Associations of genetic polymorphisms of Siglecs with human diseases

Takashi Angata

Institute of Biological Chemistry, Academia Sinica, 128 Academia Road, Section 2, Nankang District, Taipei 11529, Taiwan

Received on March 24, 2014; revised on April 27, 2014; accepted on May 12, 2014

Genetic polymorphism studies in humans provide unique opportunities to understand human biology and the mechanisms of diseases. Correlations between polymorphisms in the genes encoding human Siglecs and various diseases have been reported. Leading examples, such as the CD33 polymorphism associated with late-onset Alzheimer’s disease, are well supported by genetic replication and mechanistic studies, while some others (such as SIGLEC8 polymorphism associated with bronchial asthma and SIGLEC14 polymorphism associated with exacerbation of chronic obstructive pulmonary disease) may benefit reinforcement by independent genetic replication or mechanistic studies. In a few cases, such as MAG polymorphism associated with psychological disorder and CD22 polymorphism associated with autoimmune disease, the phenotype associated with a genetic polymorphism of a Siglec gene and that of an enzyme gene involved in the biosynthesis of Siglec ligand show some overlap, providing indirect support for the observed genotype–phenotype association. Although studies using engineered mutant mice have provided invaluable insights into the biological functions and mechanisms of diseases, it is not always possible to develop appropriate mouse model to replicate human situations because of significant species-to-species differences, which can be a major obstacle in understanding the biology of some of human CD33/Siglec-3-related Siglecs. Further studies in genetic polymorphisms of human Siglecs, combined with appropriate functional studies, may reveal unexpected biological roles of human Siglecs, and identify possible targets for prevention and/or treatment of certain diseases.

Keywords: association study / disease / polymorphism / Siglec

Introduction

Siglecs are a family of sialic acid-binding lectins in vertebrates that are expressed primarily on leukocytes (Varki and Angata 2006; Crocker et al. 2007; Pillai et al. 2012). Siglecs are thought to recognize sialic acids as a part of “self”, or self-associated molecular patterns (SAMPs), to discriminate the organism’s own cells from those of foreign origins (Varki 2011). Proteins expressed in the immune system tend to evolve more rapidly than other proteins (The Chimpanzee Sequencing and Analysis Consortium 2005; Gibbs et al. 2007), because these proteins have higher chance of encountering pathogens and thus are under selective pressure to avoid exploitation by such pathogens. Indeed, a subset of Siglecs known as CD33/Siglec-3-related Siglecs, which show high sequence similarity to CD33/Siglec-3 and are encoded in a gene cluster, show significant variations even between closely related species (Angata et al. 2004; The Chimpanzee Sequencing and Analysis Consortium 2005). This rapid evolution is likely driven both by pathogens that exploit Siglecs and by endogenous ligands (whose evolution is driven by other pathogens exploiting sialic acids; Varki and Angata 2006; Padler-Karavani et al. 2014). This rapid evolution can occasionally impose a problem for functional studies of human Siglecs, as many human CD33/Siglec-3-related Siglecs do not have clear mouse orthologs. On the other hand, rapid diversification of Siglecs also provides opportunities to correlate the polymorphisms of human Siglecs with human phenotypes, including disease susceptibilities.

In this review, I attempt to summarize the current knowledge regarding the associations between genetic polymorphisms of human Siglecs and diseases (Table 1), and correlations with studies of genetic polymorphisms of enzymes involved in Siglec ligand biosynthesis. Another table, summarizing the mutations/polymorphisms in human sialytransferase genes associated with human diseases/phenotypes, is provided as an online supplement (Supplementary data, Table SI). Nomenclature guidelines regarding gene symbols are followed in this review, namely human genes are denoted by symbols with all-Italic capital letters, without hyphen preceding numerical (e.g., CD33, MAG, SIGLEC8, ST3GAL3, etc.); mouse genes are denoted by symbols with all-Italic letters, capitalized at the initial character only, without hyphen preceding numerical (e.g., Cd33, Mag, SiglecF, St3gal3, etc.). Siglec protein nomenclatures are based on the previous recommendations (CD33/Siglec-3, MAG/Siglec-4, Siglec-F, etc.) (Crocker et al. 1998; Angata, Hingorani et al. 2001).

