UNIVERSITY OF CALIFORNIA, SAN DIEGO

Polymorphic Pseudogenization of *SIGLEC12* in Humans: Relationship to Late Stage Cancer Progression

A thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Raymond Do

Committee in charge:

Professor Nissi Varki, Chair
Professor Douglass Forbes, Co-Chair
Professor Gen-Sheng Feng

2017

The Thesis of Raymond Do is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2017

TABLE OF CONTENTS

LIST OF FIGURES

EPIGRAPH

Knowing is not enough; we must apply. Willing is not enough; we must do.

-Johann Wolfgang von Goethe

DEDICATION

I would like to dedicate this thesis to my mother and father, for theirs is a story of dedication and compassion, and boundlessly so.

ACKNOWLEDGEMENTS

I would like to thank Dr. Nissi Varki, who has been an important mentor. Her experiences and counsel have played a formant role in my development as a student, a researcher, and as an individual. She has always been willing to share her many experiences in medicine and research, and is a kind and generous person with unwavering character. Dr. Varki is a dedicated educator and researcher who has a penchant for awakening and fostering scientific curiosity amongst her students. Her intellectual tenacity has been a powerful inspiration for critical and rigorous scientific thinking in many generations of students, including myself.

I would also like to thank Dr. Douglass Forbes, who has also been an important mentor. Dr. Forbes is an exceptional educator whose teaching style and passion for biology first inspired me to go above and beyond curriculum. She inspired me to not only appreciate scientific research as an onlooker, but to actively contribute to the ever-growing knowledge base that represents our collective scientific memory. Her willingness to share her experiences and passion for biology with her pupils has played an important role in shaping the futures of many generations of students to come.

Finally, I would like to thank Dr. Shoib Siddiqui, who has shared his experience and knowledge in research with me. I could not have completed this thesis without his kind guidance and support.

ABSTRACT OF THE THESIS

Polymorphic Pseudogenization of *SIGLEC12* in Humans:
Relationship to Late Stage Cancer Progression

by

Raymond Do

Master of Science in Biology

University of California, San Diego, 2017

Professor Nissi Varki, Chair
Professor Douglass Forbes, Co-Chair

The primate *SIGLEC12* gene encodes one of the CD33-related Siglec family of signaling molecules in immune cells. Siglecs are normally expressed on hematopoietic cells. Interestingly, it was previously reported that the human *SIGLEC12* gene product Siglec-XII is expressed on some epithelium and also on prostate carcinomas. Stably transfected Siglec-XII-expressing prostate cancer cells demonstrated altered regulation of several genes associated with carcinoma

progression, as well as enhanced growth in nude mice. However, *SIGLEC12* allele status did not predict prostate carcinoma incidence. Here, we confirm and expand on previous immunohistochemical analysis of Siglec-XII expression in human epithelium, and found that Siglec-XII expression correlates with an increased risk of late stage carcinoma development. While no significant correlation was found between Siglec-XII expression and incidence of early stage or slow progressing cancers in two patient cohorts, we found that Siglec-XII expression is correlated with a significantly increased rate of late stage carcinoma progression in a late stage colorectal carcinoma patient cohort. We also found through RNA-Seq analysis that several genes associated with cancer progression are differentially expressed in a stably transfected PC-3 Siglec-XII expressing cell line. Finally, we propose a unique new noninvasive test, a dot blot using urine samples which may become a screening test for the efficient determination of an individual's Siglec-XII status. This test may thus become useful as a diagnostic screen in the future. Thus, polymorphic expression of Siglec-XII in humans may determine progression of certain late stage carcinomas.

INTRODUCTION

Sialic acids are N or O substituted derivatives of nonulsonic acid, a monosaccharide with a nine-carbon backbone. Sialic acids often compose components of oligosaccharide chains of mucins, glycoproteins and glycolipids. Most often, they occupy terminal positions of glycan chains in Deuterostomes and are potentially salient for recognition events. (Angata et al. 2002, Schauer et al. 2009, Chen et al. 2010). Siglecs (sialic acid-binding immunoglobulin-like lectins) are a class of intrinsic Sia recognition proteins in vertebrates. Siglecs are single pass transmembrane proteins with one or more extracellular C2-set Immunoglobulin-like domains and a Sia-binding site in the extracellular N-terminal Ig-like V-set domain (Varki et al. 2006, Crocker et al. 2007, von Gunten et al. 2008, Lopez et al. 2008, Varki et al. 2009, Cao et al. 2009). The majority of known Siglecs contain an immunoreceptor tyrosine-based inhibitory motif or ITIM in the cytosolic region which serves to inhibit immune cell signaling. Other Siglecs contain an immunoreceptor tyrosine-based activation motif or ITAM associated with an arginine motif. These domains and motifs are features shared by all CD33-related Siglecs.

The CD33-related Siglecs are encoded by a subset of *SIGLEC* genes clustered on chromosome 19 in humans and chimpanzees, and share high sequence homology with one another. CD33-related Siglecs are typically expressed on immune cells (Cornish et al. 1998). Comparative analysis of genomic *SIGLEC* sequences across humans, chimpanzees, baboons, rats and mice has shown that CD33-related Siglecs are undergoing rapid evolution (Angata et al. 2004).

Interestingly, the genomic analysis suggests that the Sia-recognizing V-set domain is being subjected to the greatest selective pressure.

This study focuses on Siglec-XII. It was previously shown that human Siglec-XII is unable to bind to Sias due to an Arg to Cys (R122C) substitution mutation which affects all humans (Angata et al. 2001). While the gene in all primates is referred to as *SIGLEC12*, the protein is referred to as Siglec-XII in humans to differentiate it from Siglec-12 in other primates, which retain their Sia-binding arginine. Reversing this mutation in vitro restored Sia binding, which demonstrated that the Sia-binding domain and reading frame remain otherwise intact (Angata et al. 2001). It was also shown that a second missense mutation results in complete inactivation of human *SIGLEC12* with a global allele frequency of 58% (Mitra et al. 2011).

Analysis of human epithelial tissues has shown that Siglec-XII protein is expressed on various epithelial cell surfaces, and on the cell surfaces of carcinomas (cancers of epithelial origin). Siglec-XII protein was also found to be strongly expressed on prostate carcinomas, and carcinoma cell lines, but *SIGLEC12* allele status did not predict prostate carcinoma incidence (Mitra et al. 2011). Interestingly, restoration of expression in a prostate carcinoma cell line homozygous for the frame shift mutation induced altered regulation of several genes associated with carcinoma progression. These stably transfected Siglec-XII-expressing prostate cancer cells also showed enhanced growth in nude mice (Mitra et al. 2011). Given the limited analysis of Siglec-XII expression in human epithelium, and in human carcinomas,

and the enhanced growth of the Siglec-XII transfected prostate cancer cell line, we thought it important to explore the effect of Siglec-XII on human carcinomas further.

Here we ask whether Siglec-XII expression is a prognostic factor for carcinomas in humans. Using monoclonal antibodies against Siglec-XII, we confirmed and extended earlier work which noted unexpected expression in epithelial cells and carcinomas. We also analyze three cohorts consisting of patients with low lifestyle risks for cancer, patients from the prostate cancer cohort originally analyzed in a previous *SIGLEC12* paper, and colon cancer patients participating in a phase III clinical trial, to determine whether Siglec-XII expression is correlated with carcinoma progression. Finally, we propose a method to determine the *SIGLEC12* genotype of an individual using a modified dot blot technique.

METHODS

Mouse Monoclonal Antibody against Human Siglec-XII

A fusion protein Siglec-XII-Fc including the first three Ig-like domains of human Siglec-XII and the human IgG Fc domain was prepared as described (Angata et al.). The fusion protein was used to immunize mice to generate monoclonal antibodies by BD Pharmingen. Two final clones 1130 and 276 were obtained. Specificity was confirmed by lack of cross-reactivity with Siglec-7-Fc. Studies were done using a mixture of the two clones or clone 276 or 1130 alone.

Immunohistochemistry Studies

Anonymized multi-tissue array slides were obtained from Biomax based in Rockville, Maryland. One multi-tissue array consisted of normal tissue while the other consisted of cancerous tissue with a few normal control tissue. The paraffin sections were deparaffinized, rehydrated, and blocked for endogenous peroxidases and endogenous biotin. Epitopes were revealed using heat-induced antigen retrieval at pH 6.0 in citrate buffer followed by a 5 step signal amplification method. The slides were incubated with the mouse anti-Siglec-XII antibodies, biotinylated donkey anti-mouse antibodies, horseradish peroxidase, biotinyl tyramide, and again with horseradish peroxidase. Finally, bound antibody was detected using the AEC kit from Vector following the manufacturer's instructions. Digital photographs were taken using a Keyence microscope. The images were organized using Adobe Photoshop.

Buccal Swab

Samples were obtained from healthy volunteers from the lab. Donors were asked to avoid oral intake for 1 hour before collection. Immediately before collection, a 3x3 inch sterile gauze pad was used to remove mucous from the right side of the inner cheek. Mucous removal was considered adequate once the donor confirms that the area is dry to touch. A sterile cotton tip applicator was used to abrade the dried area using an oscillating motion for 60 seconds. The cotton tips were immediately collected for DNA isolation. Genomic DNA was isolated from the samples using a ChargeSwitch Buccal Cell gDNA kit (Thermofisher) following the manufacturer's instructions. The isolated gDNA was amplified using the primers 5′-ACCCCTGCTCTGTGGGAGAGT-3′ (forward) and 5′AGGATCAGGAGGGGCATCCAAGGTGC-3′ (reverse) to amplify a 570-bp region using a Phusion High Fidelity Polymerase kit (Thermofisher) and following the manufacturer's instructions. The PCR conditions were 98 °C for 3 min after a “hot start” (94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s) for 35 cycles. The amplified DNA was purified using a Qiagen PCR purification kit following the manufacturer's instructions and sequenced using the forward primer. Sequencing was performed by Eton Biosciences based in San Diego, California.

