
SOS1	son of sevenless homolog 1	GGGGGGGGGGG	11	Del	42	Y		Y
SEC62	translocation protein SEC62	AAAAAAAAAAA	11	Del	58	Y		Y
ZCCHC6	terminal uridylyltransferase 7 isoform X4	AAAAAAAAAAA	11	Ins	24			Y
MARCKS	myristoylated alanine-rich C-kinase substrate	AAAAAAAAAAA	11	Ins	88			Y
MTA3	metastasis-associated protein MTA3	CCCCCCCCCCC	11	Del	83			Y
TSPYL1	testis-specific Y-encoded-like protein 1	CCCCCCCCCCC	11	Del	98			Y
ARID2	AT-rich interactive domain-containing protein 2 isoform X1	CCCCCCCCCCC	11	Del	37		Y	
E2F3	transcription factor E2F3 isoform X2	AAAAAAAAAAA	11	Del	18	Y		
DTYMK	thymidylate kinase	GGGGGGGGGGG	11	Ins	78	Y		
DTYMK	thymidylate kinase	GGGGGGGGGGG	11	Del	81	Y		
DDX24	ATP-dependent RNA helicase DDX24	AAAAAAAAAA	10	Ins	20	Y		Y
DDX24	ATP-dependent RNA helicase DDX24	AAAAAAAAAA	10	Del	29	Y		Y
MBNL1	muscleblind-like protein 1 isoform X9	TTTTTTTTTT	10	Ins	41	Y		Y
ACTB	actin, cytoplasmic 1 isoform X1	CCCCCCCCCC	10	Del	36	Y		Y
ANKRD12	ankyrin repeat domain-containing protein 12 isoform X3	AAAAAAAAAA	10	Del	24			Y
ECHDC1	ethylmalonyl-CoA decarboxylase isoform X2	AAAAAAAAAA	10	Del	23			Y
SLC35F5	solute carrier family 35 member F5 isoform X2	TTTTTTTTTT	10	Del	22			Y
SLC35F5	solute carrier family 35 member F5 isoform X2	TTTTTTTTTT	10	Ins	23			Y
HK1	hexokinase-1 isoform X3	AAAAAAAAAA	10	Del	21			Y
HK1	hexokinase-1 isoform X3	AAAAAAAAAA	10	Ins	33			Y
TCF7L2	transcription factor 7-like 2	CCCCCCCCCC	10	Del	17		Y	
TTF1	transcription termination factor 1 isoform X2	AAAAAAAAAA	10	Ins	18	Y		
EIF2B3	translation initiation factor eIF-2B subunit gamma isoform X2	AAAAAAAAAA	10	Del	57	Y		

Gene Expression Average	Standard deviation	mRNA Position
40.43	14.79	31.49%
5.71	1.68	2.41%
		4.21%
5.48	4.49	83.26%
5.48	4.49	83.26%
		19.95%
		19.95%
7.42	2.60	17.00%
		24.94%
		3.85%
5.59	2.00	4.12%
17.12	31.18	10.91%
		24.31%
		24.31%
		17.00%
5.19	2.12	97.63%
5.19	2.12	97.63%
50.26	13.51	0.48%
7.19	5.23	56.52%
5.46	2.39	2.31%
12.08	5.61	34.92%
8.48	4.33	3.37%
6.48	4.24	44.91%
5.62	3.76	9.26%
9.36	6.77	19.47%
		0.94%
		12.54%
		53.66%
		53.66%
25.49	13.26	18.41%
25.49	13.26	18.41%
14.17	7.88	14.58%
83.43	45.89	67.49%
10.45	4.75	35.93%
11.70	6.77	12.25%
6.66	3.83	46.73%
6.66	3.83	46.73%
12.25	10.25	15.18%
12.25	10.25	15.18%
		77.39%
		8.22%
		33.18%

Table B.3 List of MS candidates with high antibody response in dog cancer samples and meets one criteria

