Evolution of Human-Specific Alleles Protecting Cognitive Function of Grandmothers

Sudeshna Saha,[1] Naazneen Khan,[1] Troy Comi,[2] Andrea Verhagen,[1] Aniruddha Sasmal,[1] Sandra Diaz,[1] Hai Yu,[3] Xi Chen,[3] Joshua M. Akey,[2] Martin Frank,[4] Pascal Gagneux,*,[1] and Ajit Varki *,[1]

[1]Departments of Medicine, Pathology, Anthropology and Cellular and Molecular Medicine, Center for Academic Research and Training in Anthropogeny and Glycobiology Research and Training Center, University of California San Diego, San Diego, CA 92093, USA

[2]Department of Genetics, Princeton University, Princeton, NJ 08544, USA

[3]Department of Chemistry, University of California Davis, Davis, CA 95616, USA

[4]Biognos AB, Gothenburg SE-402 74, Sweden

***Corresponding authors**: E-mails: a1varki@ucsd.edu; pgagneux@ucsd.edu.

Associate editor: Evelyne Heyer

Abstract

The myelomonocytic receptor CD33 (Siglec-3) inhibits innate immune reactivity by extracellular V-set domain recognition of sialic acid (Sia)-containing "self-associated molecular patterns" (SAMPs). We earlier showed that V-set domain-deficient *CD33*-variant allele, protective against late-onset Alzheimer's Disease (LOAD), is derived and specific to the hominin lineage. We now report multiple hominin-specific CD33 V-set domain mutations. Due to hominin-specific, fixed loss-of-function mutation in the *CMAH* gene, humans lack *N*-glycolylneuraminic acid (Neu5Gc), the preferred Sia-ligand of ancestral CD33. Mutational analysis and molecular dynamics (MD)-simulations indicate that fixed change in amino acid 21 of hominin V-set domain and conformational changes related to His45 corrected for Neu5Gc-loss by switching to *N*-acetylneuraminic acid (Neu5Ac)-recognition. We show that human-specific pathogens *Neisseria gonorrhoeae* and Group B *Streptococcus* selectively bind human CD33 (huCD33) as part of immune-evasive molecular mimicry of host SAMPs and that this binding is significantly impacted by amino acid 21 modification. In addition to LOAD-protective *CD33* alleles, humans harbor derived, population-universal, cognition-protective variants at several other loci. Interestingly, 11 of 13 SNPs in these human genes (including *CD33*) are not shared by genomes of archaic hominins: Neanderthals and Denisovans. We present a plausible evolutionary scenario to compile, correlate, and comprehend existing knowledge about huCD33-evolution and suggest that grandmothering emerged in humans.

Key words: **CD33, pathogen, sialic acids, archaic genome, molecular dynamics simulation, phylogenetic analysis.**

Introduction

In keeping with the fundamental importance of reproduction for the process of biological evolution via natural selection, loss of fecundity generally coincides with the end of lifespan in almost all species studied to date. Humans and certain toothed whales like orcas are so far the only mammals known to manifest prolonged postreproductive lifespans under naturalistic conditions (Hawkes et al. 1998; Hawkes 2004, 2010; Johnstone and Cant 2010; Cant and Croft 2019; Khan, Kim, et al. 2020). One current explanation for such prolonged postreproductive survival is late life kin selection of grandmothers and other elderly caregivers of helpless young—apparently contrary to the concept of "antagonistic pleiotropy," which posits that natural selection does not operate in late life (Williams 1957; Byars and Voskarides 2020). In this regard, we previously observed several commonly occurring, derived gene variants that directly or indirectly impact postreproductive, late life cognition, and are present only in human genomes and not in those of any other "great apes" (Schwarz et al. 2016). This was considered as genomic evidence for the evolution of human postmenopausal longevity (Hawkes 2016). Here, we further explore the human-specific, derived alleles of genes that protect against late life cognitive decline, and ask when and how these emerged in hominins?

One such interesting example is a human-specific, derived allele of CD33 associated with direct or indirect protection against late-onset Alzheimer's Disease (LOAD) (Schwarz et al. 2016; Zhao 2019). In vertebrates, glycan-binding proteins of the immunoglobulin (Ig) superfamily called sialic acid (Sia)-binding Ig-like lectins (Siglecs) form a major component of the immune system (Läubli and Varki 2020). As the name indicates, Siglecs recognize

Sias on cell surface or secreted glycoproteins and glycolipids. Siglec-3, commonly known as CD33, is the eponymous member of the rapidly evolving subgroup of Siglecs called CD33-related Siglecs (or CD33rSiglecs) (Freeman et al. 1995; Siddiqui et al. 2017). In contrast, other Siglecs (Siglecs 1, 2, 4, and 15) show evolutionary conservation (Bornhöfft et al. 2018). CD33 is a type-I transmembrane protein with an amino terminal Ig-like V-set domain followed by one Ig-like C2-set domain proximal to the transmembrane region (Freeman et al. 1995). Its cytoplasmic tail contains immunoregulatory signaling motifs called immunoreceptor tyrosine-based inhibitory motif (ITIM)s, which upon ligand binding to the extracellular V-set domain, undergo phosphorylation and recruit effector molecules like tyrosine phosphatases, SHP-1/2, which inhibit the cellular immune responses. Human CD33 (huCD33) binds α2-3- and α2-6-linked *N*-acetylneuraminic acid (Neu5Ac), the predominant Sia in humans, associated either with N- and O-glycosylated molecules or sialylated glycolipids (gangliosides). More recently, a CD33 ligand composed of sialylated keratan sulfate chains carried on a minor isoform/glycoform of RPTPζ (phosphacan) was found in the extracellular milieu of the human brain parenchyma (Gonzalez-Gil et al. 2022).

HuCD33 undergoes alternative splicing, resulting in two isoforms—full-length CD33M containing the ligand-binding V-set domain and truncated D2-CD33 (or CD33m) lacking this domain (Hernandez-Caselles et al. 2006). The elimination of the terminal V-set domain is mediated through differential splicing affected by two co-inherited single nucleotide polymorphisms (SNPs) at positions rs3865444 in huCD33 promoter and rs12459419 located within exon 2 (Malik et al. 2013). The two isoforms, CD33M and D2-CD33, differ not only in their molecular weights, but also in their cellular localization and functionality which are associated with Sia-interacting V-set domain (Schwarz et al. 2016; Siddiqui et al. 2017; Saha et al. 2019).

HuCD33 is extensively studied for its role in different immune responses, under both normal and pathophysiological conditions including cancers (Lamba et al. 2009; Hollingworth et al. 2011; Naj et al. 2011; Bradshaw et al. 2013; Malik et al. 2013). Furthermore, the microglial expression of CD33 is linked with neurological pathologies like LOAD. Incidence of LOAD has been strongly associated with varied expression of CD33 isoforms in the brain of affected individuals (Hollingworth et al. 2011; Naj et al. 2011), where the LOAD-protective CD33 allele increases the ratio of D2–CD33 isoform relative to CD33M. CD33 is reported in almost all vertebrates, including nonhuman primates (Bornhöfft et al. 2018; Khan, de Manuel, et al. 2020). Although there is often high similarity in the sequence and overall genomic location, CD33 has undergone various species-specific changes. For example, murine CD33 which shows about 54% identity with huCD33 V-set and 72% identity with C2 domain, has markedly different Sia-binding and cellular expression patterns from huCD33 protein (Brinkman-Van der Linden et al. 2003). CD33 expression has greatly diverged in humans even in comparison to our closest living evolutionary relatives, the great apes. Examination of CD33 in peripheral blood showed significantly increased production of CD33M in human monocytes relative to those of chimpanzees (Schwarz et al. 2016). Furthermore, the abundance of CD33 was also markedly higher in the human brain. Interestingly, although LOAD-associated neurological pathologies, for example, buildup of Aβ proteins, hyperphosphorylated tau proteins as neurofibrillary tangles, have been observed in aged nonhuman primate brains, AD has largely been regarded as a uniquely human disease (Edler et al. 2017, 2020). Interspecies variations in CD33 have also been studied in other apes like gorilla and bonobo, in comparison to huCD33 (Padler-Karavani et al. 2014).

The presence of two physiologically significant isoforms, their distinct cellular localization, and association with uniquely human pathologies like LOAD have made huCD33 a target of much evolutionary interest. The Sia-binding V-set domain of CD33rSiglecs including CD33 itself shows high sequence variability among different species, often making it difficult to identify their orthologs. The selective pressure for this accelerated evolution of the V-set domains has been attributed to the evasion of infectious pathogens that exploit these human innate receptors. The surfaces of each vertebrate cell are layered with tens to hundreds of million Sia-terminating glycans, forming as "self-associated molecular patterns" (SAMPs), which prevent erroneous activation of innate immune responses against the body's own cells (Varki 2011). However, several human pathogens, for example, *Neisseria gonorrhoeae*, *N. meningitidis*, *Haemophilus influenzae*, *Escherichia coli* K1, Group B *Streptococcus* (GBS), and *Trypanosoma cruzi* cloak themselves with sialoglycans, effectively mimicking host SAMPs, and thereby avoiding the immune response (Varki and Gagneux 2012). Conversely, other infectious agents like influenza virus recognize SAMPs and utilize them as receptors to initiate binding and subsequent infections (Tortorici et al. 2019). CD33 has also been shown to interact with Hepatitis B viral surface sialoglycans, thereby impacting its pathogenesis (Tsai et al. 2021). SAMPs and their interacting partners, Siglecs (primarily the V-set domains) are therefore continually evolving to maintain their distinctive "self-recognition" properties, whereas also avoiding exploitation by Sia-cloaked pathogens and parasites—an example of the "Red Queen Effect" (Varki and Angata 2006).

In this work, with a focus on CD33, a postreproductive cognitive health-associated human protein, we attempt to explore the evolutionary pressures that selected for unique changes in huCD33. Using human-specific pathogens like *N. gonorrhoeae*, GBS, and *E. coli* K1, we demonstrate the differential impact of these mutations on the bacterial interactions with huCD33. We also determine the effect of these mutations on huCD33–sialoglycan binding and identify that the amino acid at position 21 within the V-set domain plays a critical role in Sia-specificity of human and chimpanzee CD33. Last but not least, we extend our study to archaic hominin genomes and show that the human-specific CD33 mutations (except the presence of truncated isoform) are shared evolutionary changes of human, Neanderthal, and Denisovan common ancestor. We also expanded our analysis to include other human-specific derived genomic changes associated with the cognitive health of postreproductive human grandmothers and other elderly caregivers. Finally, we draw an evolutionary scenario to connect the current knowledge of CD33 sialoglycan recognition and pathogen engagement to propose a role for the infectious pathogens as key selective agents in human-specific CD33 evolution, generating new alleles protective against infections, that could secondarily have come under selection for their protective effects against cognitive pathologies like LOAD.