Urine Dot Blot Procedure for Detection of Siglec-XII Protein Expression

50 mL urine samples were obtained from healthy volunteers from the lab. As this protocol was designed as a diagnostic test, the urine samples were refrigerated for 72 hours to simulate transport and storage time for patient samples sent for

testing. The urine samples were centrifuged at 500XG for 30 minutes to pellet any solid material in the urine. This spin speed was chosen to reduce the chance of damage to any cells in the samples. The supernatant was aspirated and the pellet was resuspended in 100ul of sterile PBS. A nitrocellulose membrane was washed with 10 ml of PBST. 100µl of each sampled will be pipetted onto the nitrocellulose membrane at the center of the grid, and was applied under negative vacuum pressure. Non-specific sites will be blocked by applying Licor solution-50%+50%PBST for 60 minutes at room temperature. The membrane was incubated with the primary antibody, clone 1130 of a monoclonal mouse anti-human-Siglec-XII antibody dissolved in a 10% Licor solution + 90%PBST min at RT for 1 hour. The membranes were then washed with 10ml PBST (3 x 5 min). This was then followed by incubation with the secondary antibody Licor anti-mouse-680 used at a 1:10000 dilution for 60 min at room temperature. The secondary antibody was dissolved in 10% Licor Solution + 90% PBST. The membrane was washed with 10ml of PBST (5 min x 1), then with PBS (5 min X 3 times). The membrane was then visualized using a Licor fluorescence scanning machine, and photographs were taken using the associated Licor visualization program.

Analysis of Frameshift Mutation in *SIGLEC12* locus on Adventist Healthcare Cancer and Control Groups

The frameshift mutation polymorphism was analyzed in 53 genomic DNA samples from patients with a history of cancer and on 54 control DNA samples from patients with no history of cancer. These genomic samples were isolated from whole

blood using the QIAamp DNA blood maxi kit (Qiagen). The lifestyle of these individuals included reduced consumption of red meat, no consumption of alcohol and no use of tobacco, which are all risk factor associated with cancer (Aykan et al. 2015, Bagnardi et al. 2015, Boyle et al. 2003, Doll et al. 1950). To determine the genomic status of each individual, the primers 5′-ACCCCTGCTCTGTGGGAGAGT-3′ (forward) and 5′AGGATCAGGAGGGGCATCCAAGGTGC-3′ (reverse) were used to amplify a 570-bp region using a Phusion High Fidelity Polymerase kit. The PCR conditions were 98 °C for 3 min after a “hot start” (94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s) for 35 cycles. The amplified DNA was purified using a Qiagen PCR purification kit and sequenced using the forward primer. Sequencing was performed by Eton Biosciences based in San Diego, California. The allele frequencies (presence/absence of frameshift) among the groups was compared using the chi-squared test. The odds ratio and its 95% confidence interval were estimated by unconditional logistic regression as a measure of the associations between genotypes and cancer risk, using R statistical software. Statistical tests were two-sided, and significance was set at $p < 0.05$.

RNA-Seq Analysis of Siglec-XII Stable Transfected PC-3 Prostate Cancer Cell Line

mRNA was isolated from a stably transfected Siglec-XII expressing PC-3 cell line and a Siglec-XII non-expressing PC-3 cell line using a Qiagen RNA Extraction kit following the manufacturer’s instructions. The mRNA was purified using an Invitrogen Oligo dT kit following the manufacturer’s instructions. The samples were

incubated with oligo-dT beads prepared according to the manufacturer's instructions on a thermal cycler for 65 °C for 5 minutes and on hold at 4 °C until removed. The samples were then washed and eluted following the manufacturer's instructions. FastQC was used to perform quality control checks on the RNA-Seq data. R and Bioconductor were then used to perform most of the analyses. The sequence files were read and the sequence reads mapped to a reference genome. Gene expression values were estimated by calculating either the raw read count number or Reads Per Kilobase of transcript per Million reads mapped (RPKM).

Statistical Analysis of Cohorts: Test of Association in Cohort Studies

Analysis was performed with a chi-square (X^2) statistical test. Categorical data was organized in contingency tables. Observed counts in each cell of the contingency table were compared to expected counts under the assumption of no association between the row and column categories or classifications. The chi-square statistic was calculated using the following equation, where O_i is the observed frequency in the i^{th} cell of the table, and where E_i is the expected frequency in the i^{th} cell of the table.

$$c^2 = \sum_{i=1}^{k} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

Degrees of freedom are calculated as (r-1)(c-1) where "r" is the number of rows and "c" is the number of columns. A second round of analysis was performed for each of the cohorts which combined the Siglec-XII expressing individuals into a single category, which consisted of the categories of *SIGLEC12* heterozygous, and

homozygous wildtype genotypes. The probability of observing each calculated X^2 statistic was determined by referencing a table listing the probability, or area in upper tail of the distribution, associated with the value of the chi-squared statistic for each degree of freedom.

RESULTS

Epithelial Siglec-12 Expression in Human Epithelium (Normal Tissue)

In a 2011 SIGLEC12 paper, Mitra et al. showed, using a mouse monoclonal antibody, that strong expression of Siglec-XII was seen in macrophages and also in chimpanzee prostate, pancreas, kidney, and stomach epithelium. This expression correlated with the presence of the non-frameshifted, intact *SIGLEC12* allele. It was previously shown that cell pellets of 293T cells transfected with full-length Siglec-XII were detected and that an empty vector was not detected by this monoclonal antibody. We now confirm and extend this analysis by looking for expression of Siglec-XII in a multi-tissue array consisting of about 100 paraffin sections of normal human tissue sections using immunohistochemical methods. Immunohistochemical analysis was performed by comparing nonspecific detection with a mouse IgG antibody to detection by a monoclonal mouse anti-Siglec-XII antibody (clone 276). Strong expression of Siglec-XII was seen in a subset of all epithelial cell types including squamous epithelium, columnar epithelium and cuboidal epithelium (Fig 1C). Detection of Siglec-XII expression was achieved with a chromogen which produced a red substrate. As expected, not all normal human tissue showed expression because of the frameshift insertion mutation present in a subset of the human population (Fig 1B). However, a higher frequency of detection of Siglec-XII was seen in cuboidal epithelium and squamous epithelium, including renal tubules and bladder respectively (Fig 1A). The frequency of detection of Siglec-XII expression in human epithelium overall was 34.18% (Fig 1D). This is consistent with the global WT allele frequency in humans previously measured (Mitra et al. 2011).

This suggests that the population of individuals who provided the tissue samples are representative of the overall human population. These data confirm and extend an earlier observation using the same monoclonal antibody that human Siglec-XII is expressed on epithelial cells.

