UC San Diego
UC San Diego Previously Published Works

Title

Human-specific polymorphic pseudogenization of SIGLEC12 protects against advanced cancer progression

Permalink

Journal

ISSN

Authors

Siddiqui, Shoib S
Vaill, Michael
Do, Raymond
et al.

Publication Date

DOI

RESEARCH ARTICLE

FASEB BioAdvances WILEY

Human-specific polymorphic pseudogenization of *SIGLEC12* protects against advanced cancer progression

Shoib S Siddiqui[1] | Michael Vaill[1] | Raymond Do[1] | Naazneen Khan[1] |
Andrea L Verhagen[1] | Wu Zhang[2] | Heinz-Josef Lenz[2] | Teresa L Johnson-Pais[3] |
Robin J Leach[3,4] | Gary Fraser[5] | Charles Wang[5] | Gen-Sheng Feng[1] |
Nissi Varki[1] | Ajit Varki[1]

[1]Departments of Medicine, Cellular and Molecular Medicine, and Pathology, Glycobiology Research and Training Cente and Center for Academic Research and Training in Anthropogeny, University of California, San Diego, CA, USA

[2]University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA, USA

[3]Department of Urology, University of TX Health Science Center, San Antonio, TX, USA

[4]Departments of Cell Systems and Anatomy, University of TX Health Science Center, San Antonio, TX, USA

[5]School of Public Health, Loma Linda University, Loma Linda, CA, USA

Correspondence
Ajit Varki, Glycobiology Research and Training Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0687, USA.
Email: a1varki@ucsd.edu

Present address
Shoib S Siddiqui, Department of Biotechnology, American University of Ras Al Khaimah (AURAK, American University of Ras Al Khaimah Road, Al Burairat Area, Ras Al Khaimah, UAE

Abstract

Compared with our closest living evolutionary cousins, humans appear unusually prone to develop carcinomas (cancers arising from epithelia). The *SIGLEC12* gene, which encodes the Siglec-XII protein expressed on epithelial cells, has several uniquely human features: a fixed homozygous missense mutation inactivating its natural ligand recognition property; a polymorphic frameshift mutation eliminating full-length protein expression in ~60%–70% of worldwide human populations; and, genomic features suggesting a negative selective sweep favoring the pseudogene state. Despite the loss of canonical sialic acid binding, Siglec-XII still recruits Shp2 and accelerates tumor growth in a mouse model. We hypothesized that dysfunctional Siglec-XII facilitates human carcinoma progression, correlating with known tumorigenic signatures of Shp2-dependent cancers. Immunohistochemistry was used to detect Siglec-XII expression on tissue microarrays. PC-3 prostate cancer cells were transfected with Siglec-XII and transcription of genes enriched with Siglec-XII was determined. Genomic *SIGLEC12* status was determined for four different cancer cohorts. Finally, a dot blot analysis of human urinary epithelial cells was established to determine the Siglec-XII expressors versus non-expressors. Forced expression in a *SIGLEC12* null carcinoma cell line enriched transcription of genes associated with cancer progression. While Siglec-XII was detected as expected in ~30%–40% of normal epithelia, ~80% of advanced carcinomas showed strong expression. Notably, >80% of late-stage colorectal cancers had a functional *SIGLEC12* allele, correlating with overall increased mortality. Thus, advanced carcinomas are much more likely to

occur in individuals whose genomes have an intact *SIGLEC12* gene, likely because the encoded Siglec-XII protein recruits Shp2-related oncogenic pathways. The finding has prognostic, diagnostic, and therapeutic implications.

KEYWORDS

advanced carcinoma, dot blot, immunohistochemistry, pseudogenization, *SIGLEC12*

1 | INTRODUCTION

Humans appear unusually prone to develop carcinomas (cancers arising from epithelial cells), compared with our closest evolutionary cousins ("great apes").[1-3] Here, we show an unexpected human-specific connection between advanced carcinomas and a member of the CD33-related family of Siglecs (Sialic acid-binding Ig-like lectins) receptors.[4-6] Siglecs are cell-surface receptors typically expressed on innate immune cells, and binding ligands bearing sialic acids (Sias), a family of glycans prominently present at the terminal end of the glycan chains on cell surface and extracellular glycoconjugates.[7] Most CD33-related Siglecs (CD33rSiglecs) have immunoinhibitory functions mediated by immunoreceptor tyrosine-based inhibitory motifs (ITIMs) and ITIM-like motifs in the cytosolic tail.[4-6] Upon binding sialic acid ligands, these intracellular signaling motifs recruit protein tyrosine phosphatases Shp1 and Shp2, which subsequently participate in a variety of signaling pathways in immune cells to influence cellular activation.[4-6]

Unlike other CD33-related Siglecs, the literature on Human Siglec-XII (encoded by the gene *SIGLEC12*) is sparse.[8-12] In fact, it has been largely ignored and even excluded from major reviews on Siglec biology,[6,13-15] because the protein and the locus encoding *SIGLEC12* are atypical in several ways. First, the protein has two amino-terminal V-set domains,[9] compared with only one in all other Siglecs. Second, there is a human-universal mutation of critical arginine residues in both V-set domains, rendering it unable to recognize Sias (hence, the use of the Roman numeral XII for the protein, instead of Arabic numerals for functional Siglecs). Third, the Arg—>Cys mutation of the V-set 1 domain is not present in orthologs of closely related "great apes" (chimpanzee, baboon, gorilla, and orangutan).[8] Fourth, chimpanzee Siglec-12 preferentially recognizes a form of Sia (Neu5Gc) that was lost from the human lineage due to an independent fixed mutation of *CMAH*[8]. Fifth, the *SIGLEC12* gene harbors a common polymorphic frameshift mutation causing truncation of Siglec-XII and/or alternate splicing[16] that causes loss of expression of full-length protein[11] in the majority of humans. Finally, while the wild-type protein is expressed in some tissue macrophages, it is not found on other blood cell types, and is instead more prominent on epithelial cell surfaces.[11]

At first glance, the above features suggest a nonfunctional protein in the process of being eliminated from humans by pseudogenization. However, forced expression of human Siglec-XII in a genetically null human carcinoma cell line led to enhanced tumor growth in nude mice.[11] Furthermore, while human Siglec-XII does not recognize Sias, it still has ITIM and ITIM-like domains in the cytosolic tail that can be phosphorylated to recruit Shp1 and Shp2 phosphatases.[9] Finally, genome-wide analysis of signals of selection in human populations identified polymorphisms that introduce nonsense-mediated decay into human genes, including *SIGLEC12*.[17] The human *SIGLEC12* locus appears to be undergoing selection favoring a null and/or truncated form, a possible example of the "less-is-more" hypothesis first proposed by Olson.[18]

Here, we focus on Siglec-XII expression in tumor and normal epithelia, identify genes upregulated upon Siglec-XII expression, address the predictive value of *SIGLEC12* status in cancer cohorts, and provide further evidence suggesting ongoing selection for the null state. Finally, we report a simple urine test to screen for the minority of individuals capable of full-length Siglec-XII expression.

2 | MATERIALS AND METHODS

2.1 | Immunohistochemistry studies

Multi-tissue array slides were obtained from US Biomax (Rockville, Maryland), which were completely anonymized and consisted of normal human and cancer tissues. A second set of multi-tissue array slides were obtained from Novus Bio, which contained a variety of malignancies (about 476 different types) and also a set of normal multi-tissue array. The sections were de-paraffinized and blocked for endogenous biotin and peroxidase. The heat-induced epitope retrieval was performed with citrate buffer pH 6. A five-step signal amplification method was used which includes application of mouse monoclonal anti-Siglec-XII antibody (clone 276), followed by biotinylated donkey anti-mouse, horseradish peroxidase (HRP), Streptavidin, followed by application of the enzyme biotinyl tyramide, and then, labeled Streptavidin. The AEC kit (Vector) was used as substrate, nuclear counterstain was with Mayer's hematoxylin, and the slides were

aqueous mounted for digital photographs, taken using the Olympus BH2 microscope.

