Chapter 1
Human-Specific Evolutionary Changes in the Biology of Siglecs

Flavio Schwarz *, Jerry J. Fong *, and Ajit Varki

Introduction

Sialic acid-recognizing immunoglobulin-like lectins (Siglecs) are cell surface receptors that bind sialic acids, a class of monosaccharides found on the outermost end of glycans on a variety of glycoconjugates (Varki and Angata 2006; Crocker et al. 2007; Pillai et al. 2012). Sequence similarities and evolutionary conservation place Siglecs in two categories (Fig. 1.1). Sialoadhesin (Siglec-1), CD22 (Siglec-2), myelin-associated glycoprotein (Siglec-4), and Siglec-15 have orthologs in all mammalian species and relatively low (20–25 %) sequence similarity. In contrast, the CD33-related Siglecs (CD33rSiglecs, including Siglecs-3, -5 to -14, -16, and -17 in primates) form a large, rapidly evolving subfamily of genes that expanded in mammals by duplications involving a primordial cluster of *SIGLEC* genes (Fig. 1.2). Several mechanisms were involved in the rapid evolution of the CD33rSiglec subfamily: exon shuffling, gene duplication, gene conversion, deletion leading to pseudogenization, altered expression, and adaptive amino-acid substitutions in sialic acid recognition domains (Angata et al. 2004; Altheide et al. 2006).

CD33-related Siglecs are primarily expressed on immune cells (Lock et al. 2004), but specific members of the family are also found on other cell types in humans. For instance, Siglec-XII is present on epithelial cells (Mitra et al. 2011); Siglec-6 is expressed in the placental trophoblast (Brinkman-Van der Linden et al. 2007); and, Siglec-5 and -14 are found on amniotic epithelium (Ali et al. 2014). Siglec-3 and -11

*Author contributed equally with all other contributors.

F. Schwarz • J.J. Fong • A. Varki (✉)
Departments of Medicine, and Cellular and Molecular Medicine, Glycobiology Research and Training Center, University of California San Diego, La Jolla, CA, USA
e-mail: fschwarz@ucsd.edu; jjf002@ucsd.edu; a1varki@ucsd.edu

A. Chakrabarti, A. Surolia (eds.), *Biochemical Roles of Eukaryotic Cell Surface Macromolecules*, Advances in Experimental Medicine and Biology 842,
DOI 10.1007/978-3-319-11280-0_1

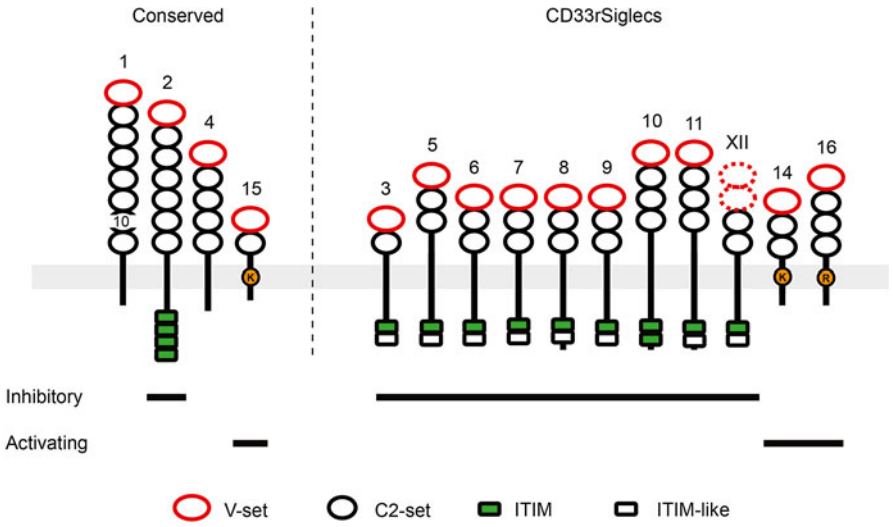

Fig. 1.1 The family of human Siglec receptors. Conserved (*left*) and CD33-related (*right*) Siglecs are cell surface receptors with a variable number of extracellular immunoglobulin-like domains. The outermost domain (V-set, in *red*) binds to sialylated molecules through a critical arginine residue. The V-set domains of Siglec-XII cannot bind sialic acid due to a mutation in a critical arginine residue, and are indicated with *dotted lines*. The transmembrane segment of Siglec-14, -15, and -16 contain a basic amino acid (lysine or arginine) that can interact with negatively charged amino acids of protein adapters. Siglecs may contain intracellular signaling motifs such as ITIM or ITIM-like. Structural elements for each protein were derived from the Uniprot database

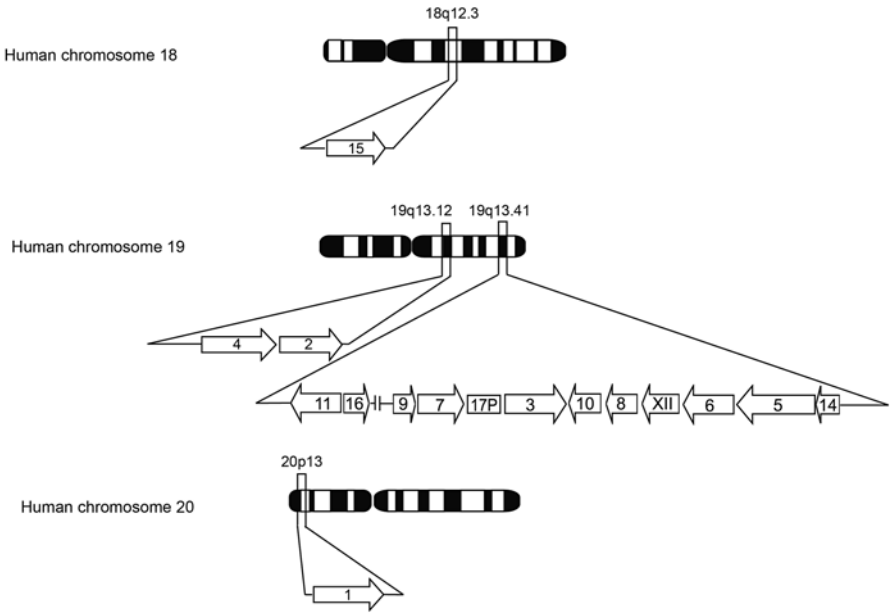

Fig. 1.2 Genomic localization of human *SIGLEC* genes. While genes encoding conserved Siglecs are found on different chromosomes, the CD33-related *SIGLEC* genes are clustered on the chromosome 19, along with multiple *SIGLEC* pseudogenes (only 17P is shown). Information on the localization, length and orientation of the genes was derived from the hg38 dataset of the UCSC Genome Browser

are also expressed in microglia, resident immune cells of the central nervous system, and influence their activity (Linnartz-Gerlach et al. 2014; Hayakawa et al. 2005; Malik et al. 2013; Griciuc et al. 2013).