Results

Sequences of huCD33 Extracellular Domains Show Many Changes Distinct from Closely Related Great Apes

Previous investigations have identified unique properties of huCD33 that influence the functionality in humans. The presence of a huCD33 V-set truncated isoform as well as its overall expression difference in microglia has been associated with the protection against the occurrence of neurological pathologies like LOAD in humans. Like other CD33rSiglecs, CD33 immunomodulatory roles depend both on its ligand-interacting extracellular domains and signaling motif-containing cytoplasmic tail. To gain a comprehensive understanding of different CD33 domain variations, we compared the amino acid residues of full-length CD33 from human and related nonhuman primates including chimpanzee, gorilla, and bonobo (fig. 1*A*). Although the regions encoding the C2-set domain and cytoplasmic tail are highly conserved, the amino acid residues within huCD33 V-set domain differ from their nonhuman counterparts in as many as ten positions. Since different amino acid residues in Sia-binding V-set domain could potentially impact huCD33–sialoglycan interactions and subsequent downstream signaling pathways, we further examined the overall frequency of these changes (fig. 1*B*). We analyzed human sequences from the 1000 Genome database (Consortium et al. 2015) and compared them with 44 gorilla, 59 chimpanzee, and 10 bonobo sequences (Prado-Martinez et al. 2013; Xue et al. 2015; de Manuel et al. 2016). Most of these amino acid residues (except at positions 66 and 148) are conserved in all the great apes and appeared to have changed only in the human lineage. Interestingly, the amino acid residues at positions 66 and 148 in huCD33 are isoleucine and leucine respectively, similar to that of chimpanzee and bonobo. The corresponding amino acids in its more distant evolutionary relative, gorilla, are phenylalanine (Phe) (at position 66) and valine (at position 148). The presence of the same amino acid in human, chimpanzee, and bonobo at these positions suggests a more ancient occurrence of these two changes, possibly before the divergence of chimpanzee about 6–8 Ma. In contrast, it has previously been shown that the two linked SNPs, resulting in the splicing of the V-set truncated isoform represent a derived evolutionary modification of the CD33 proteins in humans and are absent in chimpanzees (Schwarz et al. 2016).

To further understand the selection pressure, we calculated the nonsynonymous to synonymous substitution rate ratio (omega, $\omega = \mathrm{d}[N]/\mathrm{d}[S]$) for the CD33 V-set domains of human and other great apes. The omega value of CD33 V-set domain is greater than >1 ($\omega = 1.49$), which reflects V-set domain evolution under positive selection. Subsequently, we also analyzed the Ka/Ks ratios of exon 2 sequences in every species. Except for gorilla, the other two great apes (chimpanzee and bonobo) showed Ka/Ks ratios greater than one indicating that high Ka/Ks ratio of exon 2 is not an accidental event but an evolutionary phenomenon. Taken together, these results demonstrate that CD33 in humans has been rapidly evolving under positive selection, distinct from its orthologs in the great apes.

Archaic Neanderthal and Denisovan Genomes Share Most huCD33 Protein Changes, Except for the SNPs for the LOAD-Protective Allele

Divergence of humans from other ancient hominin lineage such as Neanderthals and Denisovans has been estimated to date back ∼0.5 Ma (Green et al. 2010). Although full-length CD33 itself is an ancient molecule, we noted that the AD-protective CD33 truncated isoform is recently derived in humans, postdating our divergence from Neanderthals and Denisovans (Schwarz et al. 2016). Since huCD33 extracellular domains showed high accumulation of changes compared with the great apes, we wanted to determine if these changes were present in the common ancestor of the hominin lineage. We, therefore, compared CD33 protein sequences from six Neanderthal and two Denisovan archaic genomes obtained from the Max Planck Institute for Evolutionary Anthropology (Prufer et al. 2014) (http://cdna.eva.mpg.de) with the corresponding human sequences of the 1000 Genome database (fig. 1*B*). Interestingly, all the amino acid residues in huCD33 that are different from the great apes are present in the ancient genomes, suggesting their occurrence in a common ancestor. These observations thus suggest that the complete loss of Sia-binding V-set domain is the latest evolutionary modification of huCD33, likely succeeding the individual amino acid changes within its extracellular domain.

A Single Amino Acid Change Facilitated CD33 Engagement to the Uniquely Human Pathogen *Neisseria gonorrhoeae*

In addition to microglial expression in the brain, CD33 is also present on tissue macrophages and peripheral blood monocytes (Schwarz et al. 2016). These cells are important components of innate immune responses throughout the body, including the reproductive tract. The human female reproductive tract is also a unique niche for the