2.2 | Buccal swab

Healthy volunteers were recruited, and their buccal swab samples were used for DNA isolation with institutional review board (IRB) approval issued by the University of California, San Diego (UCSD). Before collection of the swab, the donors were asked to remove the mucous layer of their cheek by rubbing sterile gauze against it. Subsequently a sterile cotton tip was rubbed on the inner cheek cells for genomic DNA isolation. Genomic DNA was isolated using the ChargeSwitch Buccal Cell gDNA isolation kit (Invitrogen, Cat No.-CS11021) according to the manufacturer's instructions. The PCR amplification for *SIGLEC12* gene was performed using the primers: Forward 5′-CAATGCAGAAGTCCGTGACGGTGCAGG-3′ and reverse 5′-AGGATCAGGAGGGGCATCCAAGGTGC-3′. The Phusion High-Fidelity Polymerase kit was used according to the manufacturer's instructions. The DNA amplicon was purified using QIAquick PCR purification kit (Qiagen, Cat no.-28106) and it was sent for sequencing at Eton Bio, San Diego using the sequencing primer: 5′-CTCTCTCTGGTGTCTCTGATGC-3′ (reverse).

2.3 | Dot blot using urine from healthy donors

Healthy volunteers donated 50 ml of first morning urine according to the IRB approved study. The urine sample was centrifuged at room temperature for 10 min at 500xg. The supernatant was removed, and cell pellet resuspended in 100 μl PBS. The sample was applied onto nitrocellulose membrane and immobilized by applying negative pressure. The membrane was blocked using 50% Licor solution (cat no-927-40000) +50% PBST (PBS+0.01%Tween). After blocking, primary anti-Siglec-XII antibody (clone 1130) was applied at a dilution of 1:100–1:500. This clone of antibody have been used and characterized before.[11] The primary antibody dilution was performed in 90% Licor Solution +10%PBST and incubation was carried out for 1 hour at RT. The membrane was then washed with 10 ml of PBST three times for 5 min each. After washing, the membrane was incubated with anti-mouse-Licor-800 antibody at a dilution of 1:10000 in 90% Licor Solution +10%PBST. The secondary antibody incubation was performed for 1 hour at RT in dark. After incubation the membrane was washed with PBST three times for 5 min followed by two times with PBS for 5 min. The band on the membrane was visualized by using Licor fluorescence scanning machine.

Here, only PBS was used as a negative control and Siglec-XII-Fc was used as a positive control.

2.4 | *SIGLEC12* frameshift mutation in seventh-day adventist group

The Seventh-day Adventist group is a diverse population group where the key carcinogenesis risk factors are less prevalent, such as consumption of red meat, alcohol, and smoking. The genomic DNA was isolated from the peripheral blood cells of 53 cancer patients and 54 age-matched control subjects. The frameshift deletion mutation of *SIGLEC12* was analyzed by first PCR amplifying the *SIGLEC12* locus using the primers 5′-ACCCCTGCTCTGTGGGAGAGT-3′ (forward) and 5′AGGATCAGGAGGGGCATCCAAGGTGC-3′ (reverse). The PCR was performed using Phusion High-Fidelity Polymerase kit. The amplified product was purified using the QIAquick PCR purification kit (Qiagen, cat no.-28106) and sent for sequencing to EtonBio, San Diego, USA. The sequencing was performed using the primer: 5′-CTCTCTCTGGTGTCTCTGATGC-3′ (reverse).

2.5 | RNA-sequence analysis

PC-3 and PC-3-SigXII expressing cells were cultured to confluency in T25 flasks and mRNA was extracted from the cells using the Qiagen RNeasy plus mini kit (Cat no.-74134). Transcriptomic analysis was performed on RNA libraries prepared from *SIGLEC12* and control PC3 cells using the TruSeq RNA Library Prep Kit v2. Each cell line was used to prepare four separate technical replicate libraries for sequencing. Libraries were sequenced at 1×50 bp on HiSeq 4000 (Illumina). Reads were mapped to human reference genome Hg19 using STAR v2.5.3a.[19] Mapped reads were counted at the gene level using featureCounts v1.5.2[20] and counts were analyzed using DESeq2 v1.14.1.[21] Differentially expressed genes with a p value ≤ 0.05 and fold change ≥ 2 were then selected for further examination and gene set enrichment analysis using the GSEA software,[22] the MSigDB v7.0 oncogenic signatures collection (C6),[23] and the Siglec-XII or Shp2 expression status as phenotype with 1,000 permutations.

2.6 | Statistical analysis

Graph Prism pad 5.0 was used. The chi-square test was performed on immunohistochemistry data, different cancer cohorts and a p value <0.05 was considered as significant. For the RNA-Seq the two-way ANOVA was used as the statistically

significant value. The *p* value <0.05 and fold change of 2 was used as a cutoff for assessing the differentially expressed genes.

2.7 | Population genetics analysis

Human genomes were accessed from the 1000 Genomes Project server (www.1000genomes.org/). Bed coordinates defining the *SIGLEC12* genomic regions were retrieved from build hg19 using the University of California, Santa Cruz (UCSC), genome browser. A region containing *SIGLEC12* gene in three different populations of West Africa, Northern European, and East Asian ancestry (YRI, CHB, and CEU), using the selection tools pipeline.[24] Statistical tests such as frequency-based method (Tajima's *D*) and population differentiation-based methods (F_{ST}) among three different populations were analyzed.[24] Each test is suited to detect selection at different timescales. Tajima's *D* is a commonly used summary of the site-frequency spectrum (SFS) of nucleotide polymorphism data and is based on the difference between two estimators of θ (the population mutation rate $4Ne\mu$): nucleotide diversity that is the average number of pairwise differences between sequences, and Watterson's estimator, based on the number of segregating sites. A negative Tajima's D signifies an excess of low-frequency polymorphisms, and indicates a population size expansion, selective sweep, and/or positive selection, or negative selection. A positive Tajima's D value indicates a decrease in population size and/or that balancing selection.[25] On the contrary, the estimator of population differentiation (F_{ST}), compares the variance of allele frequencies within and between populations.[26] While large values of F_{ST} at a locus indicate complete differentiation between populations, which suggests directional selection, small values indicate the lack of differentiation in populations being compared, which might be an indicator of directional or balancing selection in both.[27] Human genome raw data for *SIGLEC12*[28] was utilized for detecting Selective Sweep using SweepFinder2[29] which implements a composite likelihood ratio (CLR) test.[30] The CLR uses the variation of the SFS of a region to compute the ratio of the likelihood of a selective sweep at a given position to the likelihood of a null model without a selective sweep. Tajima's *D* and sweep scans were visualized in Excel and F_{ST} were visualized in R studio platform and examined for evidence of deviation from the null expectation.

3 | RESULTS

3.1 | Expression of genes associated with cancer progression in a Siglec-XII expressing prostate cancer cell line

Supporting the relevance of Siglec-XII expression in advanced cancer, we noted that in tissue sections where both malignant and adjacent normal tissue were present,

Siglec-XII expression was higher in the malignant cells (one such example is shown in Figure 1A). To begin to explore the mechanism of action of this cell-surface receptor, we took advantage of our earlier model system, Siglec-XII nonexpressing PC-3 prostate carcinoma cells, which were transfected with a vector causing expression of full-length Siglec-XII. We had already observed larger tumors when this PC-3-Siglec-XII cell line was injected subcutaneously into the flanks of athymic nude mice, as compared to PC-3 cells transfected with vector alone.[11] We now compared the RNA expression profiles between these two cell lines and found many genes to be differentially expressed (Figure 1B,C). Importantly, these differentially expressed genes were enriched for those known to play a role in cancer biology. A few of those upregulated were *IDO1* (Indoleamine 2,3-Dioxygenase 1)[31]; *LCP1* (Lymphocyte Cytosolic Protein 1)[32]; *BST2* (Bone Marrow Stromal Cell Antigen 2)[33]; and *CEACAM6* (Carcinoembryonic Antigen-Related Cell Adhesion Molecule 6),[34] which are all involved in cancer progression. Among the downregulated genes related to cancer progression were *CXADR* (Coxsackievirus and adenovirus receptor)[35]; *TACSTD2* (Tumor-Associated Calcium Signal Transducer 2)[36]; *CTSF* (Cathepsin F)[37]; and, *ZNF43* (Zinc Finger Protein 43)[38] (Figure 1D). Taken together these data support the notion that Siglec-XII expression may facilitate late-stage carcinoma progression in humans. A full list of differentially expressed genes is in Supplementary Data 1.