Structurally, Siglecs are type-I membrane proteins with an extracellular N-terminus, a single transmembrane span, and an intracellular C-terminus (Varki and Angata 2006). The extracellular portion is composed of a V-set immunoglobulin-like domain, which binds to sialic acid-containing ligands, and one or more underlying C2-set immunoglobulin-like domains. The intracellular segment of many Siglecs contains immunoreceptor tyrosine-based inhibitory motifs (ITIMs), which can be phosphorylated by Src family kinases upon external ligand binding and recruit SHP-1 or SHP-2 tyrosine phosphatases (Pillai et al. 2012). These events lead to blockade of MAP kinase phosphorylation and eventually attenuate the cellular inflammatory response. Inhibitory Siglecs may also contain ITIM-like motifs (Crocker et al. 2007; Crocker and Varki 2001). Phosphorylation of the tyrosine residues of ITIM-like and ITIM motifs may occur sequentially and be required for efficient recruitment of SHP-1 or SHP-2 (Tourdot et al. 2013). Some of the CD33rSiglec ITIM-like motifs also contain a consensus sequence similar to those found in the signaling lymphocytic activation molecule (SLAM) receptors (Cannons et al. 2011). However, the contribution of these SLAM-like motifs of Siglec to the modulation of signaling has not been characterized, and it is not clear whether Siglecs can interact with SAP or EAT2 proteins.

More recently, a subset of Siglecs were found to lack ITIM motifs and instead engage DNAX-activation protein of 12 kDa (DAP12) through a positively charged residue in their transmembrane domains (Angata et al. 2006; Cao et al. 2008; Kameda et al. 2013; Takamiya et al. 2013; Ishida-Kitagawa et al. 2012). Upon engaging their ligands, these immuno-activating Siglecs augment inflammation by phosphorylation of the immunoreceptor tyrosine-based activating motifs (ITAMs) of DAP12 and enhancement of the MAP kinase signaling cascade (Lanier 2009). Thus, primate Siglecs may alternatively be categorized into three groups based on the features of the transmembrane and cytoplasmic tails: Siglecs (-1 and -4) that lack standard signaling motifs and are likely involved in adhesion and/or phagocytosis; Siglecs (-2, -3, -5, -6, -7, -8, -9, 10, -11, -XII) with the general ITIM consensus I/V/L/SxYxxL/V (Ravetch and Lanier 2000); and, Siglecs (-13, -14, -15, -16) that contain a positively charged amino acid (lysine or arginine) in the transmembrane span, which supports recruitment of a homodimer of DAP12.

Interestingly, some inhibitory and activating CD33rSiglec exist as pairs in primates. The sialic acid-binding properties of Siglecs -11/-16 and of -5/-14 are kept virtually identical by gene conversion (Angata et al. 2006; Wang et al. 2012a), but each member of these pairs mediates opposing signaling events (Ali et al. 2014). Moreover, some CD33rSiglecs have inactive alleles that segregate at intermediate frequency in some human populations, but the functional significance of these alleles is not yet known (Angata 2014; Cao et al. 2008; Mitra et al. 2011; Wang et al. 2012a). The general hypothesis is that CD33rSiglec variants adjust the inflammatory responsiveness of human immune cells, which need to limit reactions against "self," but also face pathogens expressing sialic acid-containing surface polysaccharides that can subvert the inhibitory Siglecs by mimicking self signals (Cao and Crocker

2011; Varki 2010; Barclay and Hatherley 2008; Barreiro and Quintana-Murci 2010). For instance, group B *Streptococcus* serotype III (GBS-III) produces a capsular polysaccharide containing Neu5Acα2-3Galβ1-4GlcNAc, a structure that precisely mimics the terminal sequences of many human glycoproteins. Binding of GBS-III to Siglec-9 leads to a reduction of reactive oxygen species and inflammatory cytokines (Carlin et al. 2009b). Other GBS serotypes have even evolved sialic acid-independent binding to Siglecs (Carlin et al. 2009a). Activating Siglecs may have evolved to prevent pathogen hijacking of inhibitory Siglecs: they are similar targets with opposite cellular effects. However, in the absence of a pathogen, the binding of self-molecules to activating Siglecs could generate unwanted inflammation. The benefit of activating Siglecs may therefore be context dependent: advantageous in the presence of a pathogen, but costly in its absence. This could explain the polymorphic loss of function alleles observed in human activating Siglecs-14 and -16.

SIGLEC genes are rapidly evolving in all taxa where they exist (Angata et al. 2004; Padler-Karavani et al. 2014). However, the abundance of Siglec changes in humans seems unusually high compared with other species (Table 1.1). For example, mouse and rat Siglecs appear nearly identical, and differences among nonhuman hominids (NHH) and other old world primates seem limited so far. Some changes may be a consequence of uniquely human sialic acid biology, the most

Table 1.1 Human-specific changes in *SIGLEC* genes and their encoded Siglec proteins