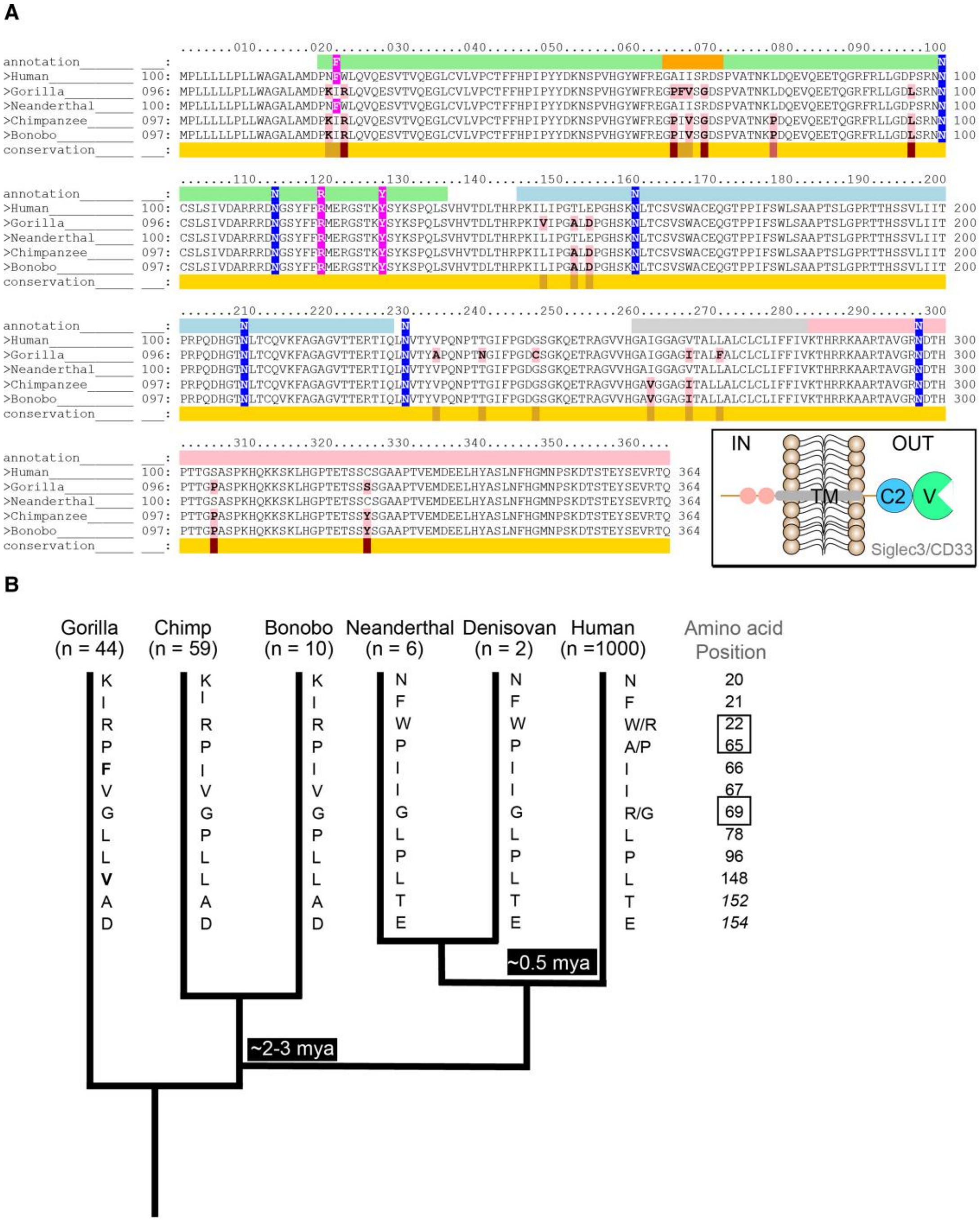

A
.......010.......020.......030.......040.......050.......060.......070.......080.......090.......100
annotation________ ___: F N
>Human____________ 100: MPLLLLLPLLWAGALAMDPNFWLQVQESVTVQEGLCVLVPCTFFHPIPYYDKNSPVHGYWFREGAIISRDSPVATNKLDQEVQEETQGRFRLLGDPSRNN 100
>Gorilla__________ 096: MPLLLLLPLLWAGALAMDPKIRLQVQESVTVQEGLCVLVPCTFFHPIPYYDKNSPVHGYWFREGPFVSGDSPVATNKLDQEVQEETQGRFRLLGDLSRNN 100
>Neanderthal______ 100: MPLLLLLPLLWAGALAMDPNFWLQVQESVTVQEGLCVLVPCTFFHPIPYYDKNSPVHGYWFREGAIISRDSPVATNKLDQEVQEETQGRFRLLGDPSRNN 100
>Chimpanzee_______ 097: MPLLLLLPLLWAGALAMDPKIRLQVQESVTVQEGLCVLVPCTFFHPIPYYDKNSPVHGYWFREGPIVSGDSPVATNKPDQEVQEETQGRFRLLGDLSRNN 100
>Bonobo___________ 097: MPLLLLLPLLWAGALAMDPKIRLQVQESVTVQEGLCVLVPCTFFHPIPYYDKNSPVHGYWFREGPIVSGDSPVATNKPDQEVQEETQGRFRLLGDLSRNN 100
conservation______ ___:
.......110.......120.......130.......140.......150.......160.......170.......180.......190.......200
annotation________ ___: N R Y N
>Human____________ 100: CSLSIVDARRRDNGSYFFRMERGSTKYSYKSPQLSVHVTDLTHRPKILIPGTLEPGHSKNLTCSVSWACEQGTPPIFSWLSAAPTSLGPRTTHSSVLIIT 200
>Gorilla__________ 096: CSLSIVDARRRDNGSYFFRMERGSTKYSYKSPQLSVHVTDLTHRPKIVIPGALDPGHSKNLTCSVSWACEQGTPPIFSWLSAAPTSLGPRTTHSSVLIIT 200
>Neanderthal______ 100: CSLSIVDARRRDNGSYFFRMERGSTKYSYKSPQLSVHVTDLTHRPKILIPGTLEPGHSKNLTCSVSWACEQGTPPIFSWLSAAPTSLGPRTTHSSVLIIT 200
>Chimpanzee_______ 097: CSLSIVDARRRDNGSYFFRMERGSTKYSYKSPQLSVHVTDLTHRPKILIPGALDPGHSKNLTCSVSWACEQGTPPIFSWLSAAPTSLGPRTTHSSVLIIT 200
>Bonobo___________ 097: CSLSIVDARRRDNGSYFFRMERGSTKYSYKSPQLSVHVTDLTHRPKILIPGALDPGHSKNLTCSVSWACEQGTPPIFSWLSAAPTSLGPRTTHSSVLIIT 200
conservation______ ___:
.......210.......220.......230.......240.......250.......260.......270.......280.......290.......300
annotation________ ___: N N N
>Human____________ 100: PRPQDHGTNLTCQVKFAGAGVTTERTIQLNVTYVPQNPTTGIFPGDGSGKQETRAGVVHGAIGGAGVTALLALCLCLIFFIVKTHRRKAARTAVGRNDTH 300
>Gorilla__________ 096: PRPQDHGTNLTCQVKFAGAGVTTERTIQLNVTYAPQNPTNGIFPGDCSGKQETRAGVVHGAIGGAGITALFALCLCLIFFIVKTHRRKAARTAVGRNDTH 300
>Neanderthal______ 100: PRPQDHGTNLTCQVKFAGAGVTTERTIQLNVTYVPQNPTTGIFPGDGSGKQETRAGVVHGAIGGAGVTALLALCLCLIFFIVKTHRRKAARTAVGRNDTH 300
>Chimpanzee_______ 097: PRPQDHGTNLTCQVKFAGAGVTTERTIQLNVTYVPQNPTTGIFPGDGSGKQETRAGVVHGAVGGAGITALLALCLCLIFFIVKTHRRKAARTAVGRNDTH 300
>Bonobo___________ 097: PRPQDHGTNLTCQVKFAGAGVTTERTIQLNVTYVPQNPTTGIFPGDGSGKQETRAGVVHGAVGGAGITALLALCLCLIFFIVKTHRRKAARTAVGRNDTH 300
conservation______ ___:
.......310.......320.......330.......340.......350.......360....
annotation________ ___:
>Human____________ 100: PTTGSASPKHQKKSKLHGPTETSSCSGAAPTVEMDEELHYASLNFHGMNPSKDTSTEYSEVRTQ 364
>Gorilla__________ 096: PTTGPASPKHQKKSKLHGPTETSSSSGAAPTVEMDEELHYASLNFHGMNPSKDTSTEYSEVRTQ 364
>Neanderthal______ 100: PTTGSASPKHQKKSKLHGPTETSSCSGAAPTVEMDEELHYASLNFHGMNPSKDTSTEYSEVRTQ 364
>Chimpanzee_______ 097: PTTGPASPKHQKKSKLHGPTETSSYSGAAPTVEMDEELHYASLNFHGMNPSKDTSTEYSEVRTQ 364
>Bonobo___________ 097: PTTGPASPKHQKKSKLHGPTETSSYSGAAPTVEMDEELHYASLNFHGMNPSKDTSTEYSEVRTQ 364
conservation______ ___:
IN
OUT
TM
C2
V
Siglec3/CD33
B
Gorilla (n = 44)
Chimp (n = 59)
Bonobo (n = 10)
Neanderthal (n = 6)
Denisovan (n = 2)
Human (n =1000)
Amino acid Position
K I R P F V G L L V A D
K I R P I V G P L L A D
K I R P I V G P L L A D
N F W P I I G L P L T E
N F W P I I G L P L T E
N F W/R A/P I I R/G L P L T E
20 21 22 65 66 67 69 78 96 148 152 154
~0.5 mya
~2-3 mya

Fig. 1. Human-specific changes in CD33 are primarily present in the Sia-binding V-set domain. (*A*) Comparison of amino acid sequences of CD33 from humans and "great apes" was performed using Conformational Analysis Tools software. The Great ape genomes included in the analysis are gorilla, chimpanzee, and bonobo and were compared against the human protein as the template. The conservation of the sequence is indicated with yellow being the most and red being the least conserved regions. Amino acids that are different from huCD33 are highlighted in pink. Amino acids encoding the different CD33 domains are indicated above the sequence with different colors corresponding to schematic in inset, namely, V-set domain in green, C2 domain in light blue, transmembrane domain in gray, and cytoplasmic end in light pink. The flexible C–C′ loop is indicated in orange. Amino acids that are in contact with Neu5Ac in huCD33 are highlighted in magenta and N-glycosylation sites in blue. (*B*) Phylogenetic analysis of the evolution of the extracellular domains of CD33 proteins from human, Great apes, and archaic genomes. The number of genomes (*n*) for each group included in the analysis is indicated. Human and great ape CD33 sequences were compared with six Neanderthal and two Denisovan genomes. Amino acid changes present in huCD33 were also present in the ancient genomes. The positions of the amino acids that are different between human and the apes are mentioned, and the identity of the amino acid present in the corresponding positions for each group is indicated by the single letter abbreviations along the branch. Amino acids at positions 152 and 154 are within the C2 domain of CD33 protein and italicized. Polymorphisms within the human population at positions 22, 65, and 69 of CD33 protein are indicated. Amino acids in gorilla CD33 at positions 66 and 148 are different from other apes and are bold. Possible timeline for the diversion of the hominin lineage is indicated in the tree. Length of the branches in the tree is not to scale.

microbiome, which can be invaded by important pathogens like *N. gonorrhoeae* (Ng). Ng is a uniquely human infectious agent, responsible for the second most prevalent, sexually transmitted infection causative for the disease gonorrhea in human populations. Gonorrhea affects both males and females and if untreated, can have detrimental effects on reproductive health (Edwards and Apicella 2004). We have previously shown that Ng interacts with huCD33 but not the chimpanzee ortholog (Landig et al. 2019). The bacterium is incapable of endogenous Neu5Ac synthesis, but instead scavenges the molecule from its host (Parsons et al. 1988; Apicella et al. 1990). Once inside the female reproductive tract, Ng utilizes the host sugar nucleotide CMP-Neu5Ac from its microenvironment to transfer Neu5Ac onto its own bacterial lipooligosaccharide. Sialylated Ng then successfully interacts with several human Siglecs including 3 (CD33), 5, 9, 11, 14, and 16 (Landig et al. 2019). However, unlike other Siglec interactions, Ng binding to CD33 appears to be entirely Sia dependent. Interestingly, of all the *Neisseria* species currently known, only Ng and *N. meningitidis* are pathogenic to humans and both are thought to be evolutionarily young compared with others (Caugant and Brynildsrud 2020). Since reproductive health/success of an organism is the key determinant of Darwinian fitness, we hypothesized that highly infectious disease like gonorrhea could potentially impact the evolution of humans, mediated through binding immune-modulating proteins like CD33.

To explore our hypothesis, we examined the binding of sialylated Ng to different recombinant CD33 protein mutants, each containing the two extracellular domains with an amino acid residue changed from human to chimpanzee at the corresponding positions identified in figure 1*B*. Fluorescently labeled Ng was allowed to interact with human recombinant Fc-chimeric constructs of the CD33 proteins that were immobilized onto protein A-coated plates (fig. 2*A* and *B*). Sia-dependence of the interaction was confirmed by comparing binding with bacteria grown in the presence and absence of CMP-Neu5Ac (supplementary fig. S1*A*, Supplementary Material online). We observed a significant reduction in bacterial binding to chimpanzee CD33 (chCD33) compared with human protein containing both V- and C2-domains. However, in the absence of the V-set domain in the truncated form of huCD33 (CD33m), bacterial binding was lost. Except for the residue at position 21, all the other amino acid alterations from human to chimpanzee CD33 maintained high bacterial binding. In fact, changing the amino acid residues at positions 22, 65 (of the V-set domain), 152, and 154 (of C2 domain) increased the binding significantly compared with wildtype huCD33. In contrast, mutating the amino acid at position 21 from human to chimpanzee residue completely abolished huCD33 binding of sialylated Ng. Interestingly, mutating the chimpanzee CD33 amino acid at position 21 to its corresponding human residue enabled Ng to now engage chimpanzee CD33 (fig. 2*C* and *D*). Considering that Ng and its closest relative meningococcus are both uniquely human pathogens thought to have evolved from commensal *Neisseria* (Seifert 2019), our data suggest important implications of CD33 amino acid change at position 21 on Ng–huCD33 interaction and their mutual evolution.

Many Amino Acid Changes in CD33 Extracellular Domains Impact *GBS* Engagement

Although the association of *Neisseria* with CD33 is a case of Sia-mediated interaction, there are other examples of human pathogens that engage Siglecs in Sia-independent manner. One such example is GBS, which has been widely studied for its various Siglec interactions (Carlin et al. 2007). GBS is an encapsulated pathogen commonly associated with pneumonia, sepsis, and meningitis in infants and neonates. It comprises nine serologically distinct groups (Ia, Ib, and II–VIII), differing in their capsular sialoglycan structures, but all containing α2-3-linked terminal Neu5Ac. Certain GBS strains have been shown to bind human Siglecs 5 and 7 in a Sia-independent manner through cell wall anchored β-protein (Fong et al. 2018), whereas some Sia-dependent binding was observed for CD33 and Siglec-9. Human Siglec-9 binding is also thought to be partially Sia independent. Interestingly, some GBS strains are also known to interact with nonhuman primate Siglecs, for example, Siglec-9 from chimpanzee (Padler-Karavani et al. 2014). Since infections by GBS mostly impact newborns and infants, we hypothesized that it could also play a role in overall Siglec evolution in humans. Similar to the Ng-CD33 binding assay (as in fig. 2*A*), we examined the interactions between the recombinant CD33 proteins and GBS group III strain, COHI (fig. 2*E*). Although the bacteria bound strongly with full-length extracellular domains of huCD33, the binding was significantly reduced in the truncated human isoform (CD33m) and the chimpanzee protein. Like Ng, GBS COHI interaction was also markedly disrupted by amino acid changes at position 21. Additionally, changing the residues at positions 20 and 65 from human to chimpanzee significantly reduced the bacterial interaction with CD33. However, GBS COHI engagement with the CD33 mutants was not entirely Sia dependent (fig. 2*F*). Using GBS COHIΔ*neuA*, a mutant strain lacking its sialyltransferase enzyme (NeuA) and hence incapable of surface sialylation, we observed that about 50% of the bacterial binding to CD33 could be attributed to Sia-independent interactions. Interestingly, the CD33 binding profile of COHI was not uniform for the other serogroups of GBS, for example, GBS group Ia strain, A909 (fig. 2*G*). None of the amino acid changes showed significant effects on CD33 interaction with GBS A909, relative to the wildtype human protein. Even the truncated huCD33 isoform (CD33m) displayed similar binding suggesting that the CD33 binding for A909 is primarily Sia independent. Unlike Ng and GBS, we did not observe any differential sialoglycan binding with *E. coli* K1, another uniquely human pathogen of newborn infants, which contains Sia polymers on its surface (supplementary fig. S1*B*, Supplementary Material online). Altogether, the data demonstrate the diverse nature of CD33-interactions in just three major