3.2 | Enrichment analysis shows that similar gene sets are upregulated in Siglec-XII expressing cell lines and Shp2-expressing cell lines

To query which molecular pathways are altered by Siglec-XII expression status, we performed gene set enrichment analysis (GSEA) on the expression profiles produced for each of the two PC3 cell lines. The GSEA result shows that transcriptional changes in Siglec-XII expressing cells affect the expression of many gene sets found in the Molecular Signature Database (MSigDB).[22] Of relevant interest are gene sets found in the Oncogenic Signatures collection, which were generated based on data produced by perturbing known cancer genes. The most dramatically enriched oncogenic signatures in Siglec-XII expressing cells include a set of genes altered in KRAS-addicted cancers,[39] and a set of TAZ-associated genes found to be enriched in high-grade tumors.[40]

To predict whether these gene set enrichments may be related to Siglec-XII activation of Shp2 signaling, we performed the same analysis on gene expression profiles of cancer cell lines that were found to be either dependent on Shp2 or independent of Shp2 in the development of resistance to MEK
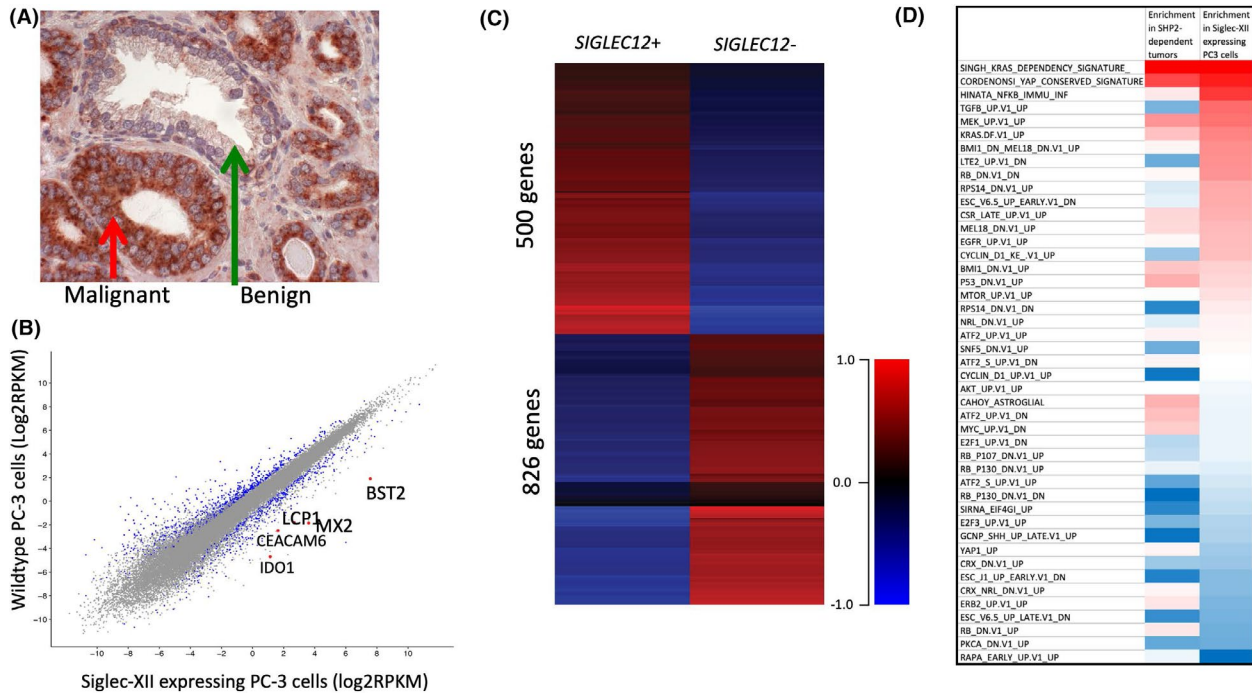
FIGURE 1 Siglec-XII Induction of expression of genes associated with cancer progression. A, Example of tissue sections with adjacent normal and malignant cells from a prostate cancer patient. B, Gene Expression in Siglec-XII transfected prostate cancer cells versus sham transfection (n = 4). Differentially expressed genes highlighted in blue, and genes not differentially expressed are in grey color. A fold change of 2 and *p* value <0.05 was used as a cutoff. C, The heatmap shows the differentially expressed genes in the Siglec-XII expressing PC-3 cell line versus parental PC-3 cells (n = 4). D, Siglec-XII GSEA shows same top pathway expression as Shp2 positive tumors

inhibition.[41] Expression data for these samples was downloaded from the Gene Expression Omnibus (NCBI GEO accession number GSE121117). This analysis revealed that many of the same sets of genes that are enriched in Shp2-dependent cancers, are also enriched in our PC-3 cells with forced expression of Siglec-XII (Figure 1D). The full results of the enrichment analysis are available in the supplementary data.
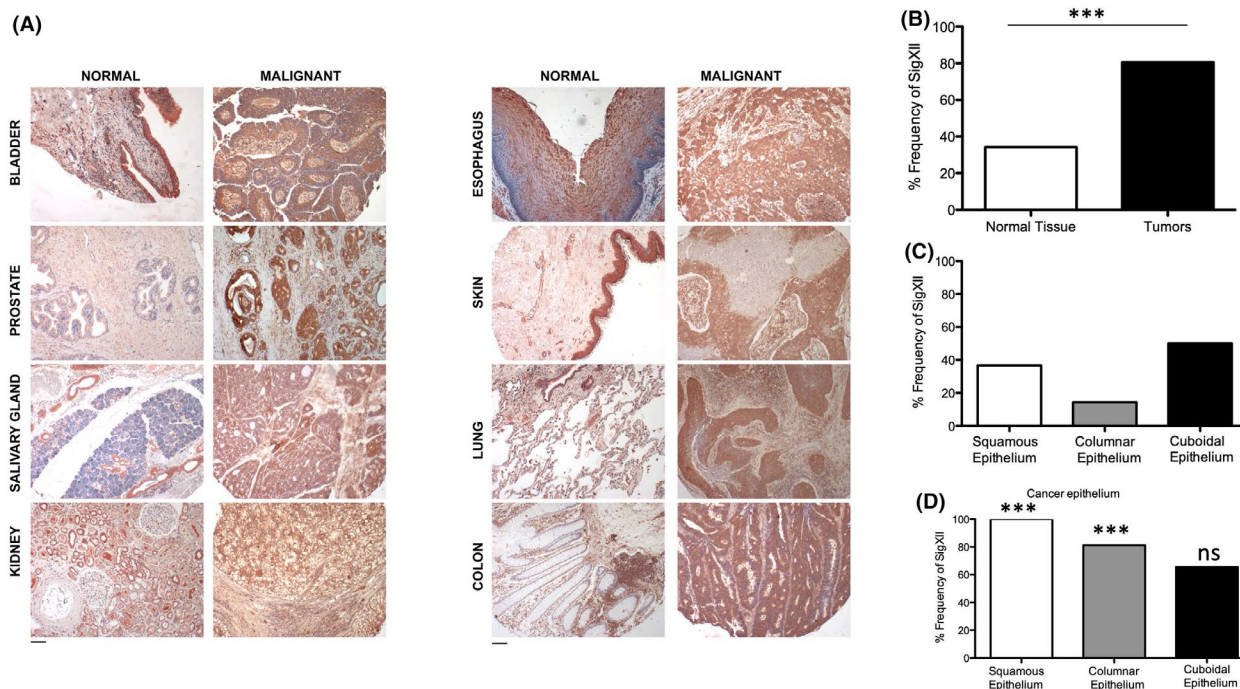
3.3 | Enhanced expression and unexpectedly high frequency of Siglec-XII in carcinomas

Immunohistochemical analyses for Siglec-XII showed low to moderate level expression in normal epithelia in a commercially available normal multi-tissue array with sample positivity of ~35%. As the majority of human genomes harbor a homozygous null-state abrogating full-length protein expression, this low frequency is as expected. In contrast, in a multi-tissue array from the same source with multiple malignancies, we found an abundance of expression in carcinomas (malignancies arising from epithelia) (see examples in Figure 2A), with a much higher than expected frequency of Siglec-XII in cancers (~80%) as compared to normal tissue (Figure 2B,C). Remarkably, 100% of the squamous cell carcinomas were positive (Figure 2D). This result was obtained from a mixed population group aged between 21

and 75 years. For this immunohistochemical analyses anti-Siglec-XII antibody clone 276 was used, which have been used and characterized earlier.[11] For analysis of subsets of tumor types, the samples were divided into squamous, columnar, cuboidal, neural, and "uncategorized" (which included endothelium, mesothelium, and endocrine glands). Between 10 and 34 of each of these categories for both carcinoma and normal tissues were included in analysis. A paired *t*-test was used to determine significance between frequency of Siglec-XII+staining in normal or carcinoma samples ($p < 0.01$). Tumors included in the "malignant" multi-tissue array are all likely to be advanced stage carcinomas. Given the prognosis of advanced carcinomas, our finding implies that minority of individuals who can express full-length Siglec-XII may be at the highest risk for dying with advanced carcinomas. A second panel of tissues, obtained from an independent source, and subjected to similar staining, confirmed this finding (Table 1).