GENE	Type of change	Fixed	Phenotypic effects	Disease relevance
SIGLEC1	Expression change	Yes?	Altered expression in spleen lymphoid follicles	Altered response to sialylated pathogens?
SIGLEC3	Alternative splicing	No	Change in relative expression of two isoforms	Alters development of late onset Alzheimer's disease
SIGLEC5	Expression change	Yes?	T-cell over-reactivity due to low expression	Susceptibility to T-cell mediated disease?
	Gene conversion		Uniquely human expression in amniotic epithelium	Fetal susceptibility to sialylated bacteria
SIGLEC6	New expression	Yes	Uniquely human expression in placental trophoblast	Up-regulation in preeclampsia
SIGLEC9	Expression and ligand preference	Yes	Simultaneous balance between recognizing "self" sialic acids and escape from sialylated pathogens	Altered response to sialylated pathogens
SIGLEC11	Gene conversion	Yes	Uniquely human expression in microglia	Modulation of neurotoxicity in neurodegenerative disease. Impact on ovarian physiology?
	Change in ligand preference	Yes	Reduced binding to sialylated glycans. Novel ligands in brain and ovary	

(continued)

Table 1.1 (continued)

GENE	Type of change	Fixed	Phenotypic effects	Disease relevance
SIGLEC12	Mutation in arginine	Yes	Loss of sialic acid binding properties	Increased expression in carcinomas
	1-bp insertion in ORF	No	Expression loss in some individuals	
SIGLEC13	*Alu*-mediated deletion	Yes	Escape from sialylated bacteria?	Not applicable—all humans null
SIGLEC14	Gene deletion by fusion	No	Change in baseline state of innate immune response	Alters response to lung inflammation
SIGLEC16	4-bp deletion in ORF	No	Change in baseline state of innate immune response?	Modulation of neurotoxicity in neurodegenerative disease?
SIGLEC17	1-bp deletion in ORF	Yes	Escape from sialylated bacteria?	Not applicable—all humans null

prominent being the loss of the sialic acid Neu5Gc. Humans have a non-functional *CMAH* gene, and thus cannot synthesize the common mammalian sialic acid Neu5Gc via hydroxylation of the precursor sialic acid *N*-acetylneuraminic acid (Neu5Ac) (Chou et al. 2002). If the primary function of CD33rSiglecs is to recognize host sialic acid as "Self-Associated Molecular Pattern" (SAMPs) and send inhibitory signals to the immune cells via cytosolic ITIMs, loss of the ability to synthesize Neu5Gc in human ancestors may have resulted in loss of the ability to dampen the immune response and thus excessive immune activation (Varki 2010, 2011). Therefore, it is reasonable to suppose that loss of Neu5Gc in humans was followed by compensatory changes in Siglecs to adapt their binding preference to Neu5Ac, and to reduce or alter the scope for immune activation in response to self-molecules. Human pathogens that interact with Siglecs would also be expected to compensate following the loss of Neu5Gc. In the sections that follow, we consider human-specific changes in several of the Siglecs.

Human-Specific Changes in Siglecs

Siglec-1/Sialoadhesin

Siglec-1, also called sialoadhesin, is a macrophage receptor with 17 extracellular domains, a single transmembrane span, and no cytosolic signaling motif. The V-set domain seems to recognize only Neu5Ac and not Neu5Gc, and prefers α2-3 and α2-8 linkages (Collins et al. 1997; Hartnell et al. 2001). This overall pattern fits what is often found on bacteria. Considering all these features, it was suggested that one likely conserved function is to eliminate sialylated pathogens (Crocker et al. 1997). In keeping with this, Siglec-1 is found in mice and in primates at sites that

would first encounter bacteria invading extracellular fluids, such as the sinuses of lymph nodes, spleen, and bone marrow (Crocker et al. 1991; Hartnell et al. 2001). However, Siglec-1 appears to be upregulated in the human spleen compared with the chimpanzee (Brinkman-Van der Linden et al. 2000). In fact, in chimpanzees, as in rodents, only a subset of splenic macrophages is Siglec-1 positive, whereas in humans the distribution is more widespread. One possible explanation is that humans are under increased selection pressure from bacteria that express sialic acids.

It has been shown that Siglec-1 on circulating monocytes binds to sialylated gp120 of HIV and facilitates entry of the virus into cells (Rempel et al. 2008; Zou et al. 2011). In this regard, it is interesting that HIV infection more often progresses to AIDS in humans.

Siglec-3/CD33

CD33 was first detected on human myeloid cells by a panel of monoclonal antibodies that recognize a 67 kDa glycoprotein uniquely expressed on cells of the hematopoietic system (Andrews et al. 1983). Later, human CD33 was found to bind to sialylated glycans, with a preference for Neu5Acα2-3Gal (Freeman et al. 1995), and catalogued as Siglec-3. This receptor is found on circulating monocytes, on subsets of B and activated T and NK cells, and on microglia (Hernandez-Caselles et al. 2006; Perez-Oliva et al. 2011). Notably, two forms of Siglec-3 are expressed (Hernandez-Caselles et al. 2006; Perez-Oliva et al. 2011). The full length human Siglec-3 (also named CD33M) is the 67 kDa protein that includes a V-set domain, a C2-set domain, and a transmembrane span followed by an ITIM domain. Alternative splicing of exon 2 generates an isoform CD33m that lacks the V-set domain, and is therefore unable to bind sialylated ligands.

Independent studies have shown that Siglec-3 levels are altered in patients with late-onset Alzheimer's disease (LOAD) (Bradshaw et al. 2013; Raj et al. 2014; Malik et al. 2013; Griciuc et al. 2013). A single nucleotide polymorphism (SNP) rs3865444 in the promoter region of human *SIGLEC3* associated with LOAD was shown to alter the ratio between the two Siglec-3 isoforms produced (Malik et al. 2013; Raj et al. 2014). Whereas the protective allele rs365444A results in a ratio CD33M:CD33m of 70:30, a higher expression of the full length protein CD33M was detected in LOAD brains (ratio 90:10). Interestingly, the rs365444 SNP was reported to be in high linkage disequilibrium with the rs12459419 SNP, which is physically found in the exon 2 and alters splicing efficiency of exon 2. Increased levels of CD33M (the form capable of binding sialic acid) in microglia are thought to suppress phagocytosis of Aβ42 peptide, possibly by blocking TREM2/DAP12-mediated activation, resulting in amyloid accumulation (Bradshaw et al. 2013; Griciuc et al. 2013; Malik et al. 2013). It was suggested that altered CD33 function could be involved in the presymptomatic phase of AD, in the middle or younger age (Bradshaw et al. 2013).