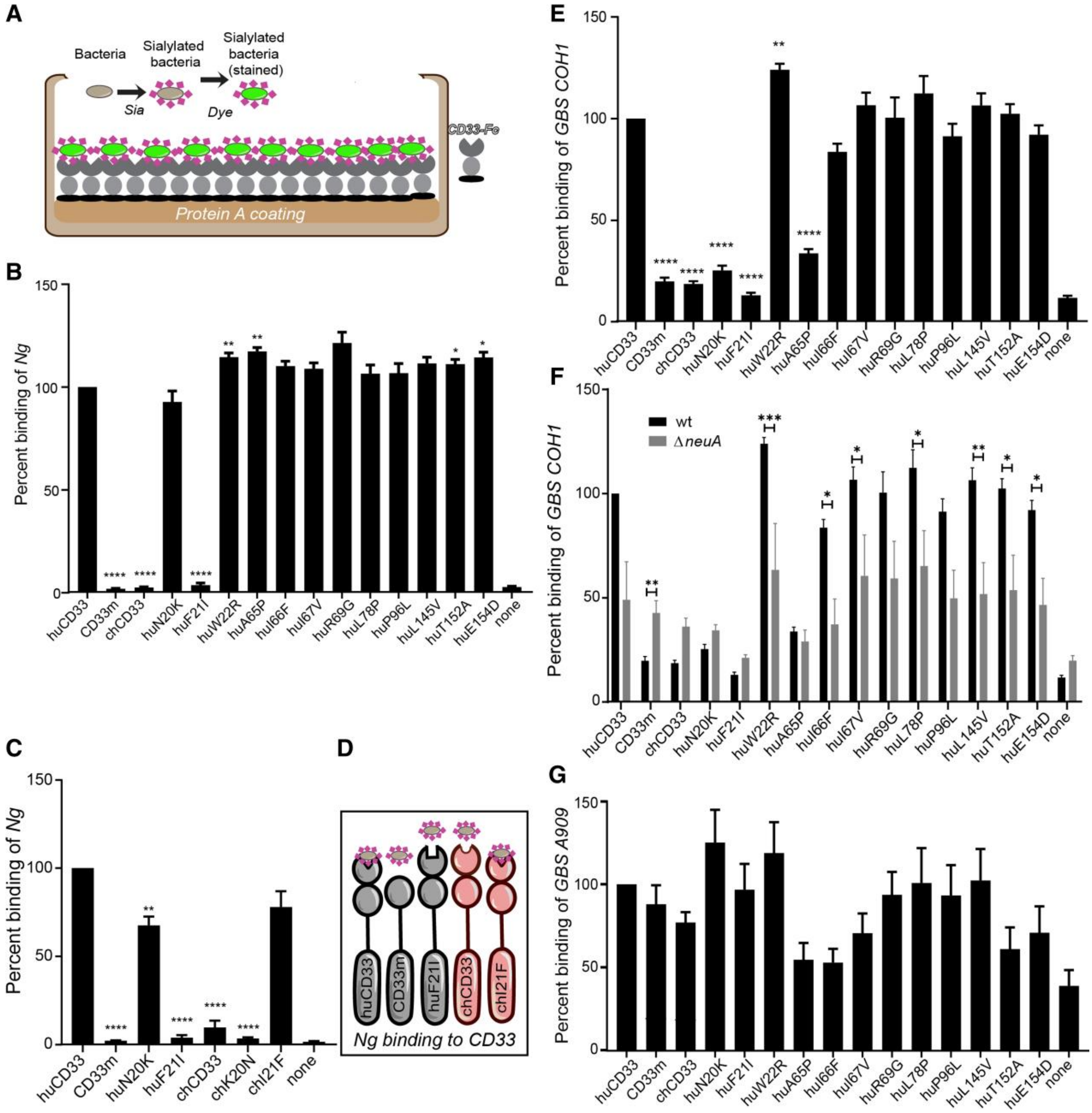


Fig. 2. Human-specific amino acid changes in CD33 affects bacterial binding. (*A*) Schematic of the ELISA-based assay using recombinant CD33-Fc chimeric proteins immobilized on protein A coated plates used to determine binding of the sialylated bacteria is shown. (*B*) Binding of fluorescently labeled *Neisseria gonorrhoeae (Ng)* was determined. The position of the amino acid different from the wildtype huCD33 protein is indicated below each bar in the *x*-axis. The bacterial binding to each individual CD33 mutant was normalized to the binding of wildtype huCD33 for that assay. "None" indicates no protein control for the background bacterial binding to the plate. (*C*) Binding of *Ng* to immobilized recombinant CD33 proteins containing the corresponding amino acid mutation (position 20 or 21) either in human or chimp CD33 protein backbone. (*D*) Pictorial summary of Ng binding to CD33 wild type and mutant proteins. (*E*) Binding of GBS COH1 strain to different CD33 mutant proteins in an ELISA-based assay with immobilized recombinant CD33 proteins. (*F*) Sialic acid dependence of the binding was determined using wildtype and Δ*neuA* mutant strains of COH1. (*G*) Interaction of CD33 proteins among different GBS strains was compared using A909 and COH1 strains. 'hu' indicates the corresponding amino acid change in huCD33 backbone and 'ch' using chimp CD33. The graphs show the cumulative result from three independent experiments, each done in triplicate. Statistical analysis was performed in Prism software using one-way ANOVA with Durrett post-comparison test. *<0.01, **<0.001, ***<0.0001.

pathogens and suggest an impact of uniquely human pathogens in the evolution of CD33 ligand-binding domain.

Ancestral Sialoglycan Preference of CD33 is Disrupted by Amino Acid Change at Position 21

A key change in the evolution of humans was the loss of CMP-Neu5Ac hydroxylase (CMAH), the enzyme that converts CMP-Neu5Ac to CMP-Neu5Gc resulting in a primarily Neu5Ac-rich sialome in humans, unlike any other Old-World primates, which express both Neu5Ac and Neu5Gc. This change is dated to ~2–3 Ma when human ancestors were evolving from ancestral hominins. Since we observed numerous changes mainly in huCD33 V-set domain, which is critical in sialoglycan interaction and therefore important for its downstream signaling pathways, we wanted to specifically understand the effect on CD33 sialoglycan interactions. We used a microarray of chemoenzymatically synthesized glycans with defined structures, terminally capped with either Neu5Ac or Neu5Gc in different glycosidic linkages (fig. 3*A* and *B*) and examined their relative interactions with recombinant, soluble CD33 proteins (fig. 3). HuCD33 with V- and C2- domains bound to both Neu5Ac and Neu5Gc-terminating sialoglycans and showed maximum binding when the Sia was α2-6-linked to an underlying lactose or lactosamine glycan (supplementary fig. S2, Supplementary Material online). Most of this binding was lost in the truncated huCD33 lacking the Sia-binding V-set domain, indicating that the interactions are Sia dependent. Conversely, the chimpanzee protein (which is identical to the bonobo orthologs and differs by only two amino acids from the gorilla) demonstrated a strong preference toward Neu5Gc-terminating sialoglycans and showed almost no binding for Neu5Ac-epitopes (fig. 3*A* and 3*C*). Considering the varied sialoglycan profiles of the two organisms, these distinct binding preferences of human and chimpanzee CD33 are interesting and suggest functional implications of the evolutionary changes in their extracellular domains.

Because the impact of amino acid residue at position 21 was most pronounced in both of our bacterial-CD33 binding assays (fig. 2*B* and *D*), we next examined the influence of this change on CD33–sialoglycan binding (fig. 3 and supplementary fig. S2, Supplementary Material online). Indeed, changing the amino acid at position 21 completely altered Sia-epitope preference of CD33 for both human and chimpanzee. The presence of human amino acid residue at position 21 enabled strong binding of Neu5Ac-epitopes by chCD33, unlike its entirely Neu5Gc-preferring wildtype counterpart (fig. 3*A* and *C*). On the other hand, the chimpanzee amino acid at the same position in huCD33 abolished its Neu5Ac binding. To determine if the Sia-binding changes are specific for position 21 and not an arbitrary effect of any amino acid change in V-set domain, we also looked at the Sia-epitopes of position 20 amino acid substitutions. Unlike position 21, amino acid modifications at position 20 did not have any major impact on the Sia-binding of CD33, which maintained the overall wildtype profile. Interestingly, modifications at position 22 demonstrated Neu5Ac-prefered binding for huCD33, whereas chimpanzee amino acid residues at 65 and 66 of huCD33 almost abolished any sialoglycan binding. Altogether, the data emphasized the importance of different amino acid changes in huCD33 V-set domain in its sialoglycan binding and identified the amino acid at position 21 to be critical in the functionality of CD33 protein.

Molecular Dynamics Simulations Provide Structural Insights for the Differences in Sia-Binding Preference Between Human and Chimpanzee CD33

We performed an extensive theoretical investigation based on molecular dynamics (MD) simulations. A detailed analysis of several available crystal structures of huCD33 revealed that the V-set domain is dynamic. For example, the C–C′ loop as well as the side chains of phenylalanine at position 21 (Phe21) and histidine at position 45 (His45) are resolved in two different conformations in PDB entry 5ihb (supplementary fig. S3, Supplementary Material online). Of all the amino acids that differ between human and chimpanzee, only the side chain of Phe21 is in direct contact with a bound Neu5Ac residue in the crystal structures of huCD33 (through the methyl group at position 5). Based on the assumption that Neu5Gc binds to the same binding site as Neu5Ac, the change in binding preference from Neu5Gc (in chimpanzee) to Neu5Ac (in human) cannot be explained by a simple I21F mutation. Both amino acids have hydrophobic side chains that cannot establish favorable interactions with the polar glycolyl group. Consequently, there is probably a more complex reason for the shift of binding preference. Based on data derived from 47 MD simulations covering an accumulated timescale of more than 100 µs, we conclude that in chCD33 His45 adopts mainly the "up" conformation (compare fig. 4*A*), which allows favorable hydrogen bonding with the glycolyl group. MD simulations (as well as X-ray crystallography) show that in huCD33 His45 can also exist in the "up" conformation (compare figs. 4*C* and supplementary fig. S4, Supplementary Material online), which would be compatible with favorable Neu5Gc binding. However, when His45 is in the "down" conformation Phe21 can stack partly with tyrosine (Tyr) at position 127 (fig. 4*B*) forming a small hydrophobic pocket, which allows the methyl group of Neu5Ac to bind favorably. To demonstrate if the binding affinity difference between Neu5Ac and Neu5Gc may be indeed correlated to the up/down conformational equilibrium of His45, we performed a series of MD simulations of chCD33 on the microsecond timescale where Neu5AcOMe or Neu5GcOMe molecules are present in the solution. The lifetimes of the complexes spontaneously formed during the MD with Neu5Gc are on average much longer when His45 is "up" (fig. 4*D* top, supplementary fig. S4, Supplementary Material online). In contrast, the lifetimes of the complexes spontaneously formed with Neu5Ac are much shorter independent of the conformational state of His45 (fig. 4*D* bottom), which would explain the lack of measurable binding affinity of Neu5Ac to chCD33. In summary, our extensive MD simulations—including unbiased simulation of carbohydrate binding and unbinding events—could provide a reasonable explanation for a change in binding specificity that is likely to be caused by an alteration of the protein–ligand interaction pattern remote from the mutated amino acid.