3.4 | No correlation between *SIGLEC12* genomic status and frequency or progression of early stage cancers

Next, we asked if Siglec-XII expression predicts early carcinoma risk or progression in a well-defined population. We

(A)

FIGURE 2 Enhanced Expression and Unexpectedly High Frequency of Siglec-XII in carcinomas. A, Expression of Siglec-XII studied in normal (benign) and cancer (malignant) human tissues using mouse monoclonal antibody clone 276 (See Materials and Methods). Representative examples of positive samples are shown. B, Frequency of Siglec-XII detection on normal and cancer tissues (n = 97 for normal tissues and n = 85 for tumor samples, ***p value <0.001). C, Normal epithelium divided into squamous (n = 35), columnar (n = 14), and cuboidal (n = 34). D, Carcinoma epithelium also divided into squamous (n = 22), columnar (n = 16), and cuboidal (n = 32)

had already reported that the incidence of prostate cancer was not different between men with different *SIGLEC12* genotypes.[11] From the same cohort, there is now a minimum of 5-year follow-up available for many of these patients categorized into no evidence of disease (NED); Biochemical recurrence (BCR) and Metastasis (Met). There was no clear correlation of *SIGLEC12* status with the progression of these early stage carcinomas (Figure 3A). Of course, most of these cases were originally diagnosed by a measuring prostate-specific antigen (PSA), which picks up many early stage cases that never progress in the lifetime of the individual.[42] Indeed, if we compare the patients with a poor outcome, versus those with no evidence of disease recurrence following prostatectomy (NED), we find that most of the patients (84 out of 122) detected by PSA screening did not have disease progression at the time of follow-up.

Seventh-day Adventists are members of a religious sect that do not smoke or consume alcohol and have a largely vegetarian diet with limited intake of red meat.[43] As usual risk factors for cancer are limited, carcinoma incidence is much lower than in the general population. We genotyped the common *SIGLEC12* frameshift insertion mutation on genomic DNA from the peripheral blood cells of 54 Seventh-day Adventist cancer patients and 53 non-cancer patients (age and sex-matched). While we found more Siglec-XII expressers in

the cancer group, this trend suggesting that Siglec-XII expressers may be more prone to develop carcinomas was not statistically significant (Figure 3B). Notably, many of these cancers were diagnosed at an early stage. Taken together, the data above suggest that the genomic status of *SIGLEC12* may not be correlated with the early cancer risk, but rather with late progression.

3.5 | High frequency of *SIGLEC12* expression in advanced colorectal cancer cohort and correlation with overall survival

Given the lack of significant correlation between *SIGLEC12* status and carcinoma risk or early stage carcinomas, we reasoned that there might instead be a correlation with late-stage cancers. Indeed, in two stage IV colorectal cancer cohorts FIRE3 (592 patients from Germany and Austria)[44] and TRIBE (508 patients from Italy)[45] >80% of patients expressed Siglec-XII based on the frameshift mutation (Figure 3C,D). This recapitulates our initial immunohistochemistry-based findings. Furthermore, to see if *SIGLEC12* status has prognostic value, we checked overall survival in relation to Siglec-XII expression. Interestingly, in FIRE3 the overall survival increased from 28 months to 51 months between Siglec-XII expressers and non-expressers (Figure 3E,F),

TABLE 1 Second tumor microarray confirms unexpected high frequency of Siglec-XII expression in carcinomas

	Total	Expressors	% Positive
Squamous Carcinomas			
Cervix	21	19	90%
Lung	18	17	94%
Skin	3	3	100%
Esophagus	3	1	33%
Head and Neck	17	15	88%
Salivary gland	4	3	75%
Adenocarcinomas			
Stomach	24	10	42%
Colon and Rectum	22	14	64%
Prostate	6	5	83%
Kidney	16	13	81%
Pancreas	5	2	40%
Breast	14	14	100%
Endometrium	16	15	94%
Other Malignancies			
Bladder	14	11	79%
Melanoma	9	8	89%
Hepatocellular carcinoma	13	4	31%
Lymphoma	16	4	25%
Bone sarcoma	17	8	47%
Esophagus	4	1	25%
Salivary gland	2	2	100%
Skin	3	1	33%
Bladder	4	1	25%
Kidney	5	1	20%
Lung	4	1	25%
Liver	3	0	0%
Breast	3	2	67%
Colon	4	0	0%
Pancreas	2	1	50%
Prostate	2	0	0%
Stomach	2	1	50%
Lymph node	3	0	0%

As shown in the main text, advanced carcinomas are more than twice as likely to express Siglec-XII protein. To independently verify this unexpected result, a second panel of tissues obtained from another source was subjected to immunohistochemical analysis, including samples categorized as squamous carcinomas, adenocarcinomas, and other malignancies. As shown in this table the total frequency of Siglec-XII expression in these carcinoma samples was 70% (166/238), while the frequency expression in normal control tissues included in the analysis was 27% (11/41), in alignment with the expected frequency of individuals in the global population with intact Siglec-XII open reading frame. These data are categorized by cell type of origin. Notably with the exception of melanomas, other non-epithelial malignancies such as lymphomas, sarcomas, and hepatocellular carcinomas had lower rates of expression, consistent with the population frequency of the frameshift allele.

that is, a correlation between Siglec-XII expression and poor prognosis in late-stage colorectal cancer.

3.6 | Further evidence for selection at the *SIGLEC12* locus

Earlier work suggested that this locus might be undergoing selection favoring the null state.[17] To test this hypothesis, we examined the population level genetic variation and evidence of natural selection in and around the *SIGLEC12* locus and carried out tests that aimed to detect a selective sweep,[29] deviation from neutrality,[25] and population differentiation[46] in three ethnic groups (YRI, CHB, and CEU).[24] Analysis of site-frequency spectrum provided a composite likelihood ratio indicative of a soft "selective sweep" acting on the gene throughout the overall human population (Figure 4A). Additionally, the common frameshift mutation (rs66949844) was present adjacent to this area. We also estimated population differentiation (measured as Wright's index of fixation; F_{ST}) and found moderately high F_{ST} values (0.3) compared to the average F_{ST} for genome-wide autosomal markers throughout the human population.[47] The high F_{ST} value indicates stark differentiation of populations, which suggests directional selection (see Figure Legend, Figure 4B). Furthermore, we found an excess of rare alleles relative to a model of neutral evolution as indicated by negative Tajima's D values (Figure 4C); especially in CHB and YRI (African ancestry population). Individually none of these signals were very strong, but together, they suggest selection for the null state (note that ongoing selection for a null state would favor inactivating mutations, which would tend to mask the usual signatures of a selective sweep).

3.7 | Dot blot analysis of bladder epithelial cells for detection of Siglec-XII status