It is interesting to note that the complete pathology of LOAD is very rare in primates other than humans (Gearing et al. 1994; Perez et al. 2013; Varki et al. 2011). Chimpanzee Siglec-3 shares similar specificities for sialylated glycoconjugates as the human counterpart, but is detected on monocytes at lower levels than humans (Padler-Karavani et al. 2014). Currently, it is unknown whether Siglec-3 is expressed in microglia in chimpanzee, or whether the Siglec-3 mRNA transcript undergoes the same type of regulation.

So far, mice deficient in CD33 have not shown major morphological or histological abnormalities and very minor differences in biochemical and erythrocyte parameters (Brinkman-Van der Linden et al. 2003). However, mouse Siglec-3 has striking differences from the human counterpart. First, it is primarily expressed on granulocytes. Secondly, it does not bind to α2-3 sialylated ligands, but shows distinctive sialic acid-dependent binding only to the short *O*-linked glycans of certain mucins and weak binding to the sialyl-Tn epitope. Furthermore, mouse CD33 includes a positively charged amino acid in the transmembrane domain similar to activating Siglecs. Lastly, alternative splicing may generate two forms with different cytosolic tails, of which only one contains a canonical ITIM motif. The signaling properties of these two potential isoforms have not been elucidated. Overall, it seems likely that murine and human Siglec-3 receptors are functionally different.

Siglec-5 and -14

Because of ongoing gene conversion, Siglec-5 and -14 are 100 % identical in their V-set sialic acid binding domain, and differ in only one amino acid in the first underlying C2 domain. However, they transmit opposite intracellular signals upon ligand engagement: while Siglec-5 suppresses the immune response, Siglec-14 augments it (Angata et al. 2006; Ali et al. 2014).

Expression studies are complicated by the fact that nearly all known high affinity monoclonal antibodies against Siglec-5 cross-react with Siglec-14 (Angata et al. 2006). On the other hand, a monoclonal antibody that specifically recognizes only Siglec-14 with no cross reactivity to Siglec-5 has been reported (Yamanaka et al. 2009). Primates express Siglec-5 and -14 on myelomonocytic cells: neutrophils display both Siglec-5 and -14, but monocytes display only Siglec-14 under normal conditions (Yamanaka et al. 2009). However, humans, but not other primates, acquired the ability to express Siglec-5 and -14 on amnion (Ali et al. 2014): one of the few known examples of Siglecs with uniquely human expression on non-hematopoietic stem cell derived lineages. Also, chimpanzee lymphocytes such as $CD19^+$ B-cells and $CD4^+$ T-cells display relatively high levels of Siglec-5, but human T-cells display low or undetectable levels of any CD33rSiglec (Nguyen et al. 2006). Although the antibody used in this study cross-reacts with both Siglec-5 and -14, the subsequent studies demonstrating the immunosuppressive nature of the Siglec receptor on T-cell receptor activation highly suggested Siglec-5 rather than Siglec-14 (Nguyen et al. 2006). This cell-intrinsic Siglec expression difference

between humans and other primates may be a contributing factor that explains why chimpanzee T-cells survive better than human T-cells after HIV-1 infection despite being equally susceptible to the virus (Soto et al. 2012). Differences in expression of Siglec-3, -7 and -9 on T-cells (high in chimpanzee, low in humans) might also affect the outcome of HIV-1 infection (Nguyen et al. 2006).

Group B *Streptococcus* (GBS), a leading cause of neonatal sepsis and death worldwide, is particularly noteworthy for its ability to engage Siglec-5 through its cell wall anchored β-protein to dampen the pro-inflammatory response (Carlin et al. 2007, 2009a, b). Such Siglec-pathogen interactions are believed to be a major driving force in the evolution of the *SIGLEC* gene cluster. *SIGLEC14* possibly emerged from a *SIGLEC5* gene duplication event, and was converted into a DAP12-binding immunoactivating receptor as an evolutionary response to combat the pathogens that subvert Siglec-5 (Angata et al. 2006; Ali et al. 2014).

Humans also have a unique *SIGLEC14* deletion polymorphism that has not so far been observed in other primates. The polymorphism occurs at varying frequencies based on the geographic origins of the population (Yamanaka et al. 2009). The deletion apparently resulted from an in-frame gene fusion that occurred between the *SIGLEC5* and *SIGLEC14* ORFs. The new gene product encodes a Siglec-5-like immunosuppressive receptor (designated as Siglec-14/5) regulated under the fully functional *SIGLEC14* gene promoter. Individuals homozygous for the wild-type ancestral allele encoding for both Siglec-5 and -14 are represented as *SIGLEC14* +/+, and those homozygous for the deletion polymorphism are represented as *SIGLEC14* −/−. Although *SIGLEC14* +/+ neutrophils display both Siglec-5 and -14, *SIGLEC14* −/− neutrophils display only the Siglec-14/5. In comparison, while *SIGLEC14* +/+ monocytes express only Siglec-14 but not -5, *SIGLEC14* −/− monocytes lose Siglec-14 but gain Siglec-14/5. *Ex vivo* experiments confirmed that *SIGLEC14* −/− myelomonocytic cells have a dampened inflammatory response in comparison to *SIGLEC14* +/+ cells when challenged with bacteria or endotoxin (Ali et al. 2014).

The selective forces responsible for establishing the *SIGLEC14* deletion polymorphism in the human genome are as yet unknown. However, *SIGLEC14* +/+ individuals are more susceptible to developing acute exacerbation of chronic obstructive pulmonary disease (COPD) compared to *SIGLEC14* −/− individuals (Angata et al. 2013). Although this complication of pulmonary inflammation has been largely linked to cigarette smoking in modern times, acute exacerbation of lung inflammation may develop for anyone with long term exposure to air pollution such as indoor smoke. The *SIGLEC14* polymorphism allele exists in all human populations worldwide, although at different frequencies. We speculate that the invention of indoor cooking, and consequently indoor smoke, could have been a selective force favoring the fusion allele. However, the ancestral *SIGLEC14* allele is not entirely evolutionary disadvantageous for modern humans. *SIGLEC14* −/− fetuses are more susceptible to preterm labor after GBS infection when compared to *SIGLEC14* +/+ infants, independent of the mother's *SIGLEC14* genotype (Ali et al. 2014). The *SIGLEC14* polymorphism might even have been maintained by heterozygote advantage since *SIGLEC14* +/− individuals would have an intermediate propensity for inflammation.