Human-Specific Polymorphisms in Cognitive Health-Related Genomic Variants are Present in All Human Populations

In an earlier study, we observed several genes, directly associated with neurodegenerative diseases or correlated with aggravation of the cognitive decline in aged-

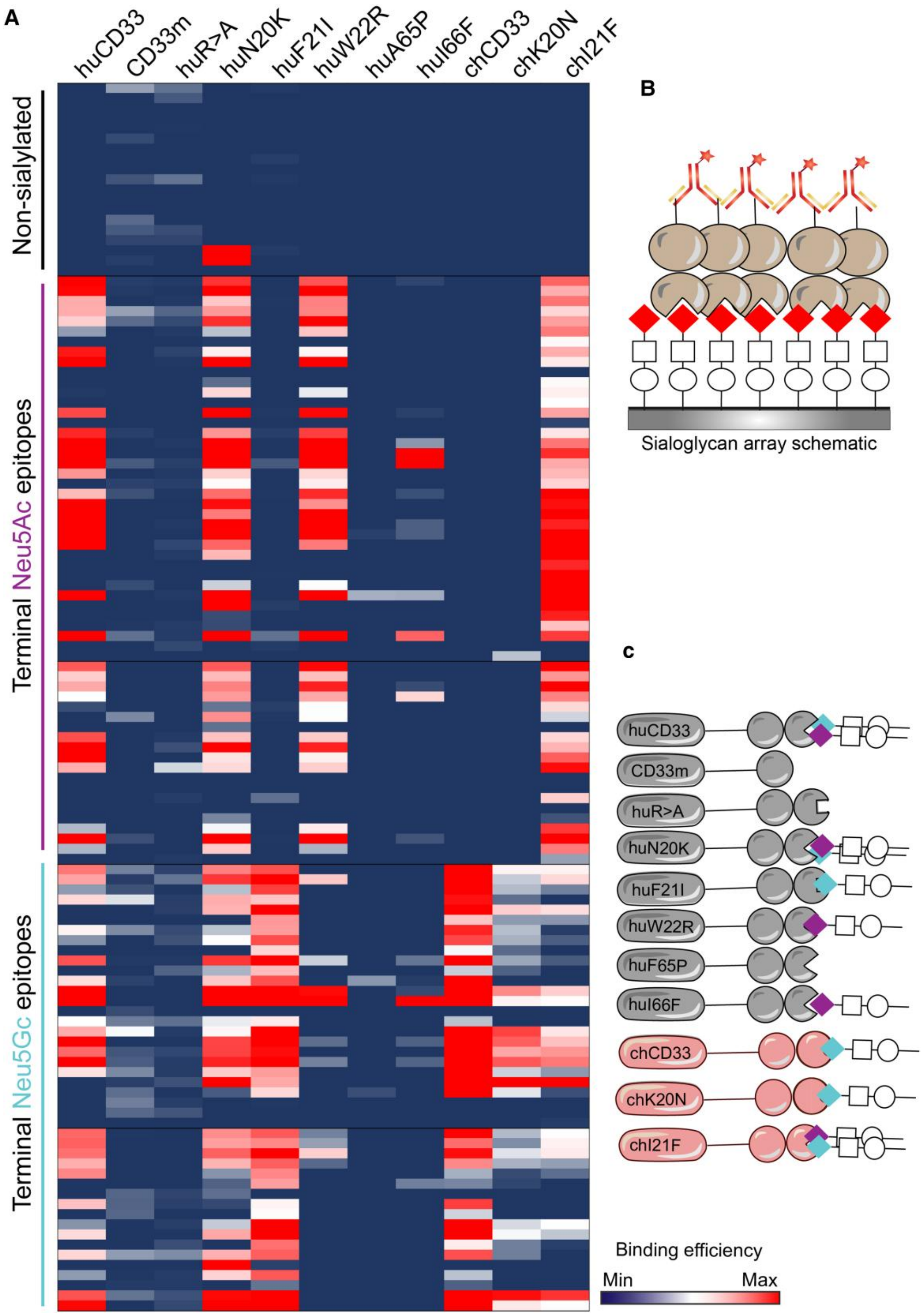


Fig. 3. Single amino acid changes affect CD33 sialoglycan binding. (*A*) Sialoglycan-binding profile of purified, soluble, recombinant CD33 proteins was determined using a sialoglycan microarray containing defined, chemically synthesized glycans (experiment schematic shown in *B*). Nonsialylated, Neu5Ac- (indicated in purple on the left) and Neu5Gc- (indicated in blue) terminating glycans were grouped together in the heatmap as shown in the left. Each column indicates the binding profile of the protein indicated on the top and each row represent a distinct glycan. Blue indicates no binding and red indicates very strong binding preferences characterized by an average relative fluorescence unit (RFU) of more than 90th percentile. The result of the heatmap is summarized in the cartoon (*C*) with the colored diamonds corresponding to the termination Sia and supplementary figure S2, Supplementary Material online. The names of the individual glycans are presented in supplementary file S1, Supplementary Material online.

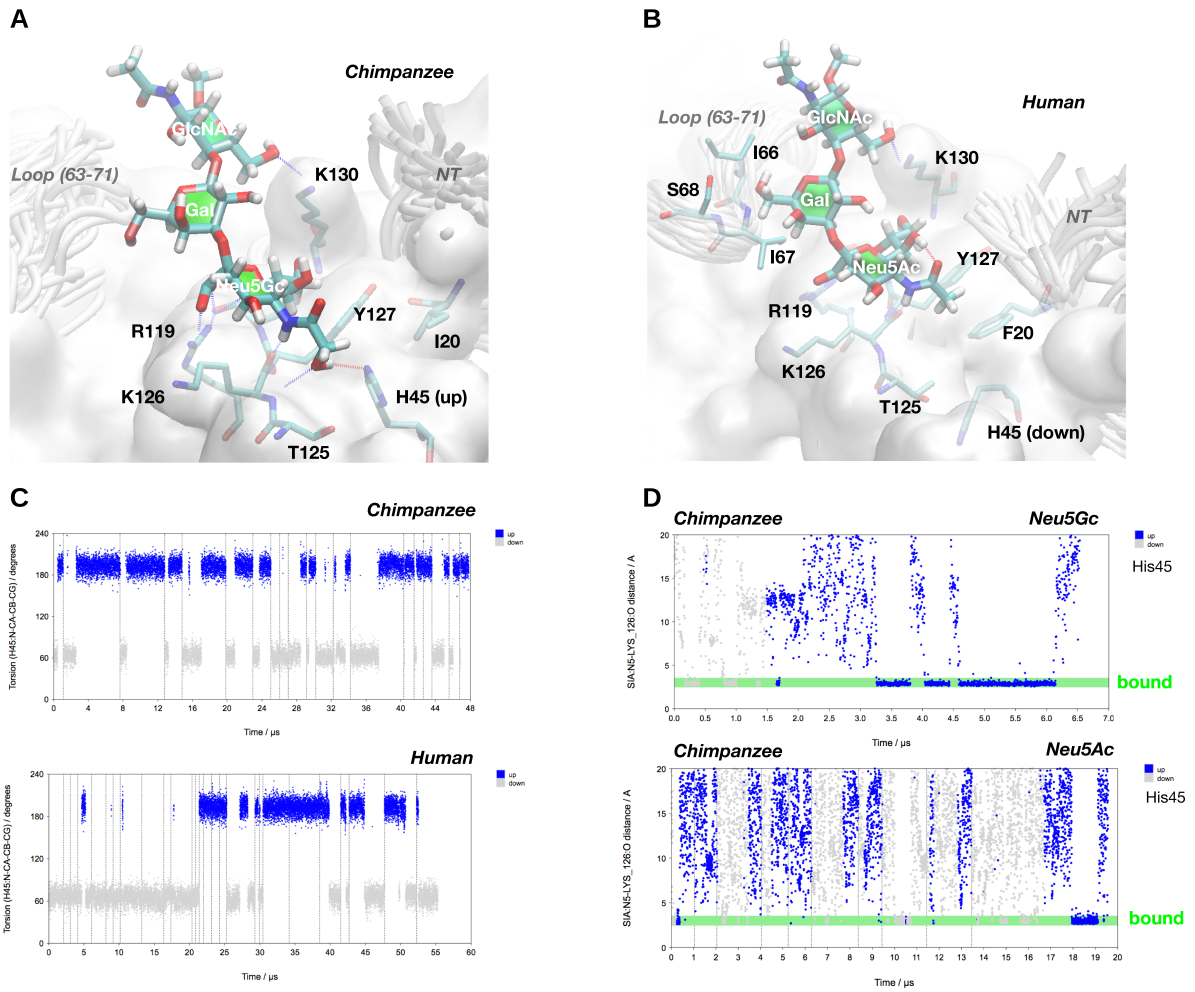


Fig. 4. Structural modeling to understand the differential binding preference of human and chimpanzee CD33 proteins. (*A*) 3D model of the complex between Neu5**Gc**α2–3Galβ1–4GlcNAcβOMe and chCD33. The increased affinity of Neu5Gc may be explained by intermolecular hydrogen bonds involving the OH-group of Gc. It should be noted that the number of favorable interactions is maximal when His45 is in "up" conformation. (*B*) 3D model of the complex between Neu5**Ac**α2–3Galβ1–4GlcNAcβOMe and huCD33. The methyl group of Ac is located in a small hydrophobic pocket formed by the side chains of Tyr127 and Phe20. It should be noted that His45 is in "down" conformation because otherwise—in the conformation shown—the bulky side chain of Phe20 would overlap partly with His45 in "up" conformation. (*C*) MD of His45 side chain orientation. Accumulated MD trajectories of torsion angle N–Cα–Cβ–Cγ are shown. The "up" conformation is present when torsion values are fluctuating around 200° and the "down" conformation is characterized by values around 70°. For chimpanzee, it can be observed reproducibly that simulations started with His45 in "down" conformation undergo a transition to the "up" conformation on the microsecond timescale. In contrast the "down" conformation appears to be more stable in huCD33, which would make Neu5Ac binding more likely. (*D*) MD simulation of unbiased binding and unbinding events of Neu5Gc (top) and Neu5Ac (bottom) to chCD33. For Neu5Gc the lifetime of the complex is significantly longer when His45 is in "up" conformation, as can be seen from the 6.5 µs MD simulation shown on the top. Also, for Neu5Ac multiple binding and unbinding events occurred on a timescale of about 20 µs, however, in general (with one exception), the lifetimes of the complexes formed are significantly shorter than for Neu5Gc.