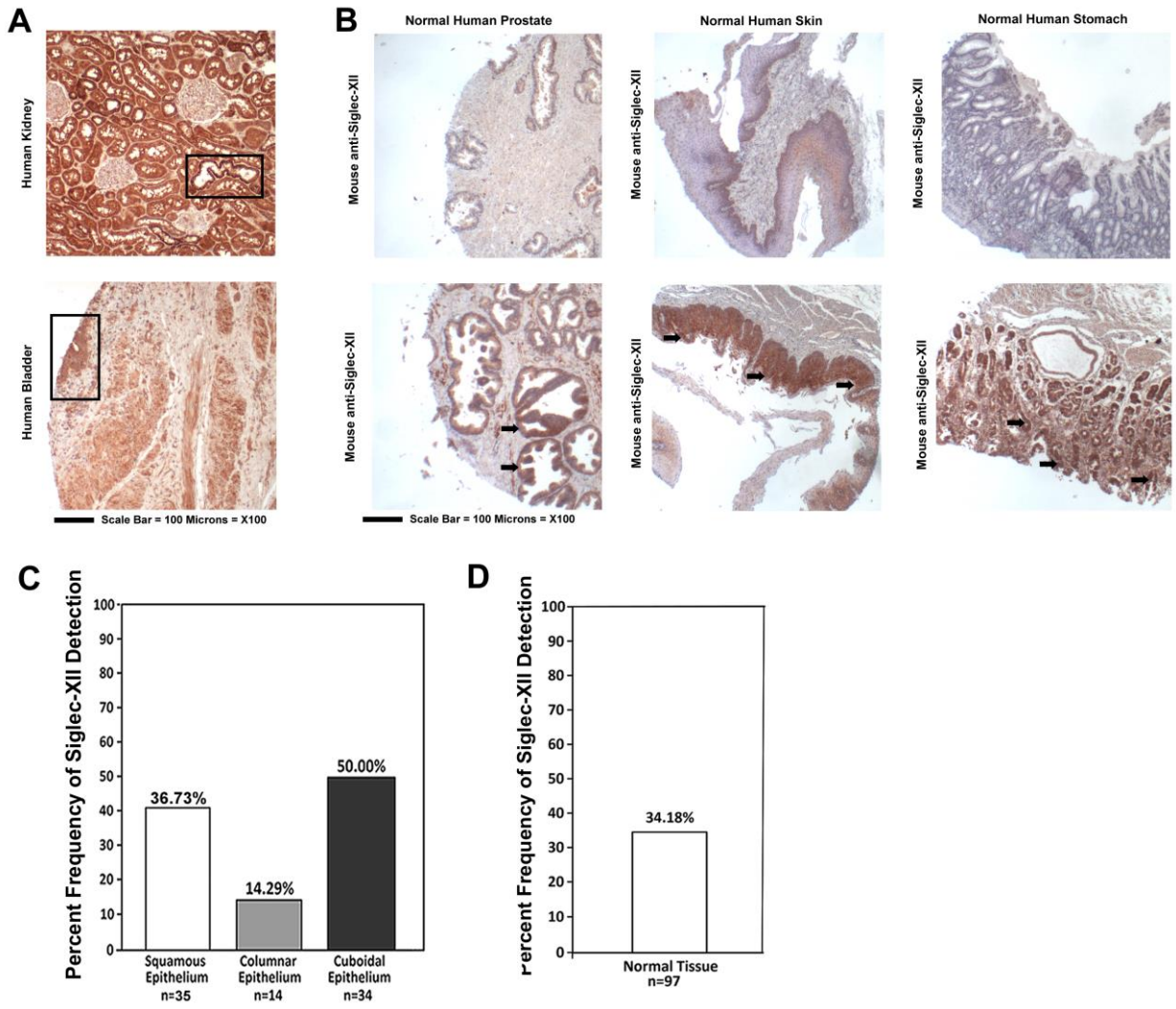


FIGURE 1. Expression of Siglec-XII in normal human epithelium. A, strong expression of Siglec-XII in human kidney tubules, but not glomeruli (top). Expression is especially strong in the distal convoluted tubules (inset). Strong expression of Siglec-XII in human bladder (bottom). Expression is especially strong in squamous epithelium lining the lumen of the bladder (inset). B, strong expression of Siglec-XII in normal human prostate, skin and stomach (bottom). No Siglec-XII expression was detected in normal human prostate, skin and stomach (top). Each sample shown was obtained from unique and separate individuals. C, Cuboidal and squamous epithelium express Siglec-XII at the highest frequency of all the epithelial subtypes. D, Human epithelial tissues express Siglec-XII protein at an overall frequency of 34.18%, which is consistent with the global *SIGLEC12* genotype distribution.

Frequency of Expression of Siglec-XII is Higher in Human Carcinomas than in Normal Human Tissues

We next analyzed the expression of Siglec-XII in human carcinomas. It was previously shown that Siglec-XII expression was seen in many human prostate carcinoma specimens and also occasionally in breast carcinoma and in melanoma (Mitra et al. 2011). We extended this observation by looking for expression of Siglec-XII in a multi-tissue array consisting of about 100 paraffin sections of human carcinoma sections using immunohistochemical methods. We found clear expression of Siglec-XII in prostate carcinomas, and occasional expression in breast carcinoma and in melanoma (Fig. 2A). Strong expression of Siglec-XII was seen in carcinomas arising from all epithelium types including prostate carcinomas arising from cuboidal epithelium, squamous carcinomas arising from squamous epithelium, and colon carcinoma arising from columnar epithelium (Fig. 2B). The frequency of detection of Siglec-XII expression in the carcinomas arising from squamous, columnar and cuboidal epithelium was significantly higher compared to the corresponding normal epithelial cell types. The frequency of detection of Siglec-XII expression in human carcinomas overall was 80.56% (Fig. 2C). This is much higher than expected, and is not consistent with the global wildtype *SIGLEC12* allele frequency in humans previously measured (Mitra et al. 2011). Our data thus extends an earlier observation using the same monoclonal antibody, that human Siglec-XII is expressed on carcinomas, but also reveals that individuals with carcinoma express Siglec-XII at more than double the frequency of individuals with no known carcinoma.

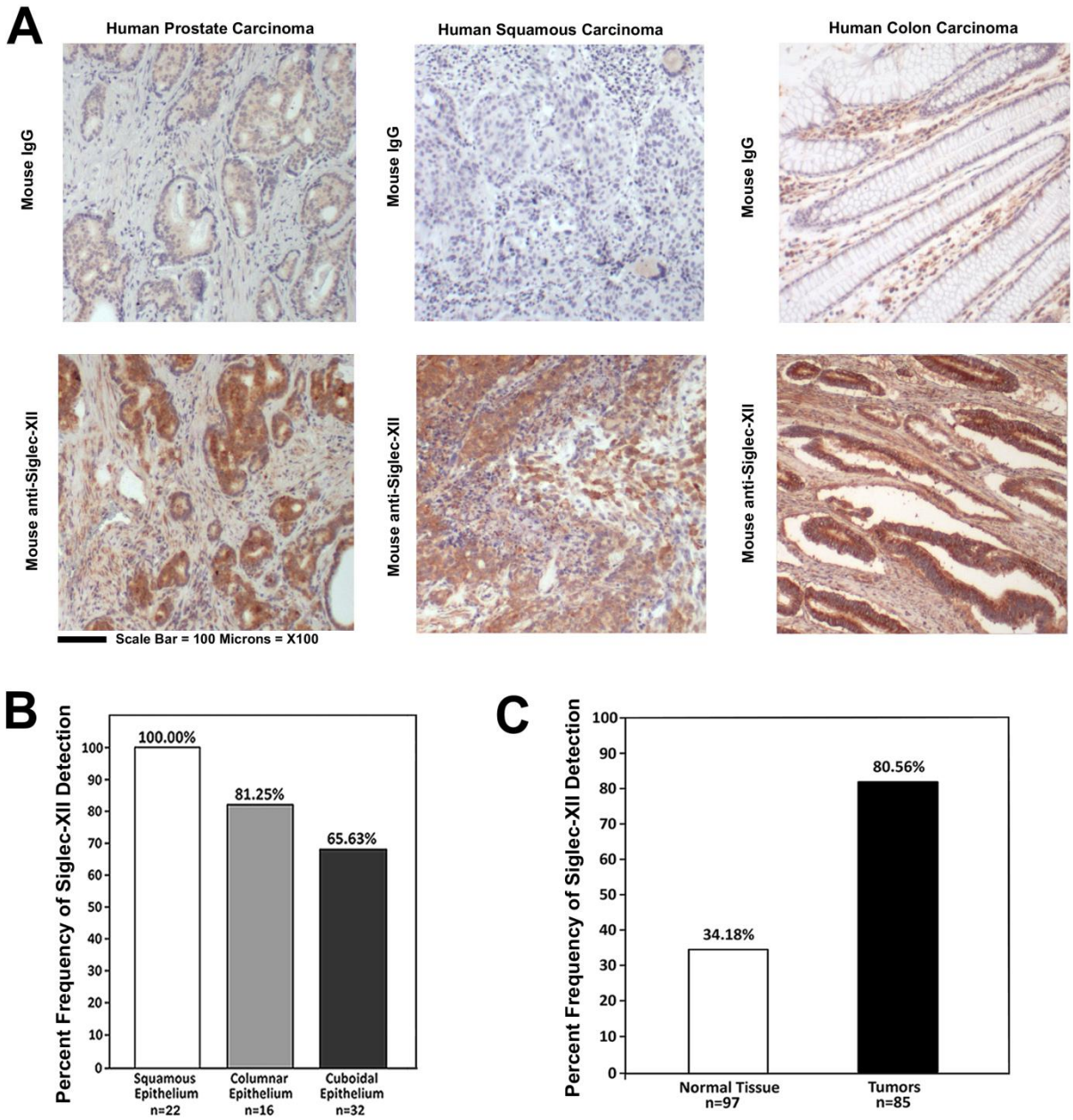


FIGURE 2. Expression of Siglec-XII in human carcinoma tissues. A, Strong epithelial expression of Siglec-XII in human prostate carcinoma, human squamous carcinoma, and human colon carcinoma. B, Human carcinomas arising from each epithelial type express Siglec-XII more frequently than the corresponding normal epithelial tissue types. C, Human carcinomas express Siglec-XII more frequently overall compared to normal tissues.

SIGLEC12 Genotype in Individuals with Low Risk for Developing Cancer

Given the unexpectedly high frequency of expression of Siglec-XII in individuals with carcinomas compared to individuals with no known carcinoma, we wondered whether the *SIGLEC12* allele status in a patient population with various cancers, not only prostate cancers, would be consistent with this finding. It was previously shown that Siglec12 allele status was not predictive of prostate carcinoma incidence. However, no genomic information was available for the individuals who had provided the tissue samples for our immunohistochemical analysis, and access to additional tissue samples for genotyping was not available either. We genotyped a low cancer risk population consisting of 53 patients with no history of carcinoma, and 54 patients with cancer detected in its early stages. We posited that this population would allow us to isolate the contribution of *SIGLEC12* to cancer incidence due to the reduction of common risk factors for cancer such as red meat consumption by individuals in this population. Statistical analysis involved performing a test of association between Siglec-XII expression status and carcinoma incidence with a Chi-Square test. In a second round of analysis, we performed a test of association between each of the *SIGLEC12* genotypes and carcinoma incidence. We found that there was no statistically significant difference in cancer incidence between individuals homozygous for an inactivating frameshift (*SIGLEC12-/-*), individuals heterozygous for (*SIGLEC12+/-*), and individuals homozygous for the wildtype alele (*SIGLEC12+/+*) (Fig. 3A). When the individuals homozygous for the inactivating frameshift (*SIGLEC12-/-*) are compared to the individuals with expression of *SIGLEC12* (*SIGLEC12-/- or SIGLEC12+/-*), there is a trend which suggests that

expression of *SIGLEC12* correlates with higher carcinoma incidence, and that inactivation of both *SIGLEC12* alleles correlates with lower carcinoma incidence, but this trend is not statistically significant either (Fig. 3B). These data did not support our hypothesis that expression of a wildtype *SIGLEC12* allele was predictive for carcinoma incidence.

A

Genotype	Cancer	Non-cancer
G+/+ (*SIGLEC12-/-*)	14	20
G+/- (*SIGLEC12+/-*)	30	27
G-/- (*SIGLEC12+/+*)	10	6
Total	**54**	**53**

B

Genotype	Cancer	Non-Cancer
SIGLEC12 -/-	14	20
SIGLEC12 +/- + *SIGLEC12 +/+*	40	33

FIGURE 3. *SIGLEC12* genotype versus cancer status in a low cancer risk population (Adventist Healthcare cohort). A, The *left column* indicates *SIGLEC12* genotype. The *middle column* shows the number of individuals with detected cancer, or a history of cancer. The *right column* shows the number of individuals with no detectable cancer or history of cancer. There was no statistically significant trend seen. B, The cancer status of the combined heterozygous (*SIGLEC12+/-*) and homozygous wildtype (*SIGLEC12+/+*) genotype individuals, which both express Siglec-XII protein, versus the homozygous inactivated (*SIGLEC12-/-*) genotype individuals. There was no statistically significant trend observed.