Genetic polymorphisms and genotype-phenotype association studies

Genetic polymorphisms include single-nucleotide polymorphisms (SNPs), nucleotide insertion, deletion, inversion and
chromosomal translocations. Two main approaches have been used to study correlations between genetic polymorphisms and human diseases/traits (i.e., phenotypes): the candidate gene-based study (hypothesis-driven approach) and the genome-wide association study (GWAS; a hypothesis-free approach). The chance of discovering genotype–phenotype correlations depends on several factors, such as the frequency of minor alleles, effect size of the polymorphism, size of the sample (population) under study and the degree of dependence of disease/trait on genetic factors, among others (Evans and Purcell 2012). One advantage of the candidate gene-based study is that it tends to have more power to detect genotype–phenotype associations, because investigators likely consider the biological effects of each polymorphic site (such as SNP) to test in their study and avoid the sites that are unlikely to have biological consequence, limiting the number of hypotheses (=polymorphic sites) tested in parallel—which dictates the threshold of statistical significance. Candidate gene-based study also tends to be easier to combine with mechanistic study, as the investigators already have some hypothesis regarding disease mechanism in mind. However, by definition, it cannot reveal any correlation between an unsuspected genetic polymorphism and the phenotype under study. Also, generally poor reproducibility is a major challenge of this approach (Lewis and Knight 2012). The GWAS is a recent trend and has been applied to many major diseases with complex etiologies. The GWAS has revealed many unsuspected associations between human genetic polymorphisms and diseases. However, its high cost and requirement for large sample size are major obstacles, making GWAS beyond the reach of most individual investigators.

Although it is commonsense, it should be emphasized that the polymorphism that shows the strongest association with the phenotype does not necessarily cause the phenotype (association does not necessarily mean causation). It is impossible to cover all human polymorphisms in a study (even with the microarray platforms being used in GWAS, which typically interrogate 0.5–1 × 10⁶ SNPs simultaneously); therefore, the polymorphism showing the strongest association with the phenotype in a study may do so just because it lies in the proximity of the true causal polymorphism (i.e., in linkage disequilibrium) and the latter was not investigated in the study.

Regardless of the initial approach, once a result suggestive of a correlation between a genotype and a phenotype is obtained, replication in an independent sample set and/or a mechanistic study is generally required to reinforce support for the association. Strictly speaking, some of the studies quoted in this review may not fulfill these criteria (and thus need careful evaluation), but are included to encourage further studies.

CD33/Siglec-3 and Alzheimer’s disease

A series of studies on the association between polymorphisms in the CD33 gene with late-onset Alzheimer’s disease may be the most successful example of research efforts on the association of Siglec polymorphism and human disease.

Human CD33/Siglec-3 protein is expressed on myeloid cells, including brain microglia. Its glycan binding preference appears to be promiscuous (Freeman et al. 1995; Brinkman-Van der Linden and Varki 2000). A GWAS with a large number of