All the population studies above were handicapped by the fact that in addition to the common frameshift insertion mutation, we found other less common mutations that would nullify Siglec-XII expression. For example, another deleterious mutation (rs16982743) was observed at a global frequency of 18.6% that changes glutamine to a stop codon (Q>*) at the 29th position.[12] Thus, biallelic whole exome sequencing of *SIGLEC12* genomic DNA would be required for rigorous population studies. Even this approach could be confounded by selection for noncoding mutations that suppress gene expression in a given allele with an intact open reading frame. In addition, there is evidence for an alternately spliced truncated form of the protein.[16] To facilitate future population and
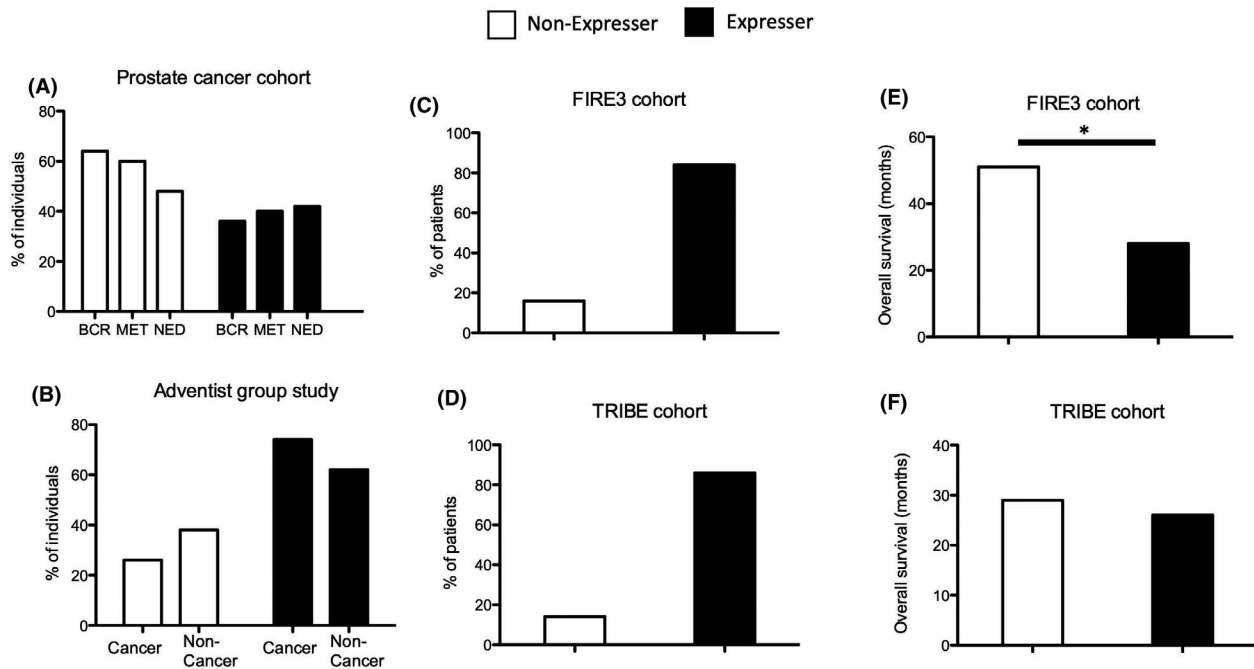
FIGURE 3 Correlation between *SIGLEC12* genomic status and frequency or progression, only of late-stage cancers. A, Prostate cancer patients diagnosed with PSA test followed up after 5 years. (NED-No evidence of disease: n = 84, BCR- Biochemical Cancer Recurrence: n = 28 and Met-Metastasis: n = 10). B, Seventh-Day Adventist population where environmental risk factors for cancer are minimal. The percentage of patients with cancer and without cancer is shown to be either *SIGLEC12*-/- (non-expresser, n for cancer = 14, n for non-cancer = 20) or *SIGLEC12*+/- and *SIGLEC12*+/+ (expresser, n for cancer = 40 and n for non-cancer = 33). C and D, Percentage of patients who are Siglec-XII expressers (*SIGLEC12*+/- and *SIGLEC12*+/+) or non-expressers (*SIGLEC12*-/-) in the FIRE3 and TRIBE stage IV colorectal cancer cohorts (FIRE3 cohort: expresser n = 85, non-expresser n = 16 and TRIBE cohort: expresser n = 177, non-expresser n = 27). E and F, Overall survival of colorectal cancer patients that are Siglec-XII expressers versus non-expressers (*p value <0.05)

cancer cohort studies, it would be useful to have a simple assay to rapidly detect all mutations abrogating expression, without the need to do whole exome sequencing. We took advantage of the fact that among normal epithelial tissues tested by IHC, Siglec-XII was expressed in bladder epithelium, kidney tubules and salivary gland ducts, and detected the expression of Siglec-XII in cells isolated from saliva and urine (Figure 5A). It was determined via buccal swab genomic analysis that the *SIGLEC12* genomic status (*SIGLEC12* +/- or -/-) correlates with either Siglec-XII expression (+/-) or no expression (-/-). As expected, Siglec-XII expression in cells obtained from the urine of multiple healthy donors showed expression of Siglec-XII in the +/- genotypes and no expression in the Siglec-XII null genotypes. While there was a significant background in samples from saliva, results from dot blot screening of urinary cells were very clean (a typical example is shown in Figure 5B).

4 | DISCUSSION

We focused our initial work on a common polymorphic frameshift mutation in human populations with an allele

frequency ranging from 38% in sub-Saharan Africans to 86% in the Native American population.[11] An earlier study[17] suggested selection on *SIGLEC12* based on the inactivating mutation rs16982743. However, another frameshift mutation (rs66949844) was present in the human population at an allele frequency of 59%.[48] Overall, this region of *SIGLEC12* showed reduced genetic diversity, which was supported by a sweep scan.[29] These findings were concordant with results from a study in six different human populations[49] showing a soft sweep in a region of *SIGLEC12*. The presence of excess rare alleles in and around a genomic region is also an indicator of a low level of population differentiation[47,50] further indicating the presence of purifying selection or balancing selection. Negative selection in S*IGLEC12* region was also evident from the result of Tajima's D (TD) especially in the YRI population (African ancestry).

Previous studies showed that while the non-Sia-binding Siglec-XII can be expressed in *SIGLEC12* mutated PC-3 human prostate cancer cells, efforts to transfect the chimpanzee version of *SIGLEC12* or the arginine-restored version of human *SIGLEC12* were not successful.[11] This could be either due to rapid turnover or selection against expression in vitro. Regardless, the non-Sia-binding full-length human Siglec-XII
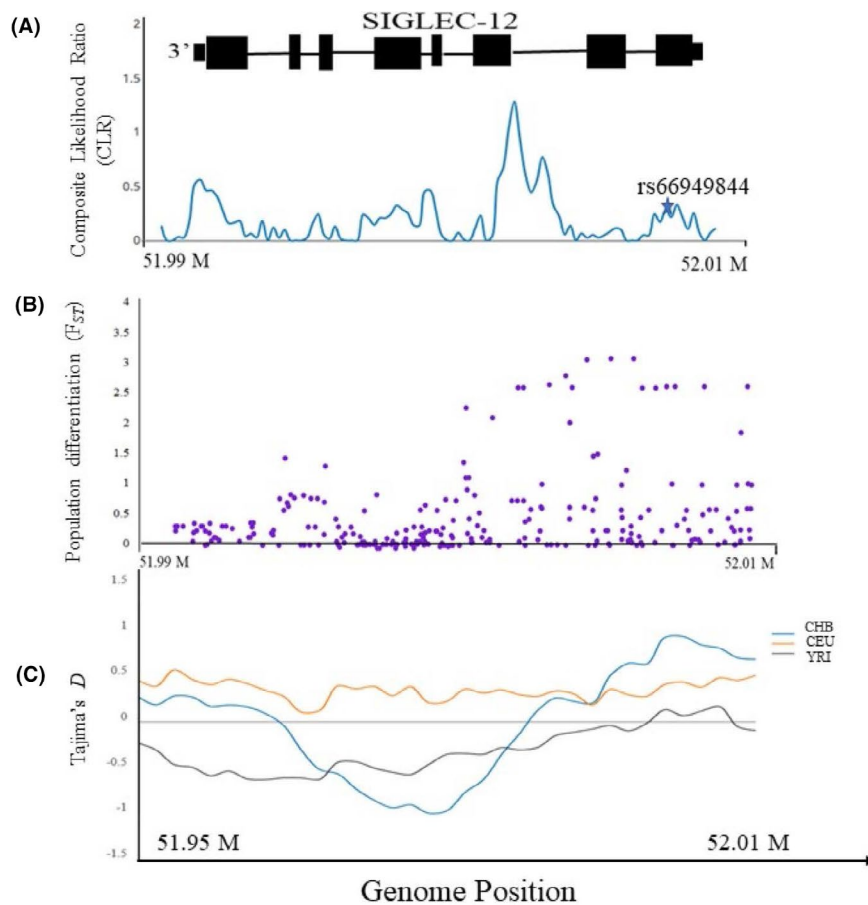
FIGURE 4 Population Genetic Analysis shows signatures of selection in and around the *SIGLEC12* locus. A, Signatures of "Selective sweep" in *SIGLEC12* in human population. The composite likelihood ratio (CLR) test of selective sweep based on the site-frequency spectrum (SFS) is shown in blue. The star shown in the figure denotes the location of frameshift mutation. (Note. Schematic representation of *SIGLEC12* gene on top). B, Estimation of Population differentiation "F_{ST}" (global) in three human populations (CHB_CEU_YRI). The purple dots represents F_{ST} values. C, Estimation of Tajima's *D* in and around region of *SIGLEC12* in three human population are shown in different colors (Blue = CHB, Orange = CEU, and Grey = YRI)