Prostate Cancer Progression in Five Year Follow-up with Individuals from a Previously Analyzed Prostate Carcinoma Cohort

In view of the lack of a significant result supporting a correlation between *SIGLEC12* expression and carcinoma incidence from the low cancer risk cohort, we next asked whether expression of wildtype *SIGLEC12* correlated with increased risk of cancer progression. We began the process of answering this question by analyzing the outcome of 122 patients from the prostate cohort originally analyzed by Mitra et al in 2011. Biochemical recurrence and metastasis were selected by that study as indicators of prostate cancer progression. Biochemical recurrence was defined as an interval increase in prostate specific antigen or PSA after radical prostatectomy (Stephenson et al. 2006). We found that 63% of individuals with prostate cancer progression carried a homozygous frameshift inactivation of *SIGLEC12*, while only 37% expressed *SIGLEC12* (*SIGLEC12-/- or SIGLEC12*+/-) (Fig. 4A). However, only 38 of the 122 patients demonstrated prostate cancer progression (Fig. 4A). Statistical analysis involved performing a test of association between Siglec-XII expression status and each of the progression outcomes with a Chi-Square test. In a second round of analysis, we performed a test of association between each of the *SIGLEC12* genotypes and each of the progression outcomes. However, no statistically significant correlation was observed. These data did not support our hypothesis that expression of a wildtype *SIGLEC12* allele is predictive of carcinoma progression.

A

Outcome	*SIGLEC12-/-*	*SIGLEC12 +/-*	*SIGLEC12+/+*
Biochemical Recurrence (BCR) (28)	64%	29%	7%
Metastasis (MET) (10)	60%	20%	20%
BCR+MET (Bad outcome) (38)	63%	26%	11%
No Evidence of disease (NED) (84)	48%	40%	12%

FIGURE 4. *SIGLEC12* genotype versus prostate cancer progression in patients with a minimum of 5 year follow-up. A, The *left column* indicates outcomes associated with prostate cancer progression (BCR or MET, or both), and outcomes with no evidence of prostate cancer (NED). The *columns to the right* indicate the percentage of individuals with each *SIGLEC12* genotype for each outcome. No calculated statistically significant trend was observed.

Cancer Progression in Individuals from a Late Stage Colorectal Cancer Cohort

Due to the relatively low rate of prostate cancer progression in the prostate cancer cohort which suggested that the cohort was predominated by early stage prostate carcinoma patients, we looked for a new cohort to determine whether Siglec-XII expression correlated with increased risk of late stage carcinoma progression. We analyzed the overall survival of a late stage colorectal carcinoma patient cohort (FIRE-3). Individuals in this cohort were participating in a phase 3 clinical drug trial, and were randomly and evenly divided between two treatment groups with FOLFOXIRI plus either Bevacizumab or Cetuximab. We found that the average survival period of patients who did not express wildtype *SIGLEC12* was 56 months, about twice the average survival period of 26 months for patients who expressed wildtype *SIGLEC12* (Fig. 5A). In addition, we found that 83% of patients in this cohort expressed wildtype *SIGLEC12*, and that only 17% of patients in this cohort did not express wildtype *SIGLEC12* which is consistent with the frequency of expression of Siglec-XII seen in immunohistochemistry performed earlier (80.56% and 19.44% respectively) (Fig. 5B). Statistical analysis involved performing a test of association between Siglec-XII expression status and overall survival with a Chi-Square test. These data support our hypothesis that expression of a wildtype *SIGLEC12* allele is predictive of late stage carcinoma progression, and is consistent with results obtained by immunohistochemistry.

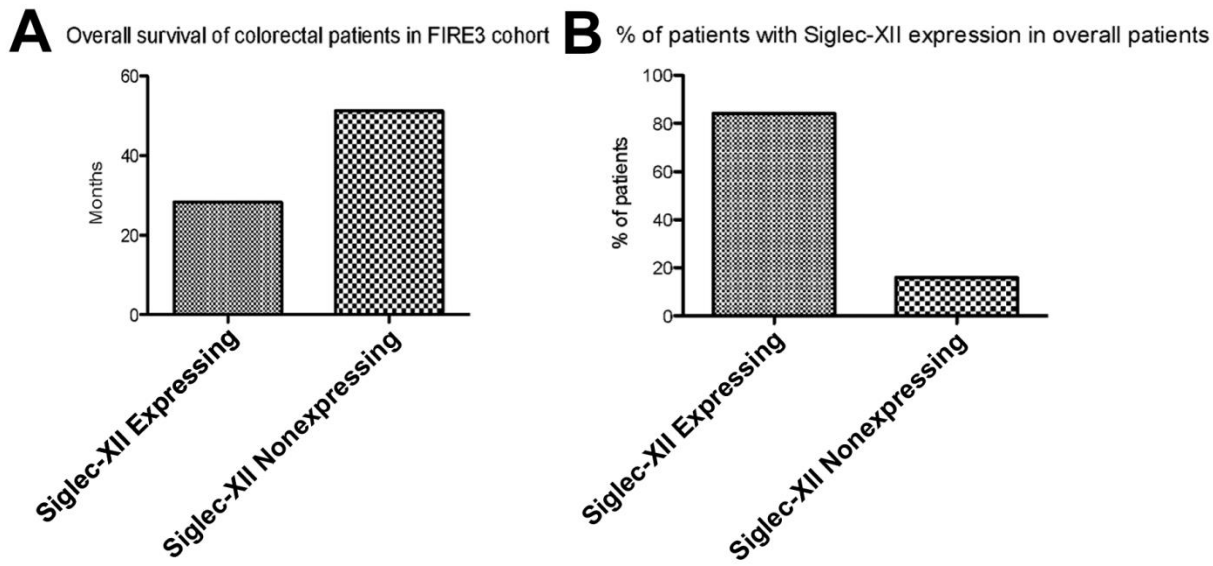
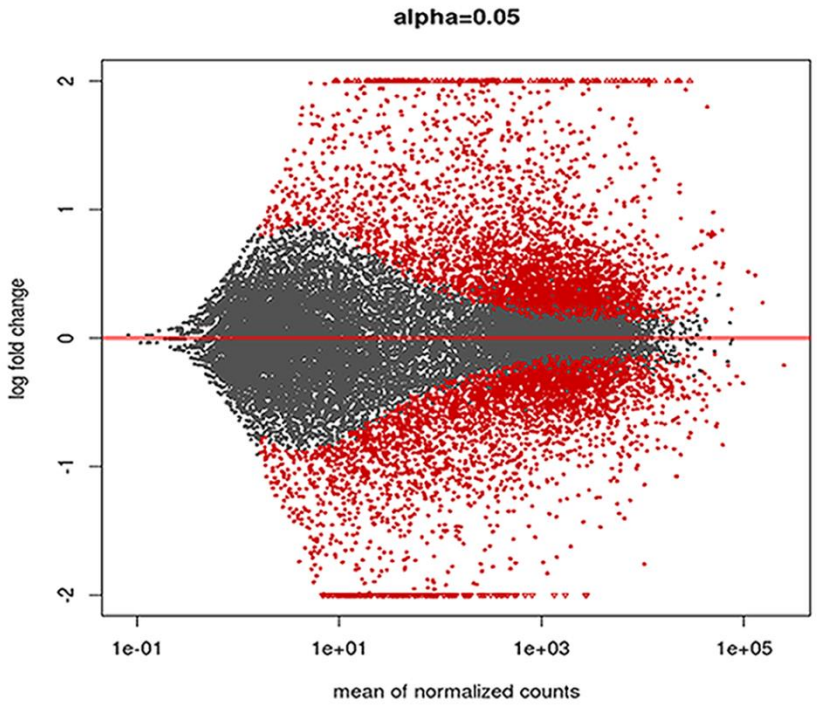


FIGURE 5. Overall survival in a late stage colorectal cancer cohort. A, Median of overall survival in months for individuals expressing Siglec-XII and for individuals who do not express Siglec-XII. Individuals who did not express Siglec-XII survived for a median of 56 months during treatment, while individuals who expressed Siglec-XII survived for a median of only 28 months during treatment. All individuals were participating in the same phase III clinical trial. B, Percentage of late stage colorectal cancer patients from the same cohort expressing Siglec-XII and not expressing Siglec-XII. The distribution of Siglec-XII expression in this cohort is consistent with the results seen using immunohistochemical analysis performed earlier (Fig. 3C).

Upregulation and Downregulation of Genes associated with Carcinoma Progression in PC-3 Prostate Carcinoma Cell Line Stably Transfected with *SIGLEC12*.

In order to better understand the mechanism of action of *SIGLEC12*, and how it contributes to late carcinoma progression, RNA-Seq analysis was performed on a PC-3 prostate carcinoma cell line stably transfected with *SIGLEC12*. It was previously shown that tumor consisting of similarly transfected PC-3 cell line demonstrated significantly increased growth compared to an untransfected PC-3 cell line (Mitra et al. 2011). Conveniently, the PC-3 cell line does not express Siglec-XII protein endogenously, as it is homozygous for frameshift inactivation of both *SIGLEC12* alleles (Mitra et al. 2011). We found many differentially expressed genes between the untransfected and stably transfected PC-3 cells (Fig. 6A). Several of genes upregulated in the stably transfected Siglec-XII expressing PC-3 cells, such as IDO1, LCP1, MX2, BST2 and CEACAM6 have been associated with cancer progression in when upregulated previous literature (Table. 6B). (Prendergast et al. 2014, Munn et al. 2016, Koide et al. 2017, Kobayashi et al 2004, Cai et al. 2009, Lewis-wambi et al. 2008). Several of the downregulated genes, such as CXADR, TACSTD2, and CTSF have been associated with cancer progression when downregulated as well (Martin et al. 2005, Wang et al. 2014, Ji et al. 2017). These data suggest that the mechanism by which *SIGLEC12* may play a role in late carcinoma progression involves the regulation of downstream targets including genes known to be associated with carcinoma progression.