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphism/marker</th>
<th>Disease</th>
<th>Methods</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD22</td>
<td>rs4805119, rs10406539, rs10413500, rs10413526, rs4805120, etc.</td>
<td>Infantile B-cell leukemia</td>
<td>Candidate gene-based association study</td>
<td>Uckun et al. (2010), Ma et al. (2012)</td>
</tr>
<tr>
<td>CD22</td>
<td>rs3865444 (promoter SNP)</td>
<td>Alzheimer’s disease</td>
<td>GWAS</td>
<td>Naj et al. (2011), Hollingworth et al. (2011), Malik et al. (2013), Raj et al. (2014)</td>
</tr>
<tr>
<td>CD22</td>
<td>rs720309 (intronic SNP)</td>
<td>Schizophrenia</td>
<td>Candidate gene-based association study</td>
<td>Wan et al. (2005), Yang et al. (2005), Jitoku et al. (2011)</td>
</tr>
<tr>
<td>CD33</td>
<td>rs12459419 (coding SNP, non-synonymous; exon skipping)</td>
<td>Cardio-vascular outcomes</td>
<td>Candidate gene-based association study</td>
<td>Naj et al. (2011), Hollingworth et al. (2011), Malik et al. (2013), Raj et al. (2014)</td>
</tr>
<tr>
<td>CD33</td>
<td>rs3865444 (promoter SNP)</td>
<td>Alzheimer’s disease</td>
<td>Candidate gene-based association study</td>
<td>Naj et al. (2011), Hollingworth et al. (2011), Malik et al. (2013), Raj et al. (2014)</td>
</tr>
<tr>
<td>CD33</td>
<td>rs10409962 (coding SNP, non-synonymous)</td>
<td>Bronchial asthma</td>
<td>Candidate gene-based association study</td>
<td>Wan et al. (2005), Yang et al. (2005), Jitoku et al. (2011)</td>
</tr>
<tr>
<td>MAG</td>
<td>rs7296917 (intronic SNP)</td>
<td>COPD exacerbation</td>
<td>Candidate gene-based association study</td>
<td>Wan et al. (2005), Yang et al. (2005), Jitoku et al. (2011)</td>
</tr>
<tr>
<td>SIGLEC12</td>
<td>rs1698747 (coding SNP)</td>
<td>Preterm delivery</td>
<td>Candidate gene-based association study</td>
<td>Wan et al. (2005), Yang et al. (2005), Jitoku et al. (2011)</td>
</tr>
<tr>
<td>SIGLEC14</td>
<td>rs10409962 (coding SNP)</td>
<td>COPD exacerbation</td>
<td>Candidate gene-based association study</td>
<td>Naj et al. (2011), Hollingworth et al. (2011), Malik et al. (2013), Raj et al. (2014)</td>
</tr>
<tr>
<td>SIGLEC14</td>
<td>rs12459419 (coding SNP, non-synonymous; exon skipping)</td>
<td>Alzheimer’s disease</td>
<td>Candidate gene-based association study</td>
<td>Naj et al. (2011), Hollingworth et al. (2011), Malik et al. (2013), Raj et al. (2014)</td>
</tr>
<tr>
<td>SIGLEC14</td>
<td>rs36498 (promoter SNP)</td>
<td>Bronchial asthma</td>
<td>Candidate gene-based association study</td>
<td>Wan et al. (2005), Yang et al. (2005), Jitoku et al. (2011)</td>
</tr>
<tr>
<td>SIGLEC14</td>
<td>rs1698747 (coding SNP)</td>
<td>COPD exacerbation</td>
<td>Candidate gene-based association study</td>
<td>Naj et al. (2011), Hollingworth et al. (2011), Malik et al. (2013), Raj et al. (2014)</td>
</tr>
</tbody>
</table>
late-onset Alzheimer’s disease patients and control population (exceeding 10,000 samples each) discovered an association between SNPs in CD33 and the disease (Hollingsworth et al. 2011; Naj et al. 2011). The SNP showing the strongest association was rs3865444, located in the upstream region of CD33, of which the minor allele (A allele) is protective. Subsequent studies found that the A allele is associated with reduced levels of CD33/Siglec-3 protein expression (Bradshaw et al. 2013; Griciuc et al. 2013). A mechanistic study using cultured microglial cells and a Cd33-null mouse discovered that CD33/Siglec-3 protein inhibits uptake of amyloid beta protein fragments (Aβ42 and Aβ40) by microglia (Griciuc et al. 2013). Amyloid beta protein fragments form insoluble amyloid plaques, a major culprit in Alzheimer’s disease. Thus, reduced expression of CD33/Siglec-3 allows more efficient phagocytic clearance of pathogenic amyloid beta protein fragments by brain microglia, and thus protects people from Alzheimer’s disease. Another study using human monocytes and brain samples corroborated this finding (Bradshaw et al. 2013).