is clearly different functionally, allowing persistent surface expression in malignant cells by as yet unknown mechanisms. Chimpanzee and arginine-restored human *SIGLEC12* both display a preference for binding *N*-Glycolylneuraminic acid (Neu5Gc) (Figure 6A)[8] which is absent in humans due to a homozygous fixed deletion in the gene *CMAH*.[51] After losing its preferred ligand in an ancestral prehuman species, it is possible that Siglec-12 was susceptible to exploitation by a Neu5Gc-presenting pathogen (Figure 6B) or some other harmful form of activation, driving the fixation of the arginine mutation to produce the non-Sia-binding full-length Siglec-XII found in humans today (Figure 6C). It is worth pointing out that the reason the arginine mutation was fixed in humans remains unknown, and the consequences of the Sia-binding chimpanzee Siglec-12 deserve further study to contribute to our understanding of this evolutionary scenario. Finally, with the unusual derived trait of post-reproductive life span in modern humans, we propose that selection for survival in late life is driving the complete loss of the human

SIGLEC12 gene, as evidenced by the genomic signatures we report. Ongoing selection for null-state alleles may be acting to relieve the increased risk of advanced carcinomas produced by the archaic nonfunctional Siglec protein (Figure 6D).

It is also important to reemphasize that the arginine and frameshift mutations of Siglec-XII do not occur in chimpanzees. Humans and chimpanzees are very similar in terms of genomic sequences but different phenotypically. Remarkably, while cancers are common in humans, few are reported in chimpanzees, and are usually lymphomas or soft tissue tumors, unlike those that arise in humans, who are instead prone to epithelial carcinomas.[1–3] Here, immunohistochemistry analyses indicate that Siglec-XII is highly expressed in advanced carcinomas, as compared to normal epithelium. Considering the multiple mutations reported[11,12] and others possible in the population, the overall expression in ~35% in normal samples seems reasonable to represent the general mixed population. On the contrary, the high abundance of Siglec-XII in advanced carcinomas is remarkable, as is the

(A)

(B)
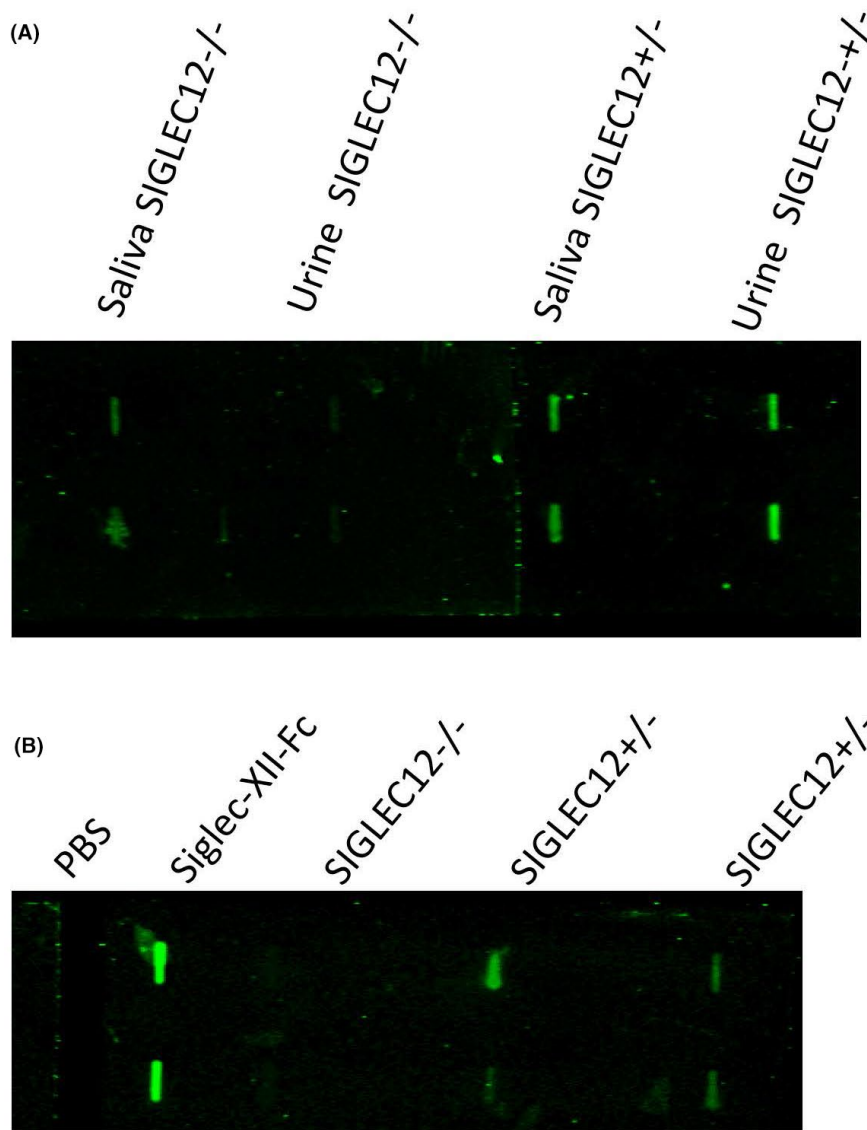
FIGURE 5 Simple analysis of urine allows screening for Siglec-XII protein expression status. A, Urine and saliva samples were obtained from healthy individuals and used for checking protein expression of Siglec-XII by the dot blot. B, Urine samples from multiple healthy donors were used to check protein expression of Siglec-XII. One typical example is shown. (The whole blot was corrected uniformly for brightness using Photoshop, to match the visual appearance)

high frequency of expression at >80%, in epithelial carcinomas. A second panel shows that certain types of carcinomas are even more likely to be found in Siglec-XII expresses; all squamous carcinomas show an 88% frequency, with lung showing 94% and cervix 90% (see Table 1).

To explore molecular mechanisms of Siglec-XII in cancer progression, we compared RNA expression patterns between *SIGLEC12* null PC-3 prostate cancer cells with and without transfection with a construct encoding Siglec-XII. One of the top hits among the upregulated genes in RNA-Seq was IDO-1 (Indoleamine 2,3-dioxygenase 1), an enzyme involved in conversion of tryptophan to kynurenine metabolites. This enzyme is highly upregulated in many types of cancers. It is known that a decrease in the levels of tryptophan and an increase in the levels of kynurenine leads to immunosuppression

and enhanced tumor growth.[31,52,53] The molecular mechanisms for the effects of IDO-1 overexpression point toward maintenance of immunosuppression in tumor microenvironment, due to depletion of effector T cells and enrichment of regulatory T cells.[31] There has been a recent focus on IDO-1 targeting through small molecule inhibitors in preclinical and clinical settings.[52,54]

While Siglec-XII has lost its Sia-binding property, it still has the ability to recruit Shp1 and Shp2.[9] Shp2 (encoded by *PTPN11*) is a well-characterized oncogene that elicits cell growth, proliferation, tumorigenesis, and metastasis.[55] Over-activation and activating mutations of Shp2 are known to be involved in breast cancer, leukemia, and gliomas.[55–57]