A



B

Upregulated Genes

Gene name	P-Value	Fold Change	Role
IDO1 (Indoleamine 2,3-Dioxygenase 1)	1.50E-53	23	Catalyzes the first and rate-limiting step in tryptophan catabolism
LCP1 (Lymphocyte Cytosolic Protein 1)	1.14E-250	21	Actin-binding protein
MX2 (MX Dynamin Like GTPase 2)	2.35E-39	18	GTP binding and GTPase activity
BST2 (Bone Marrow Stromal Cell Antigen 2)	2.25E-34	17	Pre-B-cell growth and in rheumatoid arthritis
CEACAM6 (Carcinoembryonic Antigen Related Cell Adhesion Molecule 6)	9.97E-69	13	Cell adhesion

FIGURE 6. Differential gene expression in *SIGLEC12* stably transfected PC-3 prostate cancer cell line. A, Distribution of genes by log fold change in expression. Grey points indicate transcripts with minimal change in expression between the untransfected and *SIGLEC12* transfected PC-3 cells. Red points indicate transcripts in the *SIGLEC12* transfected PC-3 cells with significant change in expression compared to the untransfected PC-3 cells. Table showing various upregulated genes which have been associated with cancer progression when upregulated. Also included are the P-Value, fold increase and role of these genes.

Diagnostic Urine Dot Blot to Determine Whether Individuals Express Siglec-XII

It was previously shown that multiple inactivating mutations affect *SIGLEC12* in humans (Mitra et al. 2011). As a result, determining the *SIGLEC12* allele status of individuals would require sequencing with multiple primers specific to each inactivating mutation. A protocol that could quickly determine in a minimally invasive manner whether an individual expressed Siglec-XII protein would be advantageous, as Siglec-XII protein expression correlates with an intact *SIGLEC12* allele (Mitra et al. 2011). Given the relatively strong expression of Siglec-XII seen in distal renal tubules and bladder epithelium, we developed a dot blot protocol which would allow the *SIGLEC12* allele status of individuals to be determined from a provided urine sample. We used a second clone of monoclonal antibody (clone 1130) which demonstrated increased binding to a prostate cancer cell line PC3 in previous FACS experiments (Mitra et al. 2011). We then compared the dot blot results for each individual to the genotype obtained by sequencing of buccal cells obtained by buccal swab. We found that Siglec-XII protein was detected in the urine of all 3 individuals who were heterozygous for wildtype *SIGLEC12*+/+ (Fig. 7A). We also found that Siglec-XII protein was not detected in the urine of all 2 individuals who were homozygous for inactivated *SIGLEC12-/-* (Fig. 7A). No Siglec-XII protein was detected in sterile PBS which served as a negative control, while Siglec-XII fc receptors which served as a positive control. These data suggest that the urine dot blot may be a useful tool to determine whether individuals express Siglec-XII protein and consequently, the *SIGLEC12* allele status of these individuals as well.

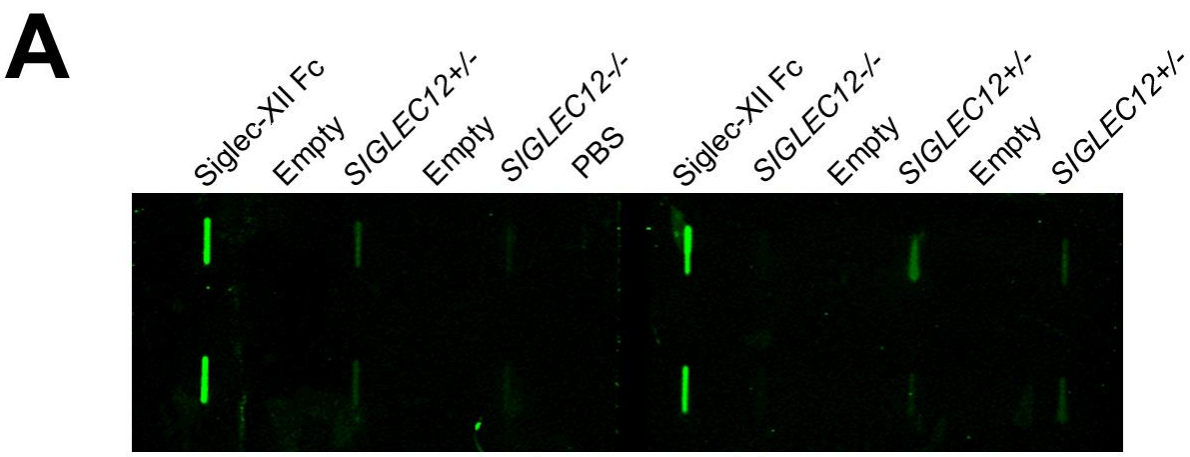


FIGURE 7. Urine dot blot to detect Siglec-XII in healthy donors. A, Fluorescence visualized detection of Siglec-XII protein. The contents of each column are indicated at the top of the figure. The genotype of the individual who provided urine samples is indicated above their respective columns. The first and second rows are replicates performed concomitantly. Siglec-XII Fc was used as a positive control, while sterile PBS was used as a negative control.

DISCUSSION

Siglecs are an essential part of the cell to cell communication system both within and between organisms such as host and pathogens. This is due to their transmembrane positioning which allows them to bind sialic acids which are found on the terminal ends of cell surface glycans. Sialic acid recognition is a defining characteristic of Siglecs (Chen et al. 2010). Due to this constraint, Siglec-XII is not considered by many to be a true Siglec as it does not recognize sialic acids. This is due to a R122C missense mutation universal to humans which affects the sialic acid binding Vset domain. Human Siglec-XII protein with the R122C missense mutation is expressed as a non-sialic acid binding, full-length protein. While reverse mutated C122R human Siglec-XII protein has been shown to bind sialic acids (Angata et. al 2001), its natural ligands remain unknown. However, previous evidence has shown that Siglec-XII may recognize sialic acids weakly. Cos-7 cells expressing Siglec-XII were shown to bind sialylated red blood cells by 4 to 5 fold greater than non-expressing Cos-7 cells (Yu et al. 2001). This may arise from secondary interactions between sialic acids and Siglecs, such as the interaction between a sialic acid glycerol group and a conserved hydrophobic amino acid which is retained in Siglec-XII (Zhuravleva et al. 2008, Attrill et al. 2006, May et al. 1998). Other Siglecs have been shown to retain some Sia binding even without the essential Arg residue in its V2 set domain (Angata et al. 2002, Brinkman et al. 2007, Vinson et al. 2001). Both the human and primate *SIGLEC12* gene product possess two ligand binding V2 domains, a characteristic which is unique amongst Siglecs, which possess only one V2 domain. It is also possible that an additional interaction is occurring between a

second V2 set domain and an unknown ligand. This evidence suggests that Siglec-XII may retain some residual sia binding ability despite the universal R122C missense mutation. Previous microarray experiments have suggested that Siglec-XII mediates large scale gene down-regulation upon stable expression in a prostate cancer cell line (Mitra et al. 2011), indicating that Siglec-XII mediates downstream signaling in the absence of the essential arginine.

It was previously shown that individuals homozygous for frameshift inactivated *SIGLEC12* lack any detectable Siglec-XII expression (Mitra et al. 2011). About 40% of humans do not express Siglec-XII due to a frameshift mutation (Mitra et al. 2011). This, along with the universal inactivating mutation affecting sialic acid binding, suggests that *SIGLEC12* is being subjected to strong selective pressure. In a large-scale survey of 169 genes in the human genome containing nonsense SNPs which were variable, others have shown that *SIGLEC12* has the 36th highest frequency of homozygous inactivation (Yngvadottir et al. 2009). This high frequency of homozygous inactivation indicates that human *SIGLEC12* is undergoing strong selection.

The universal missense mutation affecting ligand binding and the high frequency of homozygous inactivation may be evidence that such a selective sweep is currently happening to *SIGLEC12* in humans. Besides *SIGLEC12*, other Siglecs including *SIGLEC1, SIGLEC5/14, SIGLEC6, SIGLEC7, SIGLEC9, SIGLEC11, SIGLEC13 and SIGLEC16* have undergone human specific changes in expression, ligand binding and functional gene status (Varki et al. 2010). These human specific

differences indicate CD33-related *SIGLECs* are being subjected to strong selective pressure in humans (Varki et al. 2010, McMillan et al. 2008).

Unlike most other Siglecs, Siglec-XII is expressed on human and chimpanzee epithelia. Epithelial carcinomas are rare in chimpanzees in captivity, but occur at high frequencies in humans (Varki et al. 2011, Varki et al. 2010, Mcclure et al. 1973, Seibold et al. 1973). It is possible that Siglec-XII expression in humans is linked to a greater risk for developing carcinomas in humans, which would directly contribute to the strong negative selective pressure against human *SIGLEC12*. If this were the case, it would be important to determine whether expression of Siglec-XII indeed correlates with greater risk for cancer.

In order to determine whether Siglec-XII is risk factor for cancer, we looked for differential expression of Siglec-XII in both normal human tissue and human carcinomas (cancers of epithelial origin). The immunohistochemical analysis confirmed previous results, which showed that Siglec-XII was expressed in some normal epithelial tissues as well as some carcinomas (Mitra et al. 2011). It also showed that cuboidal epithelium in the distal convoluted tubule of the kidney, and transitional epithelium lining the lumen of the bladder expressed Siglec-XII relatively strongly. This suggested that Siglec-XII expressed on these epithelial cells might be detectable in urine, which we addressed later in our urine dot blot. The overall frequency of Siglec-XII expression in this cohort was 34.18%. This appears mildly consistent with the overall genotype distribution of *SIGLEC12*, with a 58% global allele frequency of an inactivating frameshift (Mitra et al. 2011). This in turn suggests that about 42% of humans should express a functional *SIGLEC12* gene product.