It was assumed that the SNP rs3865444, located in the promoter region of CD33 gene, influenced the transcription of CD33; however, this was not the case (Griciuc et al. 2013). Nevertheless, the SNP rs3865444 was associated with the CD33/Siglec-3 protein expression level. This conundrum was solved by other investigations. The polymorphism that is most likely causative of the reduced expression of CD33/Siglec-3 is another SNP (rs12459419) in exon 2, which is in linkage disequilibrium with the SNP rs3865444 (Malik et al. 2013; Raj et al. 2014). The T allele of SNP rs12459419 (which is co-inherited with the A allele of SNP rs3865444) increases the chance of exon 2 skipping, leading to the expression of a CD33/Siglec-3 protein lacking first Ig-like domain (Malik et al. 2013; Raj et al. 2014) (Figure 1A). Most anti-CD33/Siglec-3 monoclonal antibodies require the first Ig-like domain of CD33/Siglec-3 for recognition (Perez-Oliva et al. 2011); therefore, CD33/Siglec-3 without the first Ig-like domain is undetectable by most antibodies. In addition, the first Ig-like domain is essential for CD33/Siglec-3’s function to suppress amyloid fragment uptake (Griciuc et al. 2013). Taken together, reduced expression of CD33/Siglec-3 isofrom with first Ig-like domain (which inhibits phagocytic activity of myeloid cells) leads to increased clearance of amyloid beta fragments, protecting the brain from Alzheimer’s disease. Although this association between a CD33 polymorphism and late-onset Alzheimer’s disease was challenged in a study involving a larger number of case and controls (Lambert et al. 2013), the mechanistic studies mentioned above and independent genetic association studies replicating the association of SNP rs3865444 with Alzheimer’s disease (Deng et al. 2012; Omoumi et al. 2014) provide support for the initial finding.

The orthology between human and mouse CD33/Siglec-3 is somewhat ambiguous (Brinkman-Van der Linden et al. 2003). For example, human CD33/Siglec-3 has an immunoreceptor tyrosine-based inhibitory motif in the cytosolic tail and recruits protein tyrosine phosphatase SHP-1 (Taylor et al. 1999), while mouse CD33/Siglec-3 lacks this motif and is unlikely to associate with SHP-1. Nevertheless, a membrane-distal tyrosine residue in the cytoplasmic domain is conserved between human and mouse (Brinkman-Van der Linden et al. 2003), which likely maintains the regulatory function related to endocytosis.

It has been proposed that gangliosides (GM1, etc.) facilitate amyloid plaque formation and are involved in the development of Alzheimer’s disease (Yanagisawa 2007; Ariga et al. 2008; Matsuuzaki et al. 2010). Gangliosides may also serve as ligands for CD33/Siglec-3. Given the line of evidence described in a preceding paragraph, suggesting that CD33/Siglec-3 ligand recognition attenuates amyloid clearance, it may be possible to hypothesize that inhibitory signaling elicited by the engagement of CD33/Siglec-3 by gangliosides in amyloid plaque suppresses endocytic activity of microglia, leading to further accumulation of amyloids (Salminen and Kaarniranta 2009). Further studies would be needed to answer whether the interaction between gangliosides in amyloid plaque and CD33/Siglec-3 (or any other Siglec, such as Siglec-11) on microglia plays any role in Alzheimer’s disease development.

Of note, SNP rs12459419 in CD33 gene was also found to be weakly associated with the treatment outcomes of Gemtuzumab Ozogamicin (also known as Mylotarg; an anti-CD33 antibody–toxin conjugate), which was used to treat acute myeloid leukemia (Mortland et al. 2013).

CD22/Siglec-2 and autoimmunity

The CD22/Siglec-2 protein is expressed on a majority of mature B cells (Tedder et al. 1997). An association of a polymorphism of the mouse Cd22 gene with propensity toward autoimmunity has been reported (Mary et al. 2000). A line of Cd22-null mice was also reported to develop autoimmunity (O’Keefe et al. 1996; 1999), supporting the possible connection between deficiency of CD22/Siglec-2 and autoimmunity.

A study addressing the possible association of human CD22 polymorphism and autoimmune disease has been published (Hitomi et al. 2007). The authors found that a synonymous substitution at nucleotide 2304 (c.2304 C > A; rs34826052) is associated with systemic sclerosis (SSc), an autoimmune disease characterized by tissue fibrosis of the skin and visceral organs. AA homozygotes were found only in patients with limited cutaneous SSc, a subtype of the disease with lower prevalence of internal organ involvement. The authors demonstrated that this polymorphism is correlated with reduced expression of CD22/Siglec-2 protein on B cells (Hitomi et al. 2007) (Figure 1B). Although this observation was not replicated in an independent study (Dawidowicz et al. 2011), it may not negate the initial finding, considering that a population of different ancestry (Europeans) was studied in the follow-up, and that the effect size of this polymorphism is small (~17% reduction in CD22/Siglec-2 expression level on B cells).