While not the primary objective of this study, we used RNA-sequencing data from our Siglec-XII expressing PC-3
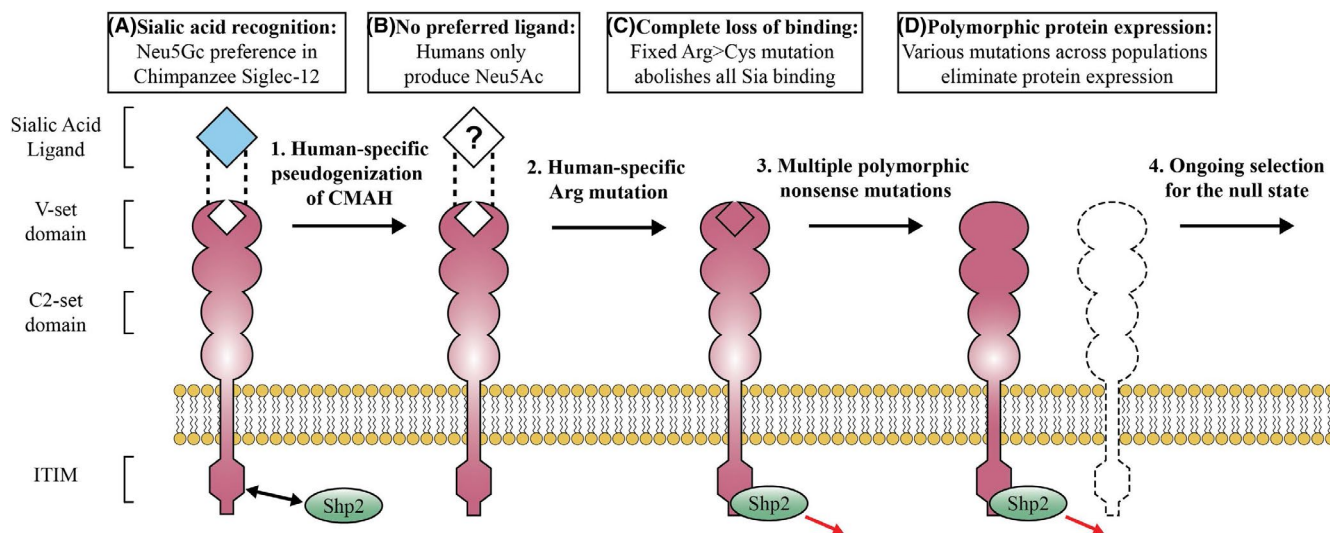
FIGURE 6 Proposed evolutionary history of human Siglec-XII. A, The last common human-chimpanzee ancestor and modern chimpanzees had a functional CMAH enzyme and an abundance of Neu5Gc-terminated cell-surface glycans. Chimpanzee Siglec-12 recognizes Neu5Gc through an arginine-dependent binding pocket in its terminal V-set domain. B, After divergence from chimpanzees CMAH was completely inactivated in human ancestor, leaving Siglec-12 with no endogenous ligand. C, Another unknown evolutionary event fixed a mutation in the critical arginine rendering human "Siglec-XII" incapable of binding any sialic acids, however, the full-length protein continues to recruit Shp2 and alter gene expression. D, Modern humans are experiencing purifying selection acting to increase the frequency of common null-state alleles across population

cells to briefly investigate the hypothesis that Siglec-XII expression enhancers tumor growth via Shp2. Using Gene Set Enrichment Analysis, we identified which pathways are altered by Siglec-XII protein expression. This analysis revealed that among the most dramatically enriched pathways are KRAS and YAP/TAZ. Notably, when we conducted a parallel analysis from a previously published transcriptomic study of Shp2-dependence the same pathways were the most highly enriched.[41] Upregulation of these well-known oncogenic pathways in individuals with an intact *SIGLEC12* allele may explain the molecular mechanisms underpinning our discovery of increased frequency of Siglec-XII protein in advanced carcinomas.

We also performed population studies on four cancer cohorts. The first was a prostate cancer cohort we had studied earlier.[11] While a 5-year follow-up for 122 patients was recorded in this cohort, we still did not see any positive correlation between *SIGLEC12* pseudogenization and outcome. One reason for this negative result may be that out of 122 patients only 10 developed metastasis (poor prognosis) and this might not be a sufficient number to find the relevance of *SIGLEC12* in prognosis. The second cohort we tested was a Seventh-day Adventist group and the lack of correlation could be due to two reasons. First, most of the cancer patients in this group represented early stage cancer, where the effect of Siglec-XII is not pronounced. Second, many cancer risk factors such as intake of red meat, smoking, drinking alcohol etc., are minimal in this cohort, so it might be that Siglec-XII plays a role only when other obvious risk factors are involved. Overall,

it appears that Siglec-XII does not play a role in early stage carcinomas. In other populations, we discovered that the null state of the gene affects the prognosis of advanced carcinomas. Therefore, Siglec-XII expression is more likely to contribute to the advancement of benign neoplasia to deadly malignancies.

According to the well-established theoretical concept, natural selection occurs in prereproductive or reproductive individuals.[58] However, humans are a rare species that have prolonged post-reproductive life span (PRLS), and according to the "grandmother hypothesis" inclusive fitness of infertile elderly caregivers can determine the fate of helpless grandchildren.[59,60] We report selection acting on the *SIGLEC12* locus in human populations. This could be caused by deleterious fitness consequences of advanced carcinomas, which mostly occur late in middle to late life. To the best of our knowledge, our work is the first potential example of inclusive fitness effects selecting for cancer suppression, supporting a function for PRLS in humans. In contrast, an expansion in the number of p53 genes maybe providing late life protection against cancer risk in long-lived elephants.[61] However, elephants do not have a PRLS, so the underlying selection mechanism must be different.

This first study of a very unusual phenomenon raises even more questions than answers. We do not know if there is still any definite ligand for Siglec-XII. It does not bind with Sias, but we cannot rule out its interaction with another unknown ligand(s). Conversely, we can also consider the hypothesis that this is a constitutively active receptor, which does not

need any ligand for its activation. This aspect of Siglec biology is not extensively studied. Second, we did not study the signaling pathways mediated by SiglecXII-Shp2 axis. Third, we have not yet done the gene expression analysis in PC-3 cells with Shp2 inhibitors. Moreover, a knockdown of *SIGLEC12* in a Siglec-XII expressing cell line will be useful. These are important aspects of Siglec-XII biology, which will be focused in further studies. Regardless, we have previously noted that triggering of endocytosis by antibodies against this receptor can deliver toxins into the cell.[11] In analogy to the targeting of Siglec-3/CD33 human leukemias,[62] a similar approach could be taken for treatment of late-stage carcinomas. Our simple urine screen should be of value in these and other clinical studies.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors declare no competing interest with the content of this manuscript.

AUTHORS CONTRIBUTION

SSS, NV, and AV conceived the idea and designed the experiments. SSS, RD, MV, NK, ALV, WZ, HJL, TLJ, RJL, GF, NV, GSF, and CW performed the experiments and analyzed the results. NV and AV supervised the work. SSS, MV, NK, and AV wrote the manuscript. All authors read the manuscript and approved it.

ORCID

Shoib S Siddiqui https://orcid.org/0000-0002-7028-290X
Michael Vaill https://orcid.org/0000-0002-4714-9254
Raymond Do https://orcid.org/0000-0003-0984-2696
Naazneen Khan https://orcid.org/0000-0002-4386-1703
Andrea L Verhagen https://orcid.org/0000-0002-0535-4006
Wu Zhang https://orcid.org/0000-0001-5477-9679
Heinz-Josef Lenz https://orcid.org/0000-0003-2178-9568
Teresa L Johnson-Pais https://orcid.org/0000-0001-7580-1206
Robin J Leach https://orcid.org/0000-0002-3201-5579
Gary Fraser https://orcid.org/0000-0002-5577-3524
Charles Wang https://orcid.org/0000-0001-8861-2121
Gen-Sheng Feng https://orcid.org/0000-0003-1255-4708
Nissi Varki https://orcid.org/0000-0001-9839-1254
Ajit Varki https://orcid.org/0000-0002-2206-975X