However, it should be noted that there is another inactivating mutation causing a premature stop with a global frequency of 18.6% (Mitra et al. 2011). The additive effect of these two mutations would indicate that the actual expected frequency of Siglec-XII expression would be slightly lower than 42%. It was previously shown that Siglec-XII expression detected by immunohistochemical analyses correlates with the *SIGLEC12* genotype of the individual (Mitra et al. 2011). As a result, the overall frequency of expression of Siglec-XII of 34.18% observed in the normal tissues is consistent with the overall *SIGLEC12* genotype distribution, which indicates that the cohort that provided the normal tissue is representative of the *SIGLEC12* genotype distribution of the general population.

Having established the frequency of expression of Siglec-XII in normal epithelium, we looked at the frequency of Siglec-XII expression in human carcinomas. If Siglec-XII expression was linked to greater cancer risk, we expected to observe a significantly higher frequency of Siglec-XII expression in the carcinomas compared to the normal tissue. This would indicate that individuals who develop cancer express Siglec-XII more frequently than those who do not. This in turn would suggest that individuals who express Siglec-XII are more prone to developing cancer. Immunohistochemical analysis showed that Siglec-XII is expressed in 80.56% of the carcinomas, which is much higher than expected. This suggests that individuals who develop carcinomas express Siglec-XII more than twice as frequently as individuals who do not develop carcinomas.

However, there are several limitations to these findings. While individuals who provided both the normal and the carcinoma tissues were evenly distributed from the

ages of 21 to 73, no information about the genotype of these individuals was available as the tissue samples were anonymized, and additional tissue was not available for in-house genotyping. Due to anonymization, there was also no information about the individual's risk factors for cancer. In addition, the evenly distributed cohorts meant that there were many younger individuals under the age of 40. The most common cancers in adults in the United States are prostate cancer and breast cancer, both of which have an onset in the early 40's (Ferlay et al. 2008). Younger individuals who express Siglec-XII who are currently normal have a significant chance of developing these common cancers in the future. As a result, any correlation between Siglec-XII expression and risk of developing cancer was likely underestimated by this data.

In order to address some of these limitations, we looked for another cohort which could provide tissue samples for in-house genotyping, and which could provide information about some of the individual's risk factors for cancer. The Adventist Health System cohort consisted of about 100 individuals, who do not smoke tobacco, consume alcohol, and consume only small amounts of red meat, which are all factors associated with cancer risk Aykan et al. 2015, Bagnardi et al. 2015, Boyle et al. 2003, Doll et al. 1950). Due to this relatively low cancer risk in this population, we posited that any contributory effect of Siglec-XII to cancer risk would be more isolated and pronounced. In addition, individuals from this cohort had provided peripheral blood samples which were previously frozen for further testing. We performed genotyping on these peripheral blood samples with the same primer that had been used to genotype prostate cancer samples in a previous paper (Mitra

et al. 2011). After genotyping was complete, we compared the genotype of each individual to their cancer status. While, no statistically significant trend was seen for this cohort when comparing the homozygous inactivated (*SIGLEC12-/-)*, heterozygous (*SIGLEC12*+/-), and homozygous wildtype (*SIGLEC12*+/+) individuals, there seemed to be a slight trend when comparing the Siglec-XII expressing and the Siglec-XII non-expressing individuals. However, this trend was also not statistically significant.

Having selected this cohort in part to isolate any contributing effect of Siglec-XII expression on cancer risk, the lack of a statistically significant trend was unexpected. One possible explanation was that the *SIGLEC12* genotype distribution in this population was not representative of the general population. Indeed, 33 of the 53 or about 62% of the individuals in this cohort who had no history of cancer expressed Siglec-XII, which is significantly higher than the percentage seen in immunohistochemistry (34.18%). This frequency is also significantly higher than the estimated frequency obtained from previous genotyping which suggests that less than 42% of individuals express Siglec-XII in the general population (Mitra et al. 2011). In spite of this, even when considering the effect this increased frequency of Siglec-XII expression might have on the outcome of this experiment, it seemed more likely that Siglec-XII did not play as great of a role in overall cancer risk as previously thought. However, the results from the immunohistochemical analysis still indicated that expressing Siglec-XII is correlated with a more than two-fold increase for developing cancer. One explanation which resolved the seemingly contradictory results obtained was based on the progression of the cancers in each of the cohorts

analyzed thus far. While the Adventist cohort patients were evenly distributed across all ages and evenly split between cancer and non-cancer groups, these patients were under active surveillance for cancers, which was the reason peripheral blood samples had been provided by these patients. As a result, many of the cancers in these patients were likely cancers which had been detected through surveillance at an early stage.

On the other hand, the carcinoma tissues analyzed in the by immunohistochemistry were obtained from biopsies obtained during surgical resection. Surgical resection is normally performed on late stage cancers. This suggests that the individuals who expressed Siglec-XII detected by immunohistochemistry were late stage cancer patients. Given the lack of a significant correlation between Siglec-XII expression and cancer risk in the Adventist cohort, which was likely over representing individuals with early stage cancers, and the statistically significant correlation between Siglec-XII expression and developing late stage cancers seen through immunohistochemical analysis, we posited that Siglec-XII expression plays an important role in cancer progression rather than cancer risk.

In order to determine whether Siglec-XII expression plays an important role in cancer progression, we decided to revisit patients from the prostate cancer cohort originally genotyped by Mitra et al. in 2011 and kindly provided by Dr. Robin Leach at the University of Texas Austin. We analyzed the prostate cancer outcome for 122 of the original 242 prostate cancer cases (not including control cases) which had a minimum of 5-year follow-up at the University of Texas Austin Medical Center.

Prostate cancer patients who demonstrated biochemical recurrence, metastasis or both were considered to have progressed while those with no evidence of disease were considered to have not progressed. We found no statistically significant correlation between Siglec-XII expression and prostate cancer progression. In fact, only 37% of individuals with prostate cancer progression expressed Siglec-XII, which is more consistent with the expected frequency of Siglec-XII expression in the general population obtained from immunohistochemical analysis (34.18%), and the global *SIGLEC12* allele distribution (less than 42%). This data seemed to suggest that Siglec-XII expression is not correlated with carcinoma progression. However, we found that the initial detection method and the chosen indicators for prostate cancer progression may have significantly downplayed any significant result.

Each of the original 242 prostate cancer cases had been detected by abnormal prostate specific antigen (PSA) testing which was recommended by some physicians and professional organizations annually for men over the age of 50 until about 2008. Since this time, as more disadvantages of PSA tests (including a high rate of false positives) was revealed, PSA testing is no longer considered a reliable means of diagnosing prostate cancer, and is used as a confirmatory test instead (Wever et al. 2010, Avery et al. 2008). In following with this, though all the 122 cases with follow-up were diagnosed with prostate cancer before 2011, 84 of these cases demonstrated no evidence of disease after a minimum of 5 year follow-up. It is therefore possible that a significant portion of these individuals had tested falsely positive on the PSA test and had never had prostate cancer. It is therefore possible that only the remaining 38 patients had developed prostate cancer. This analysis

can be extended to the 28 individuals who demonstrated biochemical recurrence. Biochemical recurrence is defined as an interval increase in PSA levels after surgical or radiation treatment (Stephenson et al. 2006). However, recent work has shown that the value for PSA levels must be greater than 0.4 ng/ml for biochemical recurrence to be considered durable after radical prostatectomy (Toussi et al. 2016). It is possible that the prostate cancer in a significant portion of the 28 individuals who demonstrated biochemical recurrence did not have progressive disease. In any case, due to the ambiguity of the outcome of some patients who were considered to have progressed, and some of the patients who had no evidence of disease, and the small sample size of individuals who were considered to have progressed (n=38), no definite conclusion about the correlation of Siglec-XII expression and prostate carcinoma progression could be drawn from this data alone.

Observation by clinicians has shown that the majority of prostate cancers, which appear as slow growing, low-grade prostate cancer at diagnosis, are unlikely to progress quickly enough to cause health problems in an adult male's lifetime, and that prostate cancers which are diagnosed with a rating of high-grade will behave very aggressively (Crawford et al. 2013). As a result, it is likely that the majority of the prostate cancer cases would never progress to later, invasive stages within the patient's lifetime. It is also likely that for the 120 individuals for which a minimum of 5 year follow-up was not available, that some of these individuals have been lost to follow-up as a result of cancer progression and death before 5 a year follow-up analysis could be completed. This suggests that this prostate cancer cohort yielded

an incomplete understanding of the relationship between Siglec-XII expression and prostate carcinoma progression.

In order to establish whether such a relationship exists, we examined a third cohort consisting of 500 late stage colorectal cancer patients who were participating in a phase III clinical trial named FIRE3. The investigators conducting the clinical trial considered a longer survival period after initiation of drug treatment, along with tumor response to be metrics for improved outcome. Results showed that the median overall survival was 33.1 months with FOLFIRI plus cetuximab vs 25.0 months with FOLFIRI plus bevacizumab (hazard ratio [HR] = 0.70, *P* = .0059) (Loupakis et al. 2014), which was considered a very significant result. We also examined the median overall survival of the entire cohort in relation to Siglec-XII expression. While we considered that each drug treatment might skew the overall survival in our analysis, Siglec-XII expressing and non-expressing individuals were distributed evenly between the two drug groups, which consisted of about 250 individuals each.