CD22/Siglec-2 shows a strong preference toward the Sia2-6Gal terminal structure (Powell et al. 1993; Sgroi et al. 1993). Genetic polymorphisms in ST6GAL1 gene, which is responsible for the biosynthesis of CD22/Siglec-2 ligands, were found to be associated with type 2 diabetes in South Asians (Kooner et al. 2011). Although type 2 diabetes is not generally considered an autoimmune disorder, recent studies suggest that it may involve an autoimmune component (Winer et al. 2011; Velloso et al. 2013).

Modification of the glycerol-like side chain of sialic acids with acetyl ester disrupts the interaction between CD22/Siglec-2 and...
sialylated oligosaccharides (Sjoberg et al. 1994). Using knock-out mice, sialic acid esterase was shown to modify B-cell receptor (BCR) signaling by regulating the interaction between BCR and CD22/Siglec-2 (Cariappa et al. 2009). Rare reduction-of-function mutations in the sialic acid esterase (SIAE) gene are more frequently found in patients with common autoimmune diseases, including rheumatoid arthritis and type 1 diabetes (Surolia et al. 2010; Chellappa et al. 2013).

These findings, related to genetic polymorphisms in the enzyme genes (ST6GAL1 and SIAE) regulating CD22/Siglec-2 ligand biosynthesis, may serve as separate lines of evidence that support involvement of CD22/Siglec-2 in autoimmune disorders.

**CD22/Siglec-2 and infantile B-cell leukemia**

A recent study (by a candidate gene-based approach) revealed that homozygous mutations in intron 12 of CD22 are associated with infantile B-precursor leukemia, an aggressive form of childhood acute lymphoblastic leukemia (Uckun et al. 2010). The same authors demonstrated that the same (or similar) mutations are also found in therapy-refractory clones of pediatric B-lineage acute lymphoblastic leukemia (Ma et al. 2012). These mutations cause skipping of exon 12 during splicing and concomitant frame shift in the CD22 transcript, with the end product (CD22/Siglec-2 protein) lacking most of the cytoplasmic tail (Figure. 1C). As SHP-1 and other signal transduction molecules interact with the cytoplasmic tail of CD22/Siglec-2 protein (Uckun et al. 2010). A further mechanistic study revealed that the transgenic expression of human CD22 mRNA lacking exon 12, under the control of B cell-specific transcriptional enhancer Eμ, induces B-cell hyperplasia in mice (Uckun et al. 2010).
It is not clear whether the mutations are found in the germline or only in the B-cell clones that are undergoing uncontrolled proliferation. Therefore, it is possible that the mutations are either inherited from parents or occurred de novo in a B-cell lineage of patients. The phenotypes observed in patients appears to be more severe than those found in Cd22-null mice, in which hyperproliferation of B cells have not been observed (Cyster and Goodnow 1997), implying the presence of other genetic factor(s) in these patients and/or the possibility that the mutant Cd22/Siglec-2 plays additional role in facilitating hyperproliferation of B cells. Although these findings may not represent a typical case of genetic polymorphism-disease association, these deserve attention from structural and functional points of view.

**MAG/Siglec-4 and neurological disorders**

Myelin-associated glycoprotein (MAG)/Siglec-4 (encoded by the MAG gene) is expressed on Schwann cells in the peripheral nerve system and on oligodendroglia in the central nervous system (Schachner and Bartsch 2000). It binds to the Neu5Acα2-3Galβ1-3GalNAc terminal structure (Kelm et al. 1994; Collins, Yang et al. 1997), which is found in glycolipids and O-glycans. The presence of sialic acid at the C6 position of GalNAc strongly enhances affinity (Collins, Kiso et al. 1997; Blixt et al. 2003).

Association of genetic polymorphisms in MAG with schizophrenia in a Chinese Han population was reported (Wan et al. 2005; Yang et al. 2005). While this association was refuted in an independent study in Europeans (Voineskos et al. 2008), another study in Japan suggested that SNPs in MAG may indeed be weakly associated with schizophrenia (Jitoku et al. 2011). Although the association of schizophrenia with the SNP (rs720309) in MAG implicated in the initial reports was not replicated, the most significantly associated SNP (rs7249617) in the Japanese study was located in the same intron and in the linkage block with the former. These studies did not offer any functional explanations regarding the association of these SNPs with transcript or protein expression level. As the association of schizophrenia and reduced oligodendroglial gene expression, with transcript or protein expression level. As the association of these SNPs and bronchial asthma has not been elucidated, it is possible that these SNPs may be associated with altered levels of MAG/Siglec-4.