Siglec-6

Siglec-6 is a CD33rSiglec found on primate B-cells but also uniquely on human placental trophoblast (Brinkman-Van der Linden et al. 2007). Several differences found between human and primate sequences of *SIGLEC6* promoter regions may explain the changed expression pattern. There are conflicting reports regarding the changes in expression pattern in relation to the onset of labor (Brinkman-Van der Linden et al. 2007; Rumer et al. 2013).

Although Siglec-6 normally binds sialyl-Tn (Neu5Acα2-6GalNAcα1-), it can also interact with the non-glycosylated adipose-derived hormone leptin (Patel et al. 1999). In this regard, immunohistochemical analysis revealed that both wild type and arginine-mutated Siglec-6 recognize ligands in and adjacent to the placenta, including uterine endometrium, suggesting the presence of both sialic acid dependent and independent ligands (Brinkman-Van der Linden et al. 2007). Since leptin is also secreted from the placenta, it is likely to be sialic acid independent ligand recognized by the arginine-mutated Siglec-6. A further increase in Siglec-6 trophoblast expression is also associated with the uniquely human condition preeclampsia (Winn et al. 2009; Rumer et al. 2013). Both mRNA and protein expression of Siglec-6 are increased in placentas obtained from women who had preeclampsia as compared to the control group of pre-term labor. Mechanistically, Siglec-6 ligation with Glycodelin-A suppresses ERK signaling and subsequently trophoblast invasiveness. Taken together, these data suggest that Siglec-6 contributes to the human-specific aspects of reproductive biology. It may also be worthwhile to investigate whether Siglec-6 expression and function directly correlates with other immunoregulatory complications in pregnancy such as spontaneous abortion and recurrent miscarriage (Chatterjee et al. 2014).

Siglec-9

Siglec-9 is an ITIM-containing CD33rSiglec found primarily on human neutrophils and monocytes (Zhang et al. 2000). It is also expressed weakly on $CD4^+$ and $CD8^+$ T-cells, and but found at more modest levels on B-cells. Recent literature also revealed that Siglec-9 expression defines a subset of cytotoxic NK cell population (Jandus et al. 2014). Differences in Siglec-9 expression level were also found on circulating monocytes between gorilla, humans, and chimpanzees in descending order, although still abundant in all three primates (Padler-Karavani et al. 2014). Furthermore, immunohistochemical analysis showed that only human splenic macrophages display Siglec-9, but not those from chimpanzee or gorilla.

Siglec-9 ligand binding preferences also changed somewhat during evolution. As previously discussed, since the *CMAH* gene was functionally inactivated during human evolution, humans are the only known primates incapable of naturally synthesizing Neu5Gc. Interestingly, while gorilla and chimpanzee Siglec-9 somewhat

preferred Neu5Gc sialylated ligands over Neu5Ac, human Siglec-9 slightly preferred Neu5Ac over Neu5Gc (Sonnenburg et al. 2004; Padler-Karavani et al. 2014). Thus, it is possible that Siglec-9 evolved to engage Neu5Ac with higher affinity after the human specific loss of synthesizing Neu5Gc.

In addition to Siglec-5, Group B *Streptococcus* (GBS) also hijacks Siglec-9's inhibitory properties through molecular mimicry of sialic acids (Carlin et al. 2007, 2009b). All three tested serotypes of GBS (Type Ia, Ib, and III) bound to human Siglec-9 at a stronger affinity over chimpanzee Siglec-9, but GBS was incapable of binding to baboon Siglec-9. Taken together, these properties may be a contributing factor for why GBS is a human-specific pathogen, as it can take advantage of human inhibitory Siglecs much easier than other primate Siglecs.

Siglec-9 may be a prime example of co-evolution between mammalian sialic acids, Siglecs, and the pathogens that exploit this receptor. Pathogens evolve to synthesize or acquire sialic acids identical or similar to ones displayed by the host to hide from the immune system by engaging inhibitory Siglecs. Meanwhile, the host evolves and alters their sialic acid repertoire to keep away from the pathogens. In order to recognize the "newly evolved" sialic acid SAMPs, Siglecs are also constantly adapting to keep up with this ongoing evolutionary arms race as well.

Siglec-11 and -16

SIGLEC11 and *SIGLEC16* genes are found head-to-head about 1 MB away from the *CD33rSIGLEC* gene cluster on human chromosome 19 (Angata et al. 2002; Hayakawa et al. 2005; Wang et al. 2012a; Cao et al. 2008). It has been suggested that *SIGLEC16* arose by an inverse duplication of inhibitory *SIGLEC11* and underwent subsequent pseudogenization. Then, two tandem and likely simultaneous gene conversions occurred from *SIGLEC16P* to the adjacent gene *SIGLEC11* with an intervening short segment being excluded, and ultimately resulting in the creation of an open reading frame. Both of the gene conversions have been dated to about 1–1.2 million years, after the emergence of the genus *Homo*, but prior to the emergence of the common ancestor of Denisovans and modern humans about 600,000 years ago (Wang et al. 2012a).

These extensive changes in the sequence of human Siglec-11 may explain the different affinity for sialylated glycans compared to the chimpanzee counterpart (Hayakawa et al. 2005), and the emergence of novel binding properties. For instance, Siglec-11 is detected on fibroblasts in ovaries of both human and chimpanzee (Wang et al. 2011). However, probing for Siglec-11 ligands revealed distinct and strong mast cell expression in human ovaries, and diffuse stromal ligands in chimpanzee ovaries. This dramatic difference in ligand specificity may have an impact on ovarian physiology.

One of the conversion events also changed the 5′ untranslated sequence, altering predicted transcription factor binding sites of *SIGLEC11* and, perhaps consequently, its expression pattern. Indeed, while Siglec-11 is found in both human and chimpanzee

tissue macrophages, it is expressed in brain microglia only in humans (Hayakawa et al. 2005). *In vitro* studies have indicated that Siglec-11 suppresses the levels of proinflammatory cytokines, reduces phagocytosis of apoptotic neurons and alleviates microglia neurotoxicity (Wang and Neumann 2010). Interestingly, the neuroprotective effects were dependent on polysialic acid, a polymer important for maintenance of brain plasticity (Rutishauser 2008).