We found that individuals in the clinical trial who expressed Siglec-XII survived for 28 months while individuals who were Siglec-XII non-expressing survived for 56 months. This suggests that late stage colorectal cancer patients who express Siglec-XII will live only half as long as patients who do not express Siglec-XII. In addition, about 82% of individuals in this cohort expressed Siglec-XII, which was consistent with the frequency of Siglec-XII expression in carcinomas seen with immunohistochemistry.

Given that all the patients from this cohort were late stage colorectal cancer patients, the colorectal cancer in each patient was already at an advanced stage and behaving aggressively at the time of the study. The difference in overall survival period seen here suggests that Siglec-XII may play an important role at this late stage of cancer development. It is possible that during the early stages of carcinoma, the effect of Siglec-XII is minimal, but as the carcinoma progresses, its role becomes increasingly salient. Previous microarray evidence has shown in a stably transfected PC-3 prostate cancer cell line that a large number of genes are downregulated, including some genes associated with cancer progression. However, the mechanism through which Siglec-XII regulates downstream targets is not known.

In order to expand on previous microarray analysis, we conducted RNA-Seq analysis of the same *SIGLEC12* stably transfected PC-3 prostate cancer cell line which had demonstrated a faster rate of tumor growth when implanted in nude mice, and which had also been used for microarray analysis (Mitra et al. 2011). We found many genes which were differentially expressed in the Siglec-XII expressing versus the Siglec-XII non-expressing PC-3 cells. Some of the genes which were upregulated as well as downregulated have been associated with cancer progression in previous literature. The gene product of one of the upregulated genes *IDO1*, indoleamine-pyrrole 2,3-dioxygenase (IDO) is one of two enzymes that catalyze the initial and rate-limiting step in the kynurenine pathway, which involves the O2-dependent oxidation of L-tryptophan to N-formylkynurenine (Dai et al. 1990). IDO has been shown to be involved in immune modulation by limiting T cell function and is involved in mechanisms of immune tolerance (Munn et al. 2016). Recent work

has also shown that IDO is activated during tumor development, and may help malignant cells survive by suppressing the growth and survival of T cells (Prendergast et al. 2014, Munn et al. 2016, Muller et al. 2005). IDO expression is normally controlled by bridging integrator-1 (Bin1), a tumor suppressor which inhibits Myc through a Myc-binding domain (Lundgaard et al. 2011). Bin1 has been shown to reverse PDL-1 mediated immune escape through inhibition of Myc and EGFR signaling pathways (Wang et al 2017). Myc is a regulator gene that encodes a protein that regulates cell cycle progression. Mutations which cause Myc to be constitutively expressed can be found in many cancers, and are associated with increased cell proliferation and tumor progression (Dang et al. 2012). In addition, previous work has shown that Siglec-XII (referred to as S2V) recruits Src homology 2 domain-containing protein tyrosine phosphatases SHP-1 and SHP-2 upon tyrosine phosphorylation (Yu et al. 2001). Specifically, it was shown that SHP-1 binds to the membrane proximal ITIM of Siglec-XII in HEK293A cells, and that SHP-2 demonstrated some interaction with Siglec-XII as well. Reduction of SHP-2 activity suppresses tumor growth in a mouse model of lung adenocarcinoma, and it is considered a potential target of cancer therapy (Scheeberger et al. 2015). SHP-1 negatively regulates cell cycle progression and the inflammatory response, which are important aspects of oncogenic transformation (Sharma et al. 2016). In hematopoietic cells, SHP-1 associates with the immunoreceptor tyrosine-based inhibitory motif (ITIM) of inhibitory receptors to perform its inhibitory regulation (Tamir I et al. 2000, Bakker et al. 2000, Bléry et al. 2000). As Siglec-XII retains its cytoplasmic ITIM signaling domain, it is likely that such an interaction is responsible

for the altered downstream regulation in carcinomas. However, as the natural ligand of Siglec-XII is unknown, it is unclear whether altered downstream regulation is due to gain of function or loss of function involving the inhibitory interaction between Siglec-XII and SHP-1 and SHP-2. In view of the universal missense mutation in human Siglec-XII which affects the sialic acid binding domain in the V2 set domain, it is more likely that the inhibitory signaling between the cytoplasmic ITIM domain of Siglec-XII and downstream targets such as SHP-1 and SHP-2 is reduced.

One of the primary advantages of a protein detection assay over sequencing is that multiple inactivating mutations affect the *SIGLEC12* locus (Mitra et al. 2011). As a result, accurate genotyping might require sequencing with multiple primers. Given the potential role that Siglec-XII may play in late carcinoma progression, we designed a dot blot protocol which would allow clinicians and researchers to quickly determine whether an individual is Siglec-XII expressing or non-expressing using a provided urine sample. We designed the dot blot protocol based on the immunohistochemical observation that Siglec-XII is strongly expressed in distal renal tubule epithelium and bladder epithelium. These epithelial cells are occasionally sloughed off by humans and appear in urine. However, as urine is typically acellular in healthy humans, we had to rule out the possibility that nonspecific detection would create false positives in the dot blot. We performed immunohistochemical analysis on pellets obtained from urine samples of a *SIGLEC12-/-* and a *SIGLEC12+/-* individuals. The immunohistochemical analysis was performed with the same clone of monoclonal anti-Siglec-XII antibody (clone 1130) which would also be used for detection in the dot blot, and was visualized with a fluorescent secondary. Strong

detection was seen on several renal casts and squamous epithelial cells likely originating from the bladder in the Siglec-XII expressing individual while no detection was seen in the Siglec-XII non-expressing individual. Having established that the urine dot blot would not result in false positives due to nonspecific detection, we performed the urine dot blot on samples taken from 5 different individuals, including the individuals who had provided the urine samples for the immunohistochemical analysis. The urine samples were refrigerated at 4 degrees Celsius for 72 hours between collection and analysis to simulate typical storage and transport times for urine samples used in diagnostic testing currently. For all 5 individuals, the Siglec-XII expression status detected using the dot blot correlated with their *SIGLEC12* genotype, as determined by buccal cell sequencing. Given the accuracy and relative simplicity of this procedure, this noninvasive test may become a very useful diagnostic tool for clinicians and researchers to determine the Siglec-XII expression status of patients in the future.

Future Directions

We are planning to conduct a study on an existing patient cohort consisting of individuals with early stage carcinomas. We plan to establish a timeline of carcinoma progression during several years of follow-up. We will also compare whether Siglec-XII expression had a greater effect in cases with aggressive cancers on a case by case basis.

We are also currently conducting analysis of full genome sequences from a cohort of 343 cancer patients. We intend to compare the *SIGLEC12* allele status of individuals with advanced carcinomas and individuals with stable, non-aggressive

carcinomas from the same patient cohort in order to further correlate Siglec-XII expression with increased risk of late carcinoma progression. It will also be interesting to compare the effect on progression between different carcinoma types, which each demonstrate unique profiles of progression and development.

REFERENCES

Angata T, Varki A. Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. Chem Rev. 2002;102(2):439-69.

Schauer R. Sialic acids as regulators of molecular and cellular interactions. Curr Opin Struct Biol. 2009;19(5):507-14.

Chen X, Varki A. Advances in the biology and chemistry of sialic acids. ACS Chem Biol. 2010;5(2):163-76.

Varki A, Angata T. Siglecs--the major subfamily of I-type lectins. Glycobiology. 2006;16(1):1R-27R.

Crocker PR, Paulson JC, Varki A. Siglecs and their roles in the immune system. Nat Rev Immunol. 2007;7(4):255-66.

Von gunten S, Bochner BS. Basic and clinical immunology of Siglecs. Ann N Y Acad Sci. 2008;1143:61-82.

Lopez PH, Schnaar RL. Gangliosides in cell recognition and membrane protein regulation. Curr Opin Struct Biol. 2009;19(5):549-57.

Varki, Ajit. *Essentials of glycobiology*. Cold Spring Harbor Laboratory Press, 2017.

Cao H, Crocker PR. Evolution of CD33-related siglecs: regulating host immune functions and escaping pathogen exploitation. Immunology. 2011;132(1):18-26.

Cornish AL, Freeman S, Forbes G, Ni J, Zhang M, Cepeda M, Gentz R, Augustus M, Carter KC, Crocker PR. Characterization of siglec-5, a novel glycoprotein expressed on myeloid cells related to CD33. Blood. 1998;92(6):2123-32.

Angata T, Margulies EH, Green ED, Varki A. Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. Proc Natl Acad Sci USA. 2004;101(36):13251-6.

Angata T, Varki NM, Varki A. A second uniquely human mutation affecting sialic acid biology. J Biol Chem. 2001;276(43):40282-7.

Mitra N, Banda K, TK Altheide, Schaffer L, Johnson-Pais TL, Beuten J, Leach RJ, Angata T, Varki N, Varki A. SIGLEC12, a human-specific segregating (pseudo)gene, encodes a signaling molecule expressed in prostate carcinomas. J Biol Chem. 2011;286(26):23003-11.

Aykan NF. Red Meat and Colorectal Cancer. Oncol Rev. 2015;9(1):288.

Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, Scotti L, Jenab M, Turati F, Pasquali E, Pelucchi C, Galeone C, Bellocco R, Negri E, Corrao G, Boffetta P, La Vecchia C. Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. Br J Cancer. 2015;112(3):580-93.

Boyle P, Autier P, Bartelink H, Baselga J, Boffetta P, Burn J, Burns HJG, Christensen L, Denis L, Dicato M, Diehl V, Doll R, Franceschi S, Gillis CR, Gray N, Griciute L, Hackshaw A, Kasler M, Kogevinas M, Kvinnsland S, La Vecchia C, Levi F, McVie JG, Maisonneuve P, Martin-Moreno JM, Newton Bishop J, Oleari F, Perrin P, Quinn M, Richards M, Ringborg U, Scully C, Siracka E, Storm H, Tubiana M, Tursz T, Veronesi U, Wald N, Weber W, Zaridze DG, Zatonski W, zur Hausen H. European Code Against Cancer and scientific justification: third version (2003). Ann Oncol. 2003;14(7):973-1005.

Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. Br Med J. 1950;2(4682):739-48.

Yu Z, Lai CM, Maoui M, Banville D, Shen SH. Identification and characterization of S2V, a novel putative siglec that contains two V set Ig-like domains and recruits protein-tyrosine phosphatases SHPs. J Biol Chem. 2001;276(26):23816-24.

Zhuravleva MA, Trandem K, Sun PD. Structural implications of Siglec-5-mediated sialoglycan recognition. J Mol Biol. 2008;375(2):437-47.

Attrill H, Takazawa H, Witt S, Kelm S, Isecke R, Brossmer R, Ando T, Ishida H, Kiso M, Crocker PR, van Aalten DM. The structure of siglec-7 in complex with sialosides: leads for rational structure-based inhibitor design. Biochem J. 2006;397(2):271-8.

May AP, Robinson RC, Vinson M, Crocker PR, Jones EY. Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 A resolution. Mol Cell. 1998;1(5):719-28.

Angata T, Kerr SC, Greaves DR, Varki NM, Crocker PR, Varki A. Cloning and characterization of human Siglec-11. A recently evolved signaling molecule that can interact with SHP-1 and SHP-2 and is expressed by tissue macrophages, including brain microglia. J Biol Chem. 2002;277(27):24466-74.

Brinkman-Van der Linden EC1, Hurtado-Ziola N, Hayakawa T, Wiggleton L, Benirschke K, Varki A, Varki N. Human-specific expression of Siglec-6 in the placenta. Glycobiology. 2007;17(9):922-31.

Vinson M, Strijbos PJ, Rowles A, Facci L, Moore S, Simmons D, Walsh FS. Myelin-associated glycoprotein interacts with ganglioside GT1b. A mechanism for neurite outgrowth inhibition. J Biol Chem. 2001;276(23):20280-5.

Yngvadottir B, Xue Y, Searle S, Hunt S, Delgado M, Morrison J, Whittaker P, Deloukas P, Tyler-Smith C. A genome-wide survey of the prevalence and evolutionary forces acting on human nonsense SNPs. Am J Hum Genet. 2009;84(2):224-34.

Varki A. Colloquium paper: uniquely human evolution of sialic acid genetics and biology. Proc Natl Acad Sci USA. 2010;107 Suppl 2:8939-46.

Mcmillan SJ, Crocker PR. CD33-related sialic-acid-binding immunoglobulin-like lectins in health and disease. Carbohydr Res. 2008;343(12):2050-6.

Varki NM, Strobert E, Dick EJ, Benirschke K, Varki A. Biomedical differences between human and nonhuman hominids: potential roles for uniquely human aspects of sialic acid biology. Annu Rev Pathol. 2011;6:365-93.

Mcclure HM. Tumors in nonhuman primates: observations during a six-year period in the Yerkes primate center colony. Am J Phys Anthropol. 1973;38(2):425-9.

Seibold HR, Wolf RH. Neoplasms and proliferative lesions in 1065 nonhuman primate necropsies. Lab Anim Sci. 1973;23(4):533-9.

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010;127(12):2893-917.

Koide N, Kasamatsu A, Endo-Sakamoto Y, Ishida Sho, Shimizu T, Kimura Y, Miyamoto I, Yoshimura S, Shiiba M, Tanzawa H, Uzawa K. Evidence for Critical Role of Lymphocyte Cytosolic Protein 1 in Oral Cancer. Sci Rep. 2017;7:43379.

Kobayashi K, Nishioka M, Kohno T, Nakamoto M, Maeshima A, Aoyagi K, Sasaki H, Takenoshita S, Sugimura H, Yokota J. Identification of genes whose expression is upregulated in lung adenocarcinoma cells in comparison with type II alveolar cells and bronchiolar epithelial cells in vivo. Oncogene. 2004;23(17):3089-96.

Cai D, Cao J, Li Z, Zheng X, Yao Y, Li W, Yuan Z. Up-regulation of bone marrow stromal protein 2 (BST2) in breast cancer with bone metastasis. BMC Cancer. 2009;9:102.

Lewis-wambi JS, Cunliffe HE, Kim HR, Willis AL, Jordan VC. Overexpression of CEACAM6 promotes migration and invasion of oestrogen-deprived breast cancer cells. Eur J Cancer. 2008;44(12):1770-9.

Martin TA, Watkins G, Jiang WG. The Coxsackie-adenovirus receptor has elevated expression in human breast cancer. Clin Exp Med. 2005;5(3):122-8.

Wang F, Liu X, Yang P, Guo L, Liu C, Li H, Long S, Shen Y, Wan H. Loss of TACSTD2 contributed to squamous cell carcinoma progression through attenuating TAp63-dependent apoptosis. Cell Death Dis. 2014;5:e1133.

Ji C, Zhao Y, Kou YW, Shao H, Guo L, Bao CH, Jiang BC, Chen X, Dai JW, Tong YX, Yang R, Sun W, Wang Q. Cathepsin F Knockdown Induces Proliferation and Inhibits Apoptosis in Gastric Cancer Cells. Oncol Res. 2017;

Stephenson AJ, Kattan MW, Eastham JA, Dotan ZA, Bianco FJ Jr, Lilja H, Scardino PT. Defining biochemical recurrence of prostate cancer after radical prostatectomy: a proposal for a standardized definition. J Clin Oncol. 2006;24(24):3973-8.

Wever EM, Draisma G, Heijnsdijk EA, Roobol MJ, Boer R, Otto SJ, de Koning HJ. Prostate-specific antigen screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer-Rotterdam. J Natl Cancer Inst. 2010;102(5):352-5.

Avery KN, Blazeby JM, Lane JA, Neal DE, Hamdy FC, Donovan JL. Decision-making about PSA testing and prostate biopsies: a qualitative study embedded in a primary care randomised trial. Eur Urol. 2008;53(6):1186-93.

Prendergast GC, Smith C, Thomas S, Mandik-Nayak L, Laury-Kleintop L, Metz R, Muller AJ. Indoleamine 2,3-dioxygenase pathways of pathogenic inflammation and immune escape in cancer. Cancer Immunol Immunother. 2014;63(7):721-35.

Munn DH, Mellor AL. IDO in the Tumor Microenvironment: Inflammation, Counter-Regulation, and Tolerance. Trends Immunol. 2016;37(3):193-207.

Toussi A, Stewart-Merrill SB, Boorjian SA, Psutka SP, Thompson RH, Frank I, Tollefson MK, Gettman MT, Carlson RE, Rangel LJ, Karnes RJ. Standardizing the Definition of Biochemical Recurrence after Radical Prostatectomy-What Prostate Specific Antigen Cut Point Best Predicts a Durable Increase and Subsequent Systemic Progression?. J Urol. 2016;195(6):1754-9.

Crawford ED, Bennett CL, Andriole GL, Garnick MB, Petrylak DP. The utility of prostate-specific antigen in the management of advanced prostate cancer. BJU Int. 2013;112(5):548-60.

Loupakis F, Cremolini C, Masi G, Lonardi S, Zagonel V, Salvatore L, Cortesi E, Tomasello G, Ronzoni M, Spadi R, Zaniboni A, Tonini G, Buonadonna A, Amoroso D, Chiara S, Carlomagno C, Boni C, Allegrini G, Boni L, Falcone A. Initial therapy with FOLFOXIRI and bevacizumab for metastatic colorectal cancer. N Engl J Med. 2014;371(17):1609-18.

Dai W, Gupta SL. Molecular cloning, sequencing and expression of human interferon-gamma-inducible indoleamine 2,3-dioxygenase cDNA. Biochem Biophys Res Commun. 1990;168(1):1-8.

Muller AJ, Duhadaway JB, Donover PS, Sutanto-ward E, Prendergast GC. Inhibition of indoleamine 2,3-dioxygenase, an immunoregulatory target of the cancer suppression gene Bin1, potentiates cancer chemotherapy. Nat Med. 2005;11(3):312-9.

Lundgaard GL, Daniels NE, Pyndiah S, Cassimere EK, Ahmed KM, Rodrigue A, Kihara D, Post CB, Sakamuro D. Identification of a novel effector domain of BIN1 for cancer suppression. J Cell Biochem. 2011;112(10):2992-3001.

Wang J, Jia Y, Zhao S, Zhang X, Wang X, Han X, Wang Y, Ma M, Shi J, Liu L. BIN1 reverses PD-L1-mediated immune escape by inactivating the c-MYC and EGFR/MAPK signaling pathways in non-small cell lung cancer. Oncogene. 2017;36(45):6235-6243.

Dang CV. MYC on the path to cancer. Cell. 2012;149(1):22-35.
Sharma Y, Ahmad A, Bashir S, Elahi A, Khan F. Implication of protein tyrosine phosphatase SHP-1 in cancer-related signaling pathways. Future Oncol. 2016;12(10):1287-98.

Tamir I, Dal porto JM, Cambier JC. Cytoplasmic protein tyrosine phosphatases SHP-1 and SHP-2: regulators of B cell signal transduction. Curr Opin Immunol. 2000;12(3):307-15.

Bakker AB, Wu J, Phillips JH, Lanier LL. NK cell activation: distinct stimulatory pathways counterbalancing inhibitory signals. Hum Immunol. 2000;61(1):18-27.
Bléry M, Olcese L, Vivier E. Early signaling via inhibitory and activating NK receptors. Hum Immunol. 2000;61(1):51-64.

Schneeberger VE, Ren Y, Luetteke N, Huang Q, Chen L, Lawrence HR, Lawrence NJ, Haura EB, Koomen JM, Coppola D, Wu J. Inhibition of Shp2 suppresses mutant EGFR-induced lung tumors in transgenic mouse model of lung adenocarcinoma. Oncotarget. 2015;6(8):6191-202.