Similarities in the phenotypes observed in Mag-deficient and in B4galnt1 (GM2/GD2 synthase gene)-deficient mice have been reported (Sheikh et al. 1999; Vyas et al. 2002; Sun et al. 2004; Pan et al. 2005). Therefore, it is possible that deficiency or polymorphism in the human enzyme involved in the biosynthesis of gangliosides GD1a and GT1b (the preferred MAG/Siglec-4 ligand abundantly expressed in adult mammalian brain) may result in similar phenotype as the MAG polymorphisms (see Supplementary data, Fig. S1 for the biosynthetic pathway of gangliosides and the enzymes involved in the pathway). In this context, similarities in the phenotypes associated with human polymorphisms in MAG (schizophrenia) and ST3GAL1 (bipolar disorder and negative symptoms in schizophrenia; Perlis et al. 2008; Zandi et al. 2008; Zhang et al. 2010; Xu et al. 2013) appears suggestive. However, the biosynthesis of the Neu5Aca2-3Galβ1-3GalNAc terminal structure in GD1a and GT1b in the mouse brain was found to be largely dependent on ST3Gal-II, and to a lesser extent, on ST3Gal-III (encoded by the ST3gal2 and ST3gal3 genes, respectively), and not on ST3Gal-I (Sturgill et al. 2012). Although the relative contributions of various ST3Gal enzymes to the biosynthesis of human brain gangliosides are unknown, it is likely to be similar to those in mouse brain. While loss-of-function mutations in ST3GAL3 were reported to cause severe neurological disorders (mental retardation and epileptic encephalopathy) (Najmabadi et al. 2007; Hu et al. 2011; Edvardson et al. 2013), which may be partly explained by the loss of interaction with MAG/Siglec-4, no loss-of-function mutation or polymorphism in human ST3GAL2 has been associated with any disease so far.

**Siglec-8 and bronchial asthma**

Human Siglec-8 is expressed on eosinophils and mast cells (Floyd et al. 2000; Kikly et al. 2000). Antibody-mediated Siglec-8 ligation induces eosinophil apoptosis (Nutku et al. 2003). Mouse Siglec-F is expressed on eosinophils and is considered to carry out a similar function to human Siglec-8 (Zhang et al. 2004; Tateno et al. 2005; Bochner 2009). Siglec-F expression on eosinophils is upregulated by allergic lung inflammation, and genetic disruption of Siglec-F enhances allergen-induced eosinophilia in mouse (Zhang et al. 2007). These studies imply that Siglec-8 may be involved in the pathogenesis of allergic diseases like bronchial asthma, and polymorphisms in SIGLEC8 may influence susceptibility of individuals toward these diseases.

Indeed, an international research group found associations of genetic polymorphisms in SIGLEC8 and bronchial asthma (Gao et al. 2010). The most significantly associated SNPs in an African-American population, rs36498 and rs10409962, lie in the upstream (promoter) region and exon 2 of SIGLEC8, respectively. These findings were essentially replicated in the study using different populations (either Brazilians or Japanese). Another SNP rs6509541, in the 3’ untranslated region, was associated with serum IgE level in African Americans and Brazilians. The authors expressed a Siglec-8 cDNA containing allelic variations at rs10409962 (translated to Ser170 or Pro170) in HEK293 cells and compared the expression levels and ligand binding properties, but found no difference. Although the mechanism connecting these SNPs and bronchial asthma has not been elucidated, it is possible that these SNPs may be associated with altered levels of Siglec-8 expression on eosinophils.

Assuming that the ligand for Siglec-8 in human lung is synthesized by ST3Gal-III, as is the case for Siglec-F ligand synthesis in mouse lung (Guo et al. 2011; Suzukawa et al. 2013; Kiwamoto et al. 2014), polymorphisms in human ST3GAL3 may be associated with bronchial asthma. However, no such association has been reported so far. As mentioned above, loss-of-function mutations in ST3GAL3 cause severe neurological disorders (Najmabadi et al. 2007; Hu et al. 2011; Edvardson et al. 2013).