Like Siglec-5 and -14, Siglec-11 and -16 are also paired receptors (Cao et al. 2008). Their V-set and the first C2-set domains are 99 % identical in the amino acid sequence, and the other two C2-set domains share about 80 % sequence identity. However, the intracellular carboxyl-terminal region of Siglec-11 contains one ITIM, whereas Siglec-16 has a transmembrane domain and a short cytosolic tail that may associate with DAP12 (Cao et al. 2008). Therefore, whereas the two proteins are likely to recognize similar ligands, this binding will allegedly lead to opposite intracellular signaling cascades. To date, Siglec-11 and -16 mediated immunomodulation has not been well characterized.

It is interesting to note that whereas *SIGLEC11* is fixed in the human population, *SIGLEC16* is often pseudogenized due to a deletion of four nucleotides in the second exon that results in frameshift and premature termination of translation (Cao et al. 2008; Wang et al. 2012a). The frequency of the functional *SIGLEC16* allele varies among populations, adding another potential layer of complexity to the modulation of the immune responses.

Siglec-XII

Human *SIGLEC12* encodes for a Siglec receptor with two V-set domains (Yu et al. 2001). However, the arginine residues critical for glycan binding are substituted in both domains (Angata et al. 2001). Therefore, this receptor is unable to bind sialic acid, and it is referred by convention as Siglec-XII. The single nucleotide substitution (C to T) at the first position of the arginine codon in the outer most V-set domain is universal in humans but not found in other primates (Angata et al. 2001). It occurred prior to the common ancestor of all modern humans and sometime after the split of the hominin lineage from that of the common ancestor of chimpanzee and bonobos. In chimpanzee, Siglec-12 is strongly expressed in macrophages, and epithelial cells of prostate, pancreas, kidney, and stomach. Human Siglec-XII is generally less expressed in the same tissues (Mitra et al. 2011).

A SNP in human *SIGLEC12* produces a single nucleotide insertion in the exon that encodes the first V-set, which changes the open reading frame and results in a polypeptide of 115 amino acids (Mitra et al. 2011). The global frequency of this mutation was found to be 0.58, with allele frequencies ranging from 0.38 in Sub-Saharan Africa to 0.86 in Native American populations. About 40 % of human individuals are homozygous for the *SIGLEC12P*; heterozygosity reaches about 30 %. Siglec-XII was also observed to be overexpressed in prostate cancer (Mitra et al. 2011). However, no association was found with the inactivating SNP and incidence

of prostate cancer. Interestingly, a study on polymorphic nonsense SNPs in the human genome found *SIGLEC12* to be one of the outliers, among 167 cases (Yngvadottir et al. 2009). In particular, it was noted that the inactivating SNP was not itself deleterious, suggesting a balancing selection or a selective sweep.

Siglec-13

Analysis of genomic BAC clones indicated that the primate *SIGLEC13* gene was missing from the human genome, but was present in chimpanzee and baboon (Angata et al. 2004). Comparative analysis of the genomic regions of human, chimpanzee, baboon and rhesus genomes identified five repetitive *Alu* elements in a 10 kb genomic region containing the *SIGLEC13* locus in the chimpanzee, baboon and rhesus genome. By contrast, the human genome includes a single composite *Alu* element occupying the region of 7 kb in the same genomic region. The composite element likely derives from recombination of *Alu* elements that resulted in the excision of *SIGLEC13* in humans. Indeed, analysis of monocytes from peripheral blood confirmed expression of Siglec-13 in chimpanzee, and absence in human monocytes (Wang et al. 2012b). Universal absence of Siglec-13 in humans was also confirmed in 28 HapMap human samples and in the common ancestral population of Neanderthals and Denisovan. It is interesting that despite recruiting DAP12, Siglec-13 reduces inflammation in response to pathogenic bacteria that can specifically interact with Siglec-13 (Wang et al. 2012b). This indicates that the ability of a Siglec receptor to alter inflammatory responses cannot be deduced only based on the primary sequence, and that the outcome of Siglec engagement might be more complex (Barrow and Trowsdale 2006; Hamerman and Lanier 2006).

Siglec-17

The primate *SIGLEC17* pseudogene (*SIGLEC17P*) exhibits high sequence similarity with *SIGLEC3* and was originally annotated as *SIGLEC3P* (Angata et al. 2004). This locus has a human-specific 1 bp deletion that alters the predicted open reading frame (ORF) and results in a truncated protein (Wang et al. 2012b). The rest of the ORF remains intact in all humans tested and the corresponding mRNA is strongly expressed in NK cells. Moreover, the human *SIGLEC17P* contains a human-unique missense mutation of the codon encoding an arginine residue that is required for sialic acid binding. The human *SIGLEC17P* allele was already present in the common ancestral populations of Neanderthals and Denisovans. Some other primate *SIGLEC17* genes underwent independent events of inactivation. *SIGLEC17* seems to be completely deleted in the rhesus and baboon genomes (Wang et al. 2012b). In contrast, New World monkeys carry a functional *SIGLEC17* ORF.

Conclusions and Perspectives

Comparative studies in mammals have revealed an expansion in the number of *SIGLEC* genes in primates, and an accumulation of multiple variations, particularly in the human lineage. Work of the last decade has revealed that pathogens exploit Siglec function by expressing sialylated and non-sialylated ligands, and thus constitute a major selective force for the evolution of *SIGLEC* genes. Therefore, pathogens likely shaped the *SIGLEC* gene family: expansion, deletion and polymorphic inactivation of activating receptors are signatures of past and ongoing selection. At the same time, current data suggest that some variations in Siglecs might have resulted in advantageous functions that were retained and contributed to human evolution. The human-specific *SIGLEC11* and *SIGLEC16* conversion events and subsequent recruitment of novel Siglec-11 dependent functions to the brain are paradigmatic and deserve further studies. The same is true of the recruitment of *SIGLEC6* expression to the placental trophoblast. Novel regulation of Siglec-3/CD33 at the V-set domain and its impact in the development of late onset alzheimer's disease calls for studies to define whether this phenomenon occurs in other primates, or if it is specific to humans. Future research should also address whether observed reduction of affinity for sialylated molecules in human Siglecs are accompanied by the emergence of alternative binding properties. Also, as the innate immune system impacts many human conditions such as cancer and obesity, it will be interesting to study the influence of Siglec polymorphic variations in the incidence and progression of various diseases. It is even possible that *SIGLEC* gene changes played key roles in population bottlenecks involved in human origins.