REFERENCES

1. Schmidt RE. Systemic pathology of chimpanzees. *J Med Primatol*. 1978;7:274-318.
2. Puente XS, Velasco G, Gutierrez-Fernandez A, Bertranpetit J, King MC, Lopez-Otin C. Comparative analysis of cancer genes in the human and chimpanzee genomes. *BMC Genom*. 2006;7:15.
3. Varki NM, Varki A. On the apparent rarity of epithelial cancers in captive chimpanzees. *Philos Trans R Soc Lond B Biol Sci*. 2015;370(1673):20140225.
4. Bochner BS, Zimmermann N. Role of siglecs and related glycan-binding proteins in immune responses and immunoregulation. *J Allergy Clin Immunol*. 2015;135:598-608.
5. Adams OJ, Stanczak MA, von Gunten S, Läubli H. Targeting sialic acid-Siglec interactions to reverse immune suppression in cancer. *Glycobiology*. 2017.
6. Varki A, Schnaar RL, Crocker PR. I-Type Lectins. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH, editors. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2017.
7. Varki A, Schnaar RL, Schauer R. Sialic Acids and Other Nonulosonic Acids. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH, editors. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2017.
8. Angata T, Varki NM, Varki A. A second uniquely human mutation affecting sialic acid biology. *J Biol Chem*. 2001;276:40282-40287.
9. 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;276:23816-23824.
10. Foussias G, Taylor SM, Yousef GM, Tropak MB, Ordon MH, Diamandis EP. Cloning and molecular characterization of two splice variants of a new putative member of the Siglec-3-like subgroup of Siglecs. *Biochem Biophys Res Commun*. 2001;284:887-899.
11. Mitra N, Banda K, Altheide TK, et al. SIGLEC12, a human-specific segregating (pseudo)gene, encodes a signaling molecule expressed in prostate carcinomas. *J Biol Chem*. 2011;286:23003-23011.
12. McDonough CW, Gong Y, Padmanabhan S, et al. Pharmacogenomic Association of Nonsynonymous SNPs in SIGLEC12, A1BG, and the Selectin Region and Cardiovascular Outcomes. *Hypertension*. 2013;62:48-54.
13. Macauley MS, Crocker PR, Paulson JC. Siglec-mediated regulation of immune cell function in disease. *Nat Rev Immunol*. 2014;14:653-666.
14. Bornhöfft KF, Goldammer T, Rebl A, Galuska SP. Siglecs: A journey through the evolution of sialic acid-binding immunoglobulin-type lectins. *Dev Comp Immunol*. 2018;86:219-231.
15. Zhou JY, Oswald DM, Oliva KD, Kreisman LSC, Cobb BA. The Glycoscience of Immunity. *Trends Immunol*. 2018;39:523--535.
16. Flores R, Zhang P, Wu W, et al. Siglec genes confer resistance to systemic lupus erythematosus in humans and mice. *Cell Mol Immunol*. 2019;16:154-164.
17. Yngvadottir B, Xue Y, Searle S, et al. A genome-wide survey of the prevalence and evolutionary forces acting on human nonsense SNPs. *Am J Hum Genet*. 2009;84:224-234.
18. Olson MV. When less is more: gene loss as an engine of evolutionary change. *Am J Hum Genet*. 1999;64:18-23.

19. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15-21.

20. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923-930.

21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.

22. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545-15550.

23. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27:1739-1740.

24. Pybus M, Dall'Olio GM, Luisi P, et al. 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. *Nucleic Acids Res*. 2014;42:D903-D909.

25. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–595.

26. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat Rev Genet*. 2009;10:639-650.

27. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. *Annu Rev Genet*. 2013;47:97-120.

28. Huber CD, DeGiorgio M, Hellmann I, Nielsen R. Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol Ecol*. 2016;25:142-156.

29. Aakhus AM, Stavem P, Hovig T, Pedersen TM, Solum NO. Studies on a patient with thrombocytopenia, giant platelets and a platelet membrane glycoprotein Ib with reduced amount of sialic acid. *Br J Haematol*. 1990;74:320-329.

30. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. *Genome Res*. 2005;15:1566-1575.

31. Zhai L, Spranger S, Binder DC, et al. Molecular pathways: targeting IDO1 and other tryptophan dioxygenases for cancer immunotherapy. *Clin Cancer Res*. 2015;21:5427-5433.

32. Koide N, Kasamatsu A, Endo-Sakamoto Y, et al. Evidence for critical role of lymphocyte cytosolic protein 1 in oral cancer. *Sci Rep*. 2017;7:43379.

33. Mahauad-Fernandez WD, DeMali KA, Olivier AK, Okeoma CM. Bone marrow stromal antigen 2 expressed in cancer cells promotes mammary tumor growth and metastasis. *Breast Cancer Res*. 2014;16:493.

34. Chiang WF, Cheng TM, Chang CC, et al. Carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6) promotes EGF receptor signaling of oral squamous cell carcinoma metastasis via the complex N-glycosylation. *Oncogene*. 2018;37:116-127.

35. Stecker K, Vieth M, Koschel A, Wiedenmann B, Röcken C, Anders M. Impact of the coxsackievirus and adenovirus receptor on the adenoma-carcinoma sequence of colon cancer. *Br J Cancer*. 2011;104:1426-1433.

36. Wang F, Liu X, Yang P, et al. Loss of TACSTD2 contributed to squamous cell carcinoma progression through attenuating TAp63-dependent apoptosis. *Cell Death Dis*. 2014;5:e1133.

37. Ji C, Zhao Y, Kou YW, et al. Cathepsin F knockdown induces proliferation and inhibits apoptosis in gastric cancer cells. *Oncol Res*. 2018;26:83-93.

38. Jen J, Wang YC. Zinc finger proteins in cancer progression. *J Biomed Sci*. 2016;23:53.

39. Singh A, Greninger P, Rhodes D, et al. A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival. *Cancer Cell*. 2009;15:489-500.

40. Cordenonsi M, Zanconato F, Azzolin L, et al. The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. *Cell*. 2011;147:759-772.

41. Ahmed TA, Adamopoulos C, Karoulia Z, et al. SHP2 drives adaptive resistance to ERK signaling inhibition in molecularly defined subsets of ERK-dependent tumors. *Cell Rep*. 2019;26:65-78.e5.

42. Eastham JA, Riedel E, Scardino PT, et al. Variation of serum prostate-specific antigen levels: an evaluation of year-to-year fluctuations. *JAMA*. 2003;289:2695-2700.

43. Beeson WL, Mills PK, Phillips RL, Andress M, Fraser GE. Chronic disease among Seventh-day Adventists, a low-risk group. Rationale, methodology, and description of the population. *Cancer*. 1989;64:570-581.

44. Heinemann V, von Weikersthal LF, Decker T, et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): a randomised, open-label, phase 3 trial. *Lancet Oncol*. 2014;15:1065-1075.

45. Loupakis F, Cremolini C, Masi G, et al. Initial therapy with FOLFOXIRI and bevacizumab for metastatic colorectal cancer. *N Engl J Med*. 2014;371:1609-1618.

46. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984;38:1358-1370.

47. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res*. 2002;12:1805-1814.

48. Auton A, Brooks LD, Durbin RM, et al, 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526:68-74.

49. Schrider DR, Kern AD. S/HIC: robust identification of soft and hard sweeps using machine learning. *PLoS Genet*. 2016;12:e1005928.

50. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. Natural selection has driven population differentiation in modern humans. *Nat Genet*. 2008;40:340-345.

51. Chou HH, Takematsu H, Diaz S, et al. A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proc Natl Acad Sci USA*. 1998;95:11751-11756.

52. Li F, Zhang R, Li S, Liu J. IDO1: An important immunotherapy target in cancer treatment. *Int Immunopharmacol*. 2017;47:70-77.

53. Zhai L, Ladomersky E, Lenzen A, et al. IDO1 in cancer: a Gemini of immune checkpoints. *Cell Mol Immunol*. 2018;15(5):447-457.

54. Prendergast GC, Malachowski WP, DuHadaway JB, Muller AJ. Discovery of IDO1 inhibitors: from bench to bedside. *Cancer Res*. 2017;77:6795-6811.

55. Bollu LR, Mazumdar A, Savage MI, Brown PH. Molecular pathways: targeting protein tyrosine phosphatases in cancer. *Clin Cancer Res*. 2017;23:2136-2142.

56. Xu R, Yu Y, Zheng S, et al. Overexpression of Shp2 tyrosine phosphatase is implicated in leukemogenesis in adult human leukemia. *Blood*. 2005;106:3142-3149.

57. Zhou X, Coad J, Ducatman B, Agazie YM. SHP2 is up-regulated in breast cancer cells and in infiltrating ductal carcinoma of the breast, implying its involvement in breast oncogenesis. *Histopathology*. 2008;53:389-402.

58. Ungewitter E, Scrable H. Antagonistic pleiotropy and p53. *Mech Ageing Dev*. 2009;130:10-17.

59. Hawkes K, Coxworth JE. Grandmothers and the evolution of human longevity: a review of findings and future directions. *Evol Anthropol*. 2013;22:294-302.

60. Coxworth JE, Kim PS, McQueen JS, Hawkes K. Grandmothering life histories and human pair bonding. *Proc Natl Acad Sci USA*. 2015;112:11806-11811.

61. Abegglen LM, Caulin AF, Chan A, et al. Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *JAMA*. 2015;314:1850-1860.

62. Lamba JK, Chauhan L, Shin M, et al. CD33 splicing polymorphism determines gemtuzumab ozogamicin response in de novo acute myeloid leukemia: report from randomized phase III Children's Oncology Group Trial AAML0531. *J Clin Oncol*. 2017;35(23):2674-2682.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Dataset S1