**Siglec-XII and cardiovascular therapeutic outcomes**

Siglec-XII (encoded by SIGLEC12) is expressed on monocytes/macrophages and luminal epithelia, including that of the prostate (Angata, Varki et al. 2001; Mitra et al. 2011). Although...
it does not function as a lectin, because it lacks the arginine residue essential for the recognition of sialic acid (thus it is numbered in Roman numerals, to indicate it does not have lectin function), it maintains a signaling function (Yu et al. 2001). Two null mutations of SIGLEC12, which appear to have occurred independently, have been discovered (Mitra et al. 2011).

In an effort to identify predictive markers of cardiovascular outcomes in patients with hypertension on antihypertensive therapy, McDonough et al. found that SNP rs16982743 in SIGLEC12 is associated with adverse outcomes, which depend on the therapeutic strategy (β-blocker or calcium channel blocker) (McDonough et al. 2013). SNP rs16982743 corresponds to one of the null mutations mentioned above (Figure 1D). Combination of the allele that allows expression of Siglec-XII and therapeutic use of calcium channel blocker raised the risk of death, myocardial infarction or stroke. The mechanism by which the presence of Siglec-XII leads to cardiovascular outcomes in a therapy-dependent manner is not clear. Some Siglecs regulate calcium signals in immune cells; therefore, it appears possible that Siglec-XII’s regulation of the calcium signal interacts with a calcium channel blocker. Further study with regard to the expression pattern and function of Siglec-XII would be necessary to solve this question.

This same SNP in SIGLEC12 was also studied in the context of prostate cancer susceptibility and outcome, but no clear correlation between genotype and phenotype was observed (Mitra et al. 2011). Another study suggested that the null allele of this SNP appears to be under selective pressure and was selected for during human evolution, implying that the absence of Siglec-XII may have been associated with favorable trait(s) (Yngvadottir et al. 2009). Another “null” mutation of SIGLEC12, rs371016684 (c.193delCinsAG, inducing frame-shift and premature termination; overlapping with rs61742108), which is very close to, but apparently not co-inherited with rs16982743 mentioned above, was not tested in these studies.

Siglec-14 and COPD exacerbation

Siglec-14 is an activating-type Siglec that associates with the DAP12 adapter protein and is expressed on monocytes and granulocytes. The gene encoding Siglec-14 (SIGLEC14) is aligned in tandem with that for Siglec-5 (SIGLEC5). Siglec-14 is absent in many individuals, because of homozygosity of an allele that lacks the major part of the SIGLEC14-coding segment (Yamanaka et al. 2009). This SIGLEC14-null allele was generated via fusion of the SIGLEC14 and SIGLEC5 genes (the protein product of this SIGLEC14/5 fusion gene is identical to Siglec-5 in its amino acid sequence) (Figure 1E).

We reported that the frequency of exacerbations among the patients of COPD is associated with the SIGLEC14-null polymorphism (Angata et al. 2013). COPD is a chronic lung disease caused by chronic exposure to noxious particles or gases (typically cigarette smoke) and is characterized by emphysema (destruction of alveolar walls) and chronic bronchitis (inflammation of airways). Exacerbation is an acute worsening of the symptoms associated with COPD (such as shortness of breath, cough and sputum) that is beyond normal day-to-day variations. Exacerbation of COPD is typically caused by airway infection by bacteria (e.g., Haemophilus influenza or Streptococcus pneumoniae) or viruses (e.g., influenza virus or rhinovirus) (Sethi and Murphy 2008). We found that the patients who do not express Siglec-14 (homozygous SIGLEC14-null) tend to suffer less episodes of COPD exacerbation on average. We also showed that a myeloid cell line expressing Siglec-5 (mimicking monocytes from homozygous SIGLEC14-null individuals) produce less pro-inflammatory cytokines than those expressing Siglec-14 (mimicking monocytes from homozygous wild-type individuals) when exposed to H. influenzae. Based on these data, we proposed a model in which Siglec-14 (along with other receptors of pathogen-associated molecular patterns, such as Toll-like receptors) trigger a pro-inflammatory signal that leads to COPD exacerbation, and the absence of Siglec-14 attenuates this cascade, and thus protects the patients from developing the excessive inflammation that leads to COPD exacerbation (Angata et al. 2013).