population, are derived alleles in humans (Schwarz et al. 2016). Increasing evidence of correlation between cognitive health and nonneurological, metabolic conditions, for example, diabetes (Peila et al. 2002, Munshi 2017) suggest that such derived alleles could be important in the maintenance of cognitive health in human grandparents. Here, we expanded this list of cognition-protective gene variants through the literature and database (https://alzoforum.org) searches (Finch and Sapolsky 1999; Altshuler et al. 2000; Nakajima et al. 2004; Thompson et al. 2004; Vander Molen et al. 2005; Helgason et al. 2007; Wen et al. 2007; Carrasquillo et al. 2011; Naj et al. 2011; Raj et al. 2014; Xu et al. 2015; Liu et al. 2018; Rathore et al. 2018) to include additional gene variants, namely *BINI*, *ARID5B*, *PICALM*, *PILRA*. Supplementary table S1, Supplementary Material online describes the characteristics of 13 human genes that are implicated in diseases including dementia, cardiovascular diseases, hypertension, and AD. Although some of these physiological abnormalities like salt retention, hypertension, diabetes, appear nonneurological, they have been associated with the aggravation of the pathologies resulting in late life cognitive decline (Iadecola 2013). Notably, the derived alleles are common and found in globally diverse human populations, indicating that they predate the common ancestor of modern humans (supplementary table S2, Supplementary Material online).

SNPs Associated with Human-Specific Cognitive-Protective Alleles are Unusual in their Absence in the Archaic Hominin Genomes

With the availability of genomes from extinct archaic hominins (Reich et al. 2010; Meyer et al. 2012; Prufer et al. 2014), a set of SNPs can be assessed as to whether their protective phenotypes arose recently in the evolutionary history of anatomically modern humans. We previously showed that many other human–chimpanzee differences were shared with archaic hominins (Denisovan/Neanderthal) genomes; for example, genomic changes in CD33rSiglecs (Khan, de Manuel, et al. 2020; Khan, Kim, et al. 2020). To gain similar insights about the evolutionary origin of these cognitive-protective loci, we analyzed the Neanderthal and Denisovan reference genomes and compared them with modern human sequences. Analysis of the 1000 Genomes dataset shows the presence of protective alleles in human populations with variable frequency (table 1). Analysis of the available genomic data from Neanderthal and Denisovan genomes showed that only two derived variants (rs2975760 and rs2588969; table 1) are present in these archaic genomes, suggesting the remaining 11 derived, protective variants arose after the divergence of modern and archaic hominins ~0.5 Ma (Green et al. 2010; Mendez et al. 2016). This is in striking contrast to most human–chimpanzee genomic differences in which the archaic hominins are similar to humans. In fact, the majority of the Sia-related genes lack positive selection signatures and rather show neutral evolution in the modern human lineage (Moon et al. 2018). To more formally assess whether the high frequency, global distribution, and recent origin observed for 11 of the 13 SNPs is unusual, we performed a resampling analysis of variants in the genome. Variants in the 1000 genomes dataset that met the following criteria were considered: (1) present in both Altai Neanderthal and Denisovan minimal filters, (2) derived in at least one modern individual from nonadmixed African populations, (3) called in both archaic samples, and (4) have an ancestral allele matching the reference

Table 1. Gene Variants Directly or Indirectly Affecting Cognitive Function.

Gene	Associated Disease[a]	SNP ID	hg19 Position	Allele	Derived Global AF[b]	Archaic Genotype	
						Altai	Denisovan
AGT	Sodium retention	rs699	1:230845794	G A	0.295	0/0	0/0
BIN1	AD	rs7561528	2:127889637	A G	0.8	0/0	0/0
SGC2	Hypertension	rs1017448	2:224466344	T C	0.879	0/0	0/0
CAPN10	Type II Diabetes	rs2975760	2:241531163	C T	0.882	1/1	1/1
PPARG	Type II Diabetes	rs1801282	3:12393125	C G	0.070	0/0	0/0
CYP3A5	Sodium retention	rs776746	7:99270539	T C	0.621	0/0	0/0
ARID5B	AD	rs2588969	10:63611354	A C	0.532	0/0	1/1
SPON1	Dementia	rs2618516	11:14021639	C T	0.341	0/0	0/0
PICALM	AD	rs10792832	11:85867875	G A	0.313	0/0	0/0
		rs3851179	11:85868640	C T	0.315	0/0	0/0
APOE	LOAD	rs429358	19:45411941	C T	0.849	0/0	0/0
		rs7412	19:45412079	C T	0.075	0/0	0/0
CD33	LOAD	rs3865444	19:51727962	C A	0.211	0/0	0/0
PILRA	AD	rs1859788	7:99971834	G A	0.341	–	–
TCFLC2	Type II Diabetes	rs7903146	10:114758349	T C	0.772	–	–
CD33	LOAD	rs12459419	19:51728477	C T	0.211	–	–

NOTE.—Allele has the derived allele as the lower, bolded entry. Archaic genotypes are reported for SNPs passing all quality filters.

[a] See supplementary table 1, Supplementary Material online for details and the primary literature citations.

[b] See supplementary table 2, Supplementary Material online for the global population distribution.

or alternative allele. To eliminate any bias in the analysis and match the allele frequency (AF) of these SNPs compared with that of any random SNPs, we first matched our universe of SNPs to the 13 SNPs of interest by AF, ± 2 derived haplotypes (fig. 5). Resampling was then performed by drawing a SNP from each of the 13 matched sets and assessing how many derived alleles were observed, resulting in a *P*-value = 0.08333 ± 0.00003. As a less conservative estimate, directly sampling from the universe of SNPs and estimating the probability of observing at most two derived SNPs and a mean AF as large as the empirical variants of interest produced a highly significant *P*-value = 0.00487 (supplementary fig. S5, Supplementary Material online). Repeating either analysis on the set of other *SIGLEC*-related SNPs indicates that they are consistent with a random draw from the genome (Khan, de Manuel, et al. 2020). Regardless of the individual limitations, taken together our phylogenetic analyses demonstrate the unique patterns of allele frequencies in worldwide populations distribution of these 13 late life cognitive decline-linked SNPs (fig. 5). Interestingly, coinherited CD33 SNPs associated with the cognitive health in LOAD are present only in modern human genomes (Schwarz et al. 2016). A noteworthy example in our list is the human gene encoding the protein, apolipoprotein E (APOE), involved in fat metabolism in mammals. *APOE* gene exists in three allelic variants (E2, E3, and E4) where APOE4 is associated with high risk of LOAD and other allele like APOE2 is protective against the cognitive decline in elderly caregivers (Reiman et al. 2020). Interestingly the presence of APOE4 is also correlated with the protection from severe diarrhea in children (Oriá et al. 2005). Although conclusive determination of the positive selection of these alleles in modern human requires further analysis, our data suggest that the evolutionary origin of most of these cognitive health-protective changes followed after the divergence of modern humans from archaic genomes. This is also supported by the presence of grandparents, uniquely in humans. Regardless, the process of evolutionary emergence of each of these alleles is likely to be distinct and deserves further investigation.

Discussion

Fossil evidence and genomic comparisons leave little doubt about the fact that our species evolved from an African hominin. However, the detailed understanding of modern human origins is plagued by numerous uncertainties, with regard to the identity of the ancestral lineage and precise geographic locations. The evolution of modern humans was accompanied by many anatomical and behavioral changes, but increasing evidence suggests it also included uniquely human-derived modifications in the genome compared with the archaic genomes (Neanderthal/Denisovan) or the genome of the great apes (Schwarz et al. 2016; Khan, de Manuel, et al. 2020). Taken together with our previous study (Schwarz et al. 2016), we have identified many such human-specific genes associated with cognitive health of grandmothers and other human elders who are often involved in the caregiving of the young. These findings, which appear paradoxical to the concept of senescence due to antagonistic pleiotropy, have lent much additional support to the "Grandmother hypothesis" (Hawkes et al. 1998) bolstering the case for the selection of human female postreproductive survival and the existence of grandmothers. Unlike in any other mammals (except orcas and some other toothed whales), the occurrence of this prolonged postreproductive life span in humans has stirred scientific interest. Although deciphering the precise evolutionary course of any gene/protein is challenging and the proposed schemes/players are not entirely verifiable, here we attempt to compile the current evolutionary and experimental findings of one such protein associated with late life cognitive decline: CD33.