					Positive Rate	Positive Rate
					Normal	Cancer Overall
Candidate	Gene	Driver	Essential	High Exp	52	116
345803468_ins	RAB12			Y	9.62%	24.14%
345790242_del	PDS5B			Y	13.46%	23.28%
345801690_del	SETD1A		Y		7.69%	23.28%
50979053_ins	TCOF1		Y		13.46%	22.41%
94962361_del	TCOF1		Y		11.54%	22.41%
359319099_del	SPEN	Y	Y	Y	11.54%	21.55%
345796542_del	SEC62		Y	Y	9.62%	19.83%
345796190_ins	CCDC80			Y	11.54%	19.83%
50979053_del	TCOF1		Y		11.54%	19.83%
359320226_del	HGS		Y		3.85%	18.97%
345777441_del	RUFY1			Y	11.54%	18.10%
345794707_del	PRDM2			Y	5.77%	17.24%
73957532_del	TCOF1		Y		11.54%	17.24%
345797019_del	SP3			Y	7.69%	16.38%
345790842_ins	DTYMK		Y		9.62%	16.38%
359320226_ins	HGS		Y		9.62%	15.52%
73977514_ins	TCOF1		Y		1.92%	15.52%
345798736_ins	RHOBTB3			Y	7.69%	14.66%
345797019_ins	SP3			Y	5.77%	14.66%
345787800_del	SWAP70			Y	3.85%	14.66%
CDC7_9a_de	CDC7		Y		9.62%	14.66%
359320691_del	HNRNPH1		Y	Y	7.69%	13.79%
TGFβR II	TGFBR2	Y		Y	15.38%	13.79%
345803468_del	RAB12			Y	11.54%	13.79%
345803795_del	SERPINA3			Y	9.62%	13.79%
345781087_ins	TMPO			Y	7.69%	13.79%
73992592_del	TCOF1		Y		3.85%	13.79%
345798777_ins	RHOBTB3			Y	1.92%	12.93%
345804985_del	NSRP1		Y		3.85%	12.93%
16605494_del	TCOF1		Y		5.77%	12.93%
73996466_del	TCOF1		Y		3.85%	12.93%
345796984_del	IRF4	Y			3.85%	12.93%
345803795_ins	SERPINA3			Y	5.77%	12.07%
345794468_ins	MFN2		Y		9.62%	12.07%
345796987_del	PRPF4B		Y	Y	11.54%	11.21%
345787800_ins	SWAP70			Y	7.69%	11.21%
345797136_ins	ZAK			Y	5.77%	11.21%
345794365_del	ZCCHC17			Y	7.69%	11.21%

APPENDIX C

INSTITUTIONAL ANIMAL CARE AND USE COMMITTEE (IACUC) APPROVAL

Institutional Animal Care and Use Committee (IACUC)
Office of Research Integrity and Assurance
Arizona State University
660 South Mill Avenue, Suite 312
Tempe, Arizona 85287-6111
Phone: (480) 965-6788 *FAX:* (480) 965-7772

Animal Protocol Review

ASU Protocol Number:	**17-1568R**
Protocol Title:	**Cancer vaccine and immunotherapy**
Principal Investigator:	**Stephen Johnston**
Date of Action:	**3/24/2017**

The animal protocol review was considered by the Committee and the following decisions were made:

The protocol was approved.

If you have not already done so, documentation of Level III Training (i.e., procedure-specific training) will need to be provided to the IACUC office before participants can perform procedures independently. For more information on Level III requirements see https://researchintegrity.asu.edu/training/animals/levelthree.

Total # of Animals:	**23,788**	
Species:	**Mice**	**Pain Category: C-13,386; D-10,402**
Protocol Approval Period:	**3/29/2017 – 3/28/2020**	
Sponsor:	**N/A**	
ASU Proposal/Award #:	**N/A**	
Title:	**N/A**	

Signature: _____ for C. Miller
IACUC Chair or ***Designee***

Date: 3/31/2017

Cc: IACUC Office
IACUC Chair

PERSONNEL CHART

ASU requires that all personnel engaged in animal research or teaching be qualified through training or experience in order to conduct the work humanely. The IACUC requires the following training:

- **Level I Basic** – Required of ALL participants (must be renewed every 4 years)
- **Level II Species-Specific** – Required for each participant that will have direct contact with that species (must be renewed every 4 years)
- **Level III Hands-on Training** – Required to perform specific procedures independently; Level III Certification form must be submitted to the IACUC office by the person providing the training within 5 days of the training

You can access the training modules at http://hazel.forest.net/latanet/client/asu/introduction.htm. See the IACUC web site (http://researchintegrity.asu.edu/training/useofanimals) for more information on training and Level III forms.

Name	Title	ASURITE name	Role in Protocol		Species with which individual will have direct contact ("none, "all", or list species)	FOR IACUC USE ONLY
			What procedures will each person be doing on live animals under supervision?)	What procedures will each person be doing on live animals independently (without supervision)?		Training Confirmation
Stephen A. Johnston, Ph.D.	PI	sajohnst	Project oversight, no animal handling			Basic 6/2013 OHSP not required
Debra T. Hansen, Ph.D.	Associate Research Professor	dhansen6		Immunizations, blood collection, euthanasia	mouse	4/2016 OHSP
Luhui Shen, Ph.D.	Assistant Research Scientist	lshen8		Immunizations, blood collection, euthanasia, anesthetization, mouse breeding, tumor cell injection, tumor monitoring	mouse	1/2016 OHSP
Jian Zhang	Graduate Student	jzhang185		Immunizations, blood collection, euthanasia, anesthetization, mouse breeding, tumor cell injection, tumor monitoring	mouse	3/2017 OHSP
Milene Tavares Batista	Visiting Researcher	mtavare1		Immunizations, blood collection, euthanasia, anesthetization, mouse breeding, tumor cell injection, tumor monitoring	mouse	Basic 12/2016 Mouse 11/2016 OHSP

For each individual, describe the individual's years of experience with all listed species and procedures they will be conducting under this protocol. For procedures for which they are not yet trained, but will likely be trained to do during the activity period of this protocol, provide a description of who will provide such training:

Stephen A. Johnston: Will be providing project oversight and not handling mice.

Debra T. Hansen: Has 5 years' experience injecting and conducting phlebotomy in mice.

Revised 7/23/2013