In a separate study, a possible correlation between the SIGLEC14 genotype and susceptibility to preterm birth in the presence of group B Streptococcus (GBS) infection was revealed (Ali et al. 2014). GBS is a major pathogen infecting newborn and is also associated with preterm delivery. GBS serotype Ia binds to Siglec-5 and Siglec-14 via a β-protein in a sialic acid-independent manner (Carlin et al. 2009; Ali et al. 2014). In the presence of GBS infection, homozygous SIGLEC14-null babies were more likely to be delivered preterm than babies who carry at least one functional SIGLEC14 allele (Ali et al. 2014). In this case, the presence of the functional Siglec-14 protein is protective.

Considering that a similar paired configuration between SIGLEC11 and SIGLEC16 has been reported (Angata et al. 2002), and SIGLE16 has a polymorphic null allele in humans (Cao et al. 2008), it is tempting to speculate that this null allele may also be associated with some human conditions (e.g., susceptibility to some diseases).

Allo-antibodies against Siglecs

The presence of allo-antibodies against Siglec-8 (von Gunten et al. 2007), Siglec-9 (von Gunten et al. 2006) and Siglec-14 (Yasui et al. 2011) in normal human blood or blood products have been described, although no direct connection with genetic polymorphisms in the genes encoding these Siglecs has been established. In the case of the Siglec-14 allo-antibodies, their presence in blood products was found to be associated with blood transfusion-related side effects, including transfusion-related acute lung injury (Yasui et al. 2011). These allo-antibodies could be generated when a pregnant mother is exposed to the antigen from her fetus, which carries an allele that is different from the mother’s.

Conclusion and perspectives

Previous studies have suggested that many Siglecs have been generated and lost during evolution. Therefore, it is not expected that polymorphisms (even loss-of-function) of human Siglecs would lead to Mendelian disorders (hereditary human diseases). Rather, weak “associations” between human Siglec polymorphisms and phenotypes are expected and are indeed observed. This is in contrast with sialyltransferases, which are relatively well conserved among vertebrates (Harduin-Lepers et al. 2008) and their
loss-of-function mutations sometimes result in Mendelian disorders (Simpson et al. 2004; Najmabadi et al. 2007; Hu et al. 2011; Edvardson et al. 2013; Fragaki et al. 2013; Boccuto et al. 2014; InanlooRahatloo et al. 2014). Nevertheless, investigations of possible correlations between genetic polymorphisms of human Siglec genes and phenotypes have yielded clues to understand biological functions of human Siglecs and have revealed possible targets for disease interventions.

Given that many Siglecs are involved in interactions with pathogenic microbes, studies connecting infectious disease susceptibility and genetic polymorphisms of Siglecs may reveal further functions of human Siglecs in immunological defense. Although simple correlation studies might be unwarranted in developed countries with reduced infectious disease burdens, some clinical conditions, such as opportunistic infections associated with poor outcomes of chronic inflammatory diseases, may provide opportunities for such studies. In a broader context, microbial infection and colonization constitute a major factor in gene–environment interactions, as does medication in pharmacogenetics/genomics studies (although in a less well-defined manner). Also, unexpected findings from GWASs may elicit a sudden surge of interest in a member of Siglecs and reveal its cryptic function, as was the case for the association between CD33 and Alzheimer’s disease. Thus, attention should be paid to the future development of association studies.

Supplementary Data
Supplementary data for this article are available online at http://glycob.oxfordjournals.org/.

Acknowledgment
Although I attempted to include as many reports on Siglec-related polymorphisms and correlations with diseases as possible in an unbiased manner, some important references may be inadvertently left out. If so, I apologize to the authors of these reports.

Conflict of interest statement
None declared.

Abbreviations
BCR, B-cell receptor; GBS, group B Streptococcus; GWAS, genome-wide association study; SAMPs, self-associated molecular patterns; SNPs, single-nucleotide polymorphisms; SSc, systemic sclerosis

References

Genetic polymorphisms of human Siglecs and disease