A ratio of high wildtype huCD33 and a low truncated isoform of CD33 have been implicated in the progression of LOAD associated with the cognitive health of elderly population. In contrast, LOAD is unknown in chimpanzees, although evidence of LOAD pathologies has been observed in some chimpanzee brains. We found that huCD33, which

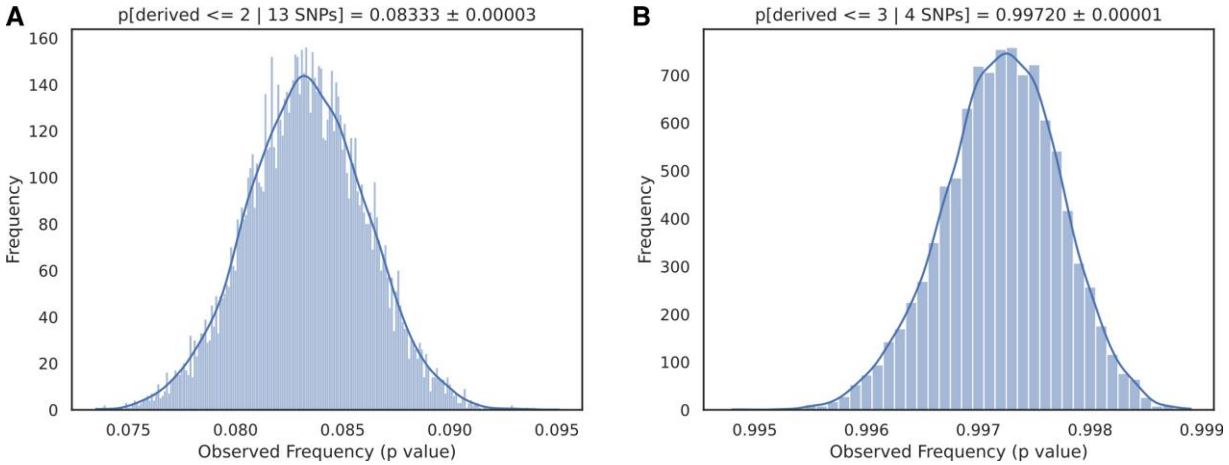


Fig. 5. Resampling analysis with matched AF SNPs from 1000 genomes variants. Frequency distribution of SNPs with similar properties to LOAD-protective set (*A*) and other *SIGLEC* SNPs (*B*).

is highly expressed in microglia of the human but not chimpanzee brain, recognizes Neu5Ac—the predominant Sia synthesized in humans—as SAMPs. In contrast, our closest evolutionary relative, the apes and other old-world primates contain both Neu5Ac and Neu5Gc. We found that the ancestral form of CD33 in chimpanzees and other great apes selectively recognizes only Neu5Gc-glycans as SAMPs (fig. 3). Notably, Neu5Gc—the ligand recognized by chCD33 is rare in chimpanzee brain, and there is also significantly less chCD33 protein compared with CD33 in humans (Schwarz et al. 2016). On the other hand, SNPs resulting in the truncated CD33 have only been observed in the human genome and not any of the archaic or great ape genomes. We also find that the truncated huCD33 does not interact with Sia (fig. 3). Taken together, these observations suggest that full-length CD33–Sia interactions are stronger in human brain compared with chimpanzee and the human-specific SNPs in CD33 resulting in the truncated protein abolish this interaction.

The CD33–Sia (i.e., protein–ligand) interaction in the human brain is important, given the immunoinhibitory function of CD33 in the microglia and the overall roles of microglia in the maintenance of brain homeostasis, including the immune surveillance to clear damaged cells and amyloid plaques comprising of β-amyloid (Aβ) peptides (Colonna and Butovsky 2017). Extracellular accumulation of Aβ peptides and intracellular phosphorylation of the tau proteins are pathophysiology associated with AD brains. Microglial expression of high full-length CD33 protein and/or low truncated CD33m isoform indicates increased interactions with the Sia-ligands and hence, inhibition of the immune activation of microglia mediated by CD33 cytosolic ITIM (or ITIM-like) motifs and therefore, low phagocytosis of the damaged proteins, including amyloid plaques rich in Sia residues (Szumanska et al. 1987; Salminen and Kaarniranta 2009). This is further evident in studies with CD33 null mice where reduced expression of CD33 in the microglia and lower number of CD33 positive microglial cells are correlated with lower brain accumulation of insoluble Aβ (Griciuc et al. 2013). Similar observations were made with CD33 knockout in human microglia and macrophages (Wißfeld et al. 2021), where CD33 absence resulted in increased amyloid plaque clearance. In fact, CD33 is among one of the most common genes linked with AD in genome-wide studies (Naj et al. 2011) and is reviewed as potential candidate for therapeutic interventions against LOAD (Zhao 2019; Griciuc et al. 2020). Notably, cognitive health has also been associated with changes in the inflammatory activity in the brain (Simen et al. 2011; Tangestani Fard and Stough 2019), with increased neuroinflammation observed in aged brain and cognitive pathologies like AD (Eikelenboom and van Gool 2004; Eikelenboom et al. 2006). Altogether, these findings also excite the evolutionary interest in the genomic changes of immunomodulatory proteins like CD33 and their role in the cognitive health in human, especially elders.

In addition to the brain microglia, CD33 is an important immune modulator of hematopoietic cells where this innate receptor prevents the immune activation of the peripheral macrophages against SAMPs on the surface of self-cells. The absence of the ligand-binding V-set domain in truncated CD33m, therefore, prevents the suppression of macrophage phagocytic activity against the self-cells and contributes to more active immune cells. In this regard, it is important to note that unlike full-length CD33, CD33m isoform is not expressed on the cell surface, but instead stored intracellularly (Siddiqui et al. 2017; Saha et al. 2019). Therefore, the question remains what could have led to the selection of the truncated isoform of huCD33 that does not interact with Sia? In this regard, CD33 on macrophages plays crucial roles in different immune responses as well as during infections. HuCD33 has also recently been shown to be involved in immunomodulation during infection with hepatitis B virus (Tsai et al. 2021). Our previous and current data show that uniquely human pathogens like *Neisseria* and GBS display Neu5Ac, that is, recognized as "self" by human but not chimp CD33 (Landig et al. 2019). In the current work, we further found that the Sia-binding-domain-depleted, truncated huCD33 isoform does not bind and thus escaped exploitation by sialylated pathogens (fig. 2). This suggests that this truncated CD33 may have been an adaptation to counter the CD33-exploiting, immune-evasive behavior of pathogens like *Neisseria* and GBS.

Taking together all currently available experimental data (including this study), we attempt to draw a plausible evolutionary scenario for CD33 protein evolution in humans and present in the context of relevant evolutionary events (fig. 6*A*). We hypothesize that the scarcity of the strongly preferred Neu5Gc ligand of ancestral CD33 in the brains of chimpanzee (and other great apes) was associated with low microglial expression. Subsequent hominin loss of CMAH (i.e., complete loss of Neu5Gc ligand) could then have been selected for the upregulation of CD33 levels perhaps to compensate for the loss of ligands, a change to Neu5Acbinding preference, and functional recruitment of CD33 to human microglia. Alongside the microglial CD33, the corresponding changes in the tissue macrophage proteins might have facilitated the emergence of Neu5Ac-coated pathogens (for example, *N. gonorrhoeae* and GBS) that evolved "molecular mimicry" of Neu5Ac-SAMP ligands to manipulate the immune response (fig. 6*B*). Appearance of the truncated isoform lacking the ligand-binding domain (CD33m), then probably allowed CD33 to escape the immune evasion by these sialylated pathogens (fig. 2). This selection pressure to stop manipulation by sialylated pathogens could have also altered splicing toward a higher level of truncated CD33, which also gets diverted to peroxisomes (Siddiqui et al. 2017). Although the significance of this diversion is unclear, decrease of full-length CD33 would facilitate escape from Neu5Ac-coated, CD33-engaging pathogens.

Finally, sometime during the last 2 million years, increased brain size (both relative and absolute) must have resulted in an increasingly prolonged period of development and maturation (Charnov et al. 2001). The selective forces underlying this increase remain to be identified and could range from