Acknowledgments We thank the members of the Varki lab group for fruitful discussions, and Yuko Naito-Matsui, Anel Lizcano, Shoib Siddiqui, Corinna Landig, Stevan Springer and John T. Ngo for comments on the manuscript. Research in the Varki laboratory is supported by grants from the NIH and by the NHLBI Program of Excellence in Glycosciences. F.S. is supported by a fellowship from the Novartis Foundation for medical-biological research.

References

Ali SR et al (2014) Siglec-5 and Siglec-14 are polymorphic paired receptors that modulate neutrophil and amnion signaling responses to group B Streptococcus. J Exp Med 211:1231–1242

Altheide TK et al (2006) System-wide genomic and biochemical comparisons of sialic acid biology among primates and rodents: evidence for two modes of rapid evolution. J Biol Chem 281:25689–25702

Andrews RG, Torok-Storb B, Bernstein ID (1983) Myeloid-associated differentiation antigens on stem cells and their progeny identified by monoclonal antibodies. Blood 62:124–132

Angata T (2014) Associations of genetic polymorphisms of Siglecs with human diseases. Glycobiology 24(9):785–793

Angata T, Varki NM, Varki A (2001) A second uniquely human mutation affecting sialic acid biology. J Biol Chem 276:40282–40287

Angata T et al (2002) 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 277:24466–24474
Angata T et al (2004) Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. Proc Natl Acad Sci U S A 101:13251–13256
Angata T et al (2006) Discovery of Siglec-14, a novel sialic acid receptor undergoing concerted evolution with Siglec-5 in primates. FASEB J 20:1964–1973
Angata T et al (2013) Loss of Siglec-14 reduces the risk of chronic obstructive pulmonary disease exacerbation. Cell Mol Life Sci 70:3199–3210
Barclay AN, Hatherley D (2008) The counterbalance theory for evolution and function of paired receptors. Immunity 29:675–678
Barreiro LB, Quintana-Murci L (2010) From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet 11:17–30
Barrow AD, Trowsdale J (2006) You say ITAM and I say ITIM, let's call the whole thing off: the ambiguity of immunoreceptor signalling. Eur J Immunol 36:1646–1653
Bradshaw EM et al (2013) CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. Nat Neurosci 16:848–850
Brinkman-Van der Linden ECM et al (2000) Loss of N-glycolylneuraminic acid in human evolution—implications for sialic acid recognition by siglecs. J Biol Chem 275:8633–8640
Brinkman-Van der Linden EC et al (2003) CD33/Siglec-3 binding specificity, expression pattern, and consequences of gene deletion in mice. Mol Cell Biol 23:4199–4206
Brinkman-Van der Linden EC et al (2007) Human-specific expression of Siglec-6 in the placenta. Glycobiology 17:922–931
Cannons JL, Tangye SG, Schwartzberg PL (2011) SLAM family receptors and SAP adaptors in immunity. Annu Rev Immunol 29:665–705
Cao H, Crocker PR (2011) Evolution of CD33-related siglecs: regulating host immune functions and escaping pathogen exploitation? Immunology 132:18–26
Cao H et al (2008) SIGLEC16 encodes a DAP12-associated receptor expressed in macrophages that evolved from its inhibitory counterpart SIGLEC11 and has functional and non-functional alleles in humans. Eur J Immunol 38:2303–2315
Carlin AF et al (2007) Group B streptococcal capsular sialic acids interact with siglecs (immunoglobulin-like lectins) on human leukocytes. J Bacteriol 89:1231–1237
Carlin AF et al (2009a) Group B Streptococcus suppression of phagocyte functions by protein-mediated engagement of human Siglec-5. J Exp Med 206:1691–1699
Carlin AF et al (2009b) Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. Blood 113:3333–3336
Chatterjee P et al (2014) Regulation of the anti-inflammatory cytokines interleukin-4 and interleukin-10 during pregnancy. Front Immunol 5:253
Chou HH et al (2002) Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. Proc Natl Acad Sci U S A 99:11736–11741
Collins BE et al (1997) Binding specificities of the sialoadhesin family of I-type lectins. Sialic acid linkage and substructure requirements for binding of myelin-associated glycoprotein, Schwann cell myelin protein, and sialoadhesin. J Biol Chem 272:16889–16895
Crocker PR, Varki A (2001) Siglecs, sialic acids and innate immunity. Trends Immunol 22:337–342
Crocker PR et al (1991) Purification and properties of sialoadhesin, a sialic acid-binding receptor of murine tissue macrophages. EMBO J 10:1661–1669
Crocker PR et al (1997) The potential role of sialoadhesin as a macrophage recognition molecule in health and disease. Glycoconj J 14:601–609
Crocker PR, Paulson JC, Varki A (2007) Siglecs and their roles in the immune system. Nat Rev Immunol 7:255–266
Freeman SD et al (1995) Characterization of CD33 as a new member of the sialoadhesin family of cellular interaction molecules. Blood 85:2005–2012

Gearing M et al (1994) Neuropathology and apolipoprotein E profile of aged chimpanzees: implications for Alzheimer disease. Proc Natl Acad Sci U S A 91:9382–9386
Griciuc A et al (2013) Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. Neuron 78:631–643
Hamerman JA, Lanier LL (2006) Inhibition of immune responses by ITAM-bearing receptors. Sci STKE 2006:re1
Hartnell A et al (2001) Characterization of human sialoadhesin, a sialic acid binding receptor expressed by resident and inflammatory macrophage populations. Blood 97:288–296
Hayakawa T et al (2005) A human-specific gene in microglia. Science 309:1693
Hernandez-Caselles T et al (2006) A study of CD33 (SIGLEC-3) antigen expression and function on activated human T and NK cells: two isoforms of CD33 are generated by alternative splicing. J Leukoc Biol 79:46–58
Ishida-Kitagawa N et al (2012) Siglec-15 protein regulates formation of functional osteoclasts in concert with DNAX-activating protein of 12 kDa (DAP12). J Biol Chem 287:17493–17502
Jandus C et al (2014) Interactions between Siglec-7/9 receptors and ligands influence NK cell-dependent tumor immunosurveillance. J Clin Invest 124(4):1810–1820
Kameda Y et al (2013) Siglec-15 regulates osteoclast differentiation by modulating RANKL-induced phosphatidylinositol 3-kinase/Akt and Erk pathways in association with signaling adaptor DAP12. J Bone Miner Res 28(12):2463–2475
Lanier LL (2009) DAP10- and DAP12-associated receptors in innate immunity. Immunol Rev 227:150–160
Linnartz-Gerlach B, Kopatz J, Neumann H (2014) Siglec functions of microglia. Glycobiology 24(9):794–799
Lock K et al (2004) Expression of CD33-related siglecs on human mononuclear phagocytes, monocyte-derived dendritic cells and plasmacytoid dendritic cells. Immunobiology 209:199–207
Malik M et al (2013) CD33 Alzheimer's risk-altering polymorphism, CD33 expression, and exon 2 splicing. J Neurosci 33:13320–13325
Mitra N et al (2011) SIGLEC12, a human-specific segregating (pseudo)gene, encodes a signaling molecule expressed in prostate carcinomas. J Biol Chem 286:23003–23011
Nguyen DH et al (2006) Loss of Siglec expression on T lymphocytes during human evolution. Proc Natl Acad Sci U S A 103:7765–7770
Padler-Karavani V et al (2014) Rapid evolution of binding specificities and expression patterns of inhibitory CD33-related Siglecs in primates. FASEB J 28:1280–1293
Patel N et al (1999) OB-BP1/Siglec-6—a leptin- and sialic acid-binding protein of the immunoglobulin superfamily. J Biol Chem 274:22729–22738
Perez SE et al (2013) Alzheimer's disease pathology in the neocortex and hippocampus of the western lowland gorilla (Gorilla gorilla gorilla). J Comp Neurol 521:4318–4338
Perez-Oliva AB et al (2011) Epitope mapping, expression and post-translational modifications of two isoforms of CD33 (CD33M and CD33m) on lymphoid and myeloid human cells. Glycobiology 21:757–770
Pillai S et al (2012) Siglecs and immune regulation. Annu Rev Immunol 30:357–392
Raj T et al (2014) CD33: increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. Hum Mol Genet 23:2729–2736
Ravetch JV, Lanier LL (2000) Immune inhibitory receptors. Science 290:84–89
Rempel H et al (2008) Sialoadhesin expressed on IFN-induced monocytes binds HIV-1 and enhances infectivity. PLoS One 3:e1967
Rumer KK et al (2013) Siglec-6 expression is increased in placentas from pregnancies complicated by preterm preeclampsia. Reprod Sci 20:646–653
Rutishauser U (2008) Polysialic acid in the plasticity of the developing and adult vertebrate nervous system. Nat Rev Neurosci 9:26–35
Sonnenburg JL, Altheide TK, Varki A (2004) A uniquely human consequence of domain-specific functional adaptation in a sialic acid-binding receptor. Glycobiology 14:339–346
Soto PC et al (2012) Cell-intrinsic mechanism involving Siglec-5 associated with divergent outcomes of HIV-1 infection in human and chimpanzee CD4 T cells. J Mol Med (Berl) 91:261–270

Takamiya R et al (2013) The interaction between Siglec-15 and tumor-associated sialyl-Tn antigen enhances TGF-beta secretion from monocytes/macrophages through the DAP12-Syk pathway. Glycobiology 23:178–187

Tourdot BE et al (2013) Immunoreceptor tyrosine-based inhibitory motif (ITIM)-mediated inhibitory signaling is regulated by sequential phosphorylation mediated by distinct nonreceptor tyrosine kinases: a case study involving PECAM-1. Biochemistry 52:2597–2608

Varki A (2010) Colloquium paper: uniquely human evolution of sialic acid genetics and biology. Proc Natl Acad Sci U S A 107(Suppl 2):8939–8946

Varki A (2011) Since there are PAMPs and DAMPs, there must be SAMPs? Glycan "self-associated molecular patterns" dampen innate immunity, but pathogens can mimic them. Glycobiology 21:1121–1124

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

Varki NM et al (2011) Biomedical differences between human and nonhuman hominids: potential roles for uniquely human aspects of sialic acid biology. Annu Rev Pathol 6:365–393

Wang Y, Neumann H (2010) Alleviation of neurotoxicity by microglial human Siglec-11. J Neurosci 30:3482–3488

Wang X et al (2011) Expression of Siglec-11 by human and chimpanzee ovarian stromal cells, with uniquely human ligands: implications for human ovarian physiology and pathology. Glycobiology 21:1038–1048

Wang X et al (2012a) Evolution of Siglec-11 and Siglec-16 genes in hominins. Mol Biol Evol 29:2073–2086

Wang X et al (2012b) Specific inactivation of two immunomodulatory SIGLEC genes during human evolution. Proc Natl Acad Sci U S A 109:9935–9940

Winn VD et al (2009) Severe preeclampsia-related changes in gene expression at the maternal-fetal interface include sialic acid-binding immunoglobulin-like lectin-6 and pappalysin-2. Endocrinology 150:452–462

Yamanaka M et al (2009) Deletion polymorphism of SIGLEC14 and its functional implications. Glycobiology 19:841–846

Yngvadottir B et al (2009) A genome-wide survey of the prevalence and evolutionary forces acting on human nonsense SNPs. Am J Hum Genet 84:224–234

Yu Z et al (2001) 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 276: 23816–23824

Zhang JQ et al (2000) Siglec-9, a novel sialic acid binding member of the immunoglobulin superfamily expressed broadly on human blood leukocytes. J Biol Chem 275:22121–22126

Zou Z et al (2011) Siglecs facilitate HIV-1 infection of macrophages through adhesion with viral sialic acids. PLoS One 6:e24559