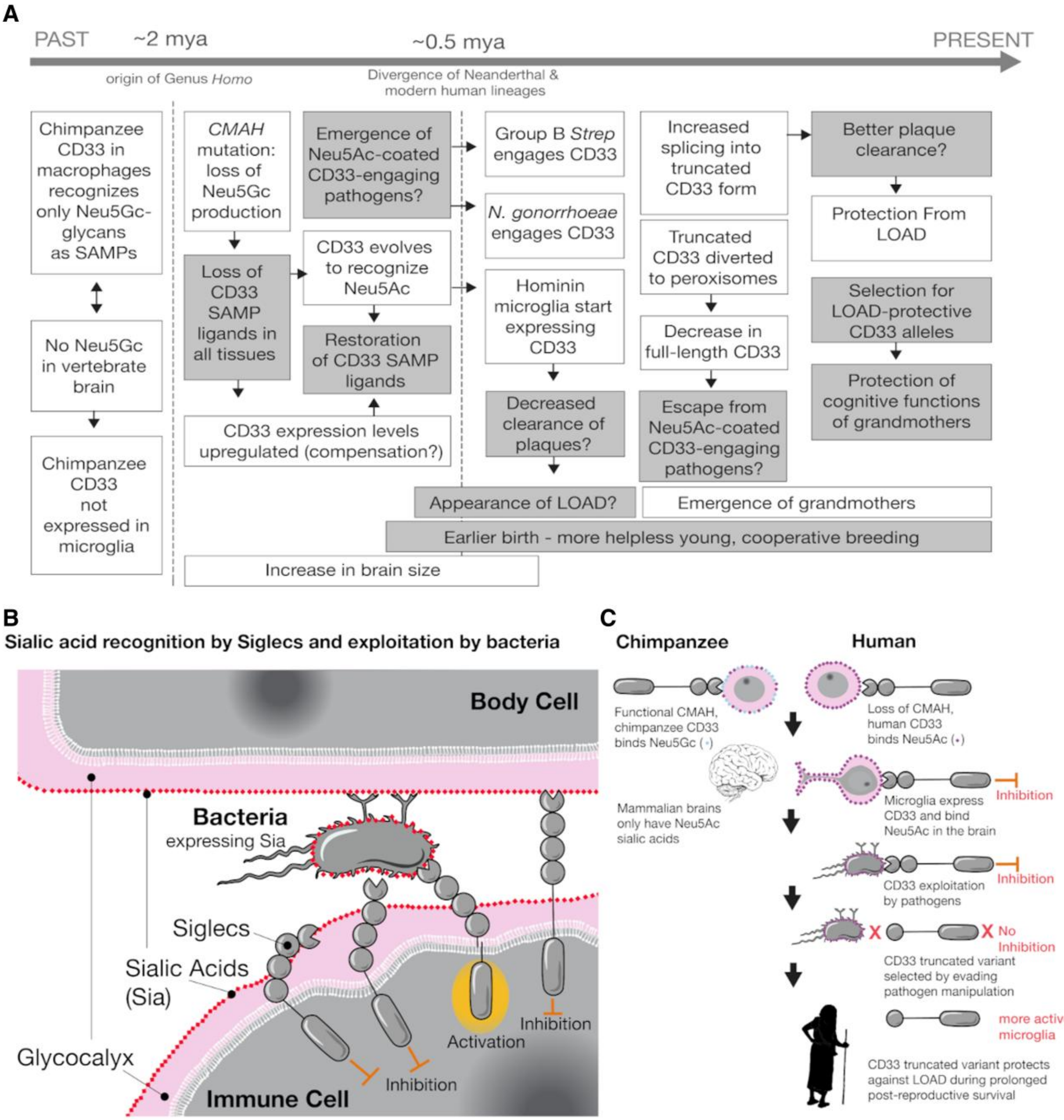

A
PAST ~2 mya ~0.5 mya PRESENT
origin of Genus Homo
Divergence of Neanderthal & modern human lineages
Chimpanzee CD33 in macrophages recognizes only Neu5Gc-glycans as SAMPs
No Neu5Gc in vertebrate brain
Chimpanzee CD33 not expressed in microglia
CMAH mutation: loss of Neu5Gc production
Loss of CD33 SAMP ligands in all tissues
Emergence of Neu5Ac-coated CD33-engaging pathogens?
CD33 evolves to recognize Neu5Ac
Restoration of CD33 SAMP ligands
CD33 expression levels upregulated (compensation?)
Group B Strep engages CD33
N. gonorrhoeae engages CD33
Hominin microglia start expressing CD33
Decreased clearance of plaques?
Increased splicing into truncated CD33 form
Truncated CD33 diverted to peroxisomes
Decrease in full-length CD33
Escape from Neu5Ac-coated CD33-engaging pathogens?
Better plaque clearance?
Protection From LOAD
Selection for LOAD-protective CD33 alleles
Protection of cognitive functions of grandmothers
Appearance of LOAD?
Emergence of grandmothers
Earlier birth - more helpless young, cooperative breeding
Increase in brain size
B
Sialic acid recognition by Siglecs and exploitation by bacteria
Body Cell
Bacteria expressing Sia
Siglecs
Sialic Acids (Sia)
Glycocalyx
Activation
Inhibition
Inhibition
Immune Cell
C
Chimpanzee
Human
Functional CMAH, chimpanzee CD33 binds Neu5Gc (•)
Loss of CMAH, human CD33 binds Neu5Ac (•)
Mammalian brains only have Neu5Ac sialic acids
Microglia express CD33 and bind Neu5Ac in the brain
Inhibition
CD33 exploitation by pathogens
Inhibition
No Inhibition
CD33 truncated variant selected by evading pathogen manipulation
more active microglia
CD33 truncated variant protects against LOAD during prolonged post-reproductive survival

Fig. 6. Scenario for evolution of huCD33 in relationship to cell surface sialic acids, infectious disease, brain microglia, and cognitive maintenance of grandmothers and other elderly caregivers. (*A*) This schematic presentation combines the known/likely facts (white boxes, including data from this manuscript) as well as suggested possibilities (gray boxes) into the most likely evolutionary scenario for human-specific evolution of CD33. Starting from the left, the likely chronological order of occurrence is indicated (by arrowheads) with the approximate timeline on the top, along the dotted lines. "?" indicates our reasonable assumption leading to the event. (*B*) Pictorial summary of the Siglec interactions with sialylated bacteria and the effect in the immune modulation. (*C*) Occurrence of key evolutionary events presented in the chronological order based on the currently available data. See text for further discussion.

ecological, social, and technocultural, or more likely complex combinations thereof, including important contributions from language, a diagnostic feature of our species without clearly determined age of appearance. This also resulted in more helpless young and would have greatly benefited from cooperative breeding and caregiving (allomothering). It has been proposed that the origin of grandmothers may date back to the beginning of the genus *Homo* and may have directly contributed to altered anatomy and life history in our genus, long before the advent of *Homo sapiens*. One important effect of grandmother subsidies for the young could have been its contribution to the reduction in inter-birth interval, characteristic of our species, as compared with living great apes. Another great impact of allomothering by grandmothers and postreproductive elders would have been the transmission of knowledge and norms, which crucially rely on the capacity for language in our species and remains without demonstrated counterpart in any other species (fig. 6C). Importantly, both types of contribution would also have helped reduce extrinsic mortality, which would have in turn led to further increases in longevity. However, the value of postmenopausal grandmothers and other elderly caregivers would then have been seriously blunted by the appearance of LOAD or any other cognitive decline. The synthesis of the truncated isoform of CD33 protects from *Neisseria* during reproductive age and a higher ratio of truncated to full-length isoforms correlates to a decrease of LOAD in grandmothers. However, a small amount of the full-length isoform remains, likely to downregulate hyper-inflammation that might arise during prolonged absence of SAMP recognition. Notably, when an elderly caregiver gets LOAD, not only are the evolutionary benefits of the individual lost, but this also presents an increased burden on younger group members to care for that elder individual. Under this proposed scenario, the current state in the evolution of huCD33 protein represents a trade-off between the evolutionary response to exploitation by pathogens in early life and cognitive maintenance in postreproductive late life. A similar evolutionary scenario appears to underly the case of the human *APOE* gene where variants include both risk alleles (*APOE4*) and protective alleles (*APOE2* and *APOE3*) for cardiovascular disease and LOAD (Reiman et al. 2020). In this instance, the ancestral *APOE4* allele is associated with increased risks of loss of cognitive functions and the derived alleles may serve to protect the cognition of the elderly caregivers. Interestingly, the *APOE4* allele is also correlated with the protection from severe diarrhea in early years of life (Oriá et al. 2005). Given these examples, it remains to be seen how widespread this pattern is wherein variants conveying survival advantages in early life coexist with other variants that protect cognition late in life, a case of coexisting variants with opposite patterns of antagonistic pleiotropy.

Materials and Methods

Bacterial Culture and Cell Lines

The bacterial strains were *N. gonorrhoeae* F62Δ*lgtD* (generous gift from Sanjay Ram, University of Massachusetts Worcester), GBS strains (generous gifts from Victor Nizet, University of California San Diego). *Neisseria* were grown overnight on chocolate II agar and GBS on Todd Hewitt agar at 37 °C and 5% CO_2 from respective frozen glycerol stocks. Before assay, GBS was grown in Todd Hewitt broth. The *E. coli* K1 strain was grown in LB. HEK293A or U937 cells in DMEM or RPMI1640 media with 10% FCS, respectively, were used.

Sialylation of *Neisseria*

Following overnight growth, *Neisseria* were grown in GC broth with IsoVitaleX, with or without 30 μM CMP-Neu5Ac (Nacalai USA. Inc.) until OD600 of 0.4–0.5.

Bacterial Staining

The bacteria were washed with prewarmed Hanks′ Balanced Salt solution (HBSS) and stained with 2 μM SYTO13 (Thermo Scientific) for 30 min at 37 °C and shaking at 200 rpm in dark. The stained bacteria were washed with HBSS and resuspended to a final concentration of $OD_{600} = 1$/ml in HBSS for the binding assay.

Generation of CD33 Mutant Proteins

A genomic fragment (1228 bp) of human or chimpanzee CD33(M), including the first four exons (two Ig domains) was fused with pcDNA3.1(−) containing C-terminal FLAG(EK) sequence followed by hIgG1-Fc genomic fragment (hinge+2 Ig-like domains) and described elsewhere (Angata and Varki 2000; Angata et al. 2002). The mutant variants were made from either construct using Q5 site-directed mutagenesis kit (NEB) according to the manufacturer's instructions (supplementary table S3, Supplementary Material online). Mutagenesis primers listed were designed using NEBaseChanger software.

Truncated CD33_EK_Fc Construction

Total mRNA was isolated from U937 cells using Oligotex Direct mRNA Mini Kit (QIAGEN). CD33m was PCR amplified by SuperScript III One-Step RT-PCR (Invitrogen) and primers 5′-TTATAT*GCTAGC*GCCACCATGCCGCTGCTGCTACTGCTGC-3′, NheI site underlined and 5′-GCGCGC*GATATC*ATGAACCACTCCTGCTCTGGTCTCTTG-3′, EcoRV site underlined. PCR products were run on 2% agarose gel and 396 bp band corresponding to CD33m excised and cut with NheI/EcoRV. Digested bands were subcloned into pcDNA3.1(−) containing a C-terminal FLAG (EK) sequence followed by a hIgG1-Fc genomic fragment (hinge + 2 Ig-like domains).

Purification of CD33 Mutants

Transfection supernatants were spun down at 500 *g* for 5 min to remove cellular debris. The pH of the supernatant was adjusted to 8.0 for optimal binding of protein A-sepharose beads to hIgG-Fc fusion proteins. Protein A-Sepharose 4 Fast Flow suspension (GE Healthcare) was washed with Tris-Buffered Saline (TBS) pH 8.0, and 1:500 ratio of beads:media added to the supernatant and incubated for 24 h at 4 °C. Thereafter, supernatants with beads were transferred to disposable columns until all liquid has run thru. Beads were washed thrice with TBS and proteins eluted directly in 0.3 ml of 1 M Tris–HCl pH 8.0 using 0.1 M Glycine Buffer pH 2.8. Each eluate was centrifuged in Amicon Ultra-15 filter unit—Molecular Weight Cut-Off (MWCO) 30 K for full-length variants (MWCO 10 K for CD33m) and washed with TBS. After the final wash, each retentate was recovered from the column and stored at −80 °C.

Binding Assay

Bacterial binding was done with recombinant Fc-chimeric proteins of CD33. Briefly, protein A coated black 96-well plate (Pierce, Thermo Scientific) was washed with TBS containing 0.05% Tween 20 and coated with 200 ng/well of the respective CD33 protein. SYTO13 (Thermo Scientific) stained bacteria were incubated in the wells for 30 min at 37 °C and 5% CO_2 without shaking. Fluorescence measurements to indicate bound bacteria were done at 488 nm (excitation) and 530 nm (emission). Data were analyzed using excel and Prism software.

Evolutionary Analysis and Detection of Positive Selection

CD33 protein-coding sequences were aligned in CLUSTALW program implemented in MEGA7 and back translated to obtain a codon alignment. The phylogenetic tree of CD33 protein-coding sequences was reconstructed with neighbor-joining method, which was implemented in MEGA7 (fig. 1), 1000 bootstrap replicates (Kumar et al. 2016). The unrooted neighbor-joining tree was used for the subsequent analysis.

To assess the significance of the relative absence of LOAD-protective SNPS in archaic genomes, a resampling analysis was performed from SNPS in the 1000 genomes variant call set. Variants were filtered based on archaic quality filters and grouped by global allele frequencies before randomly selecting SNPS based on the query AF. The frequency of observing these random SNPS in archaic genomes was used to estimate an empirical probability.

Nonsynonymous/synonymous substitution ratios ($\omega =$ dN/dS, or Ka/Ks) have become a useful means for quantifying the impact of natural selection on molecular evolution. In general, the ratio $\omega =$ dN/dS is less than one if the gene is undergoing purifying selection, equal to one if the gene is evolving neutrally, and greater than one if positive selection has accelerated the fixation of nonsynonymous substitutions that resulted in amino acid changes. The pair-wise computation of Ka/Ks between V-set exon of each species was performed using the program DnaSp v.0 6.0. The initial unrooted tree fed to the program in the format of Newick was: ([Chimpanzee5:0.00000000, Bonobo:0.00000000]:0.00222522, Gorilla:0.00979959, Human:0.01351877).

Molecular Simulation

Starting structures of the V-type domain (residues 18–142) of huCD33 were built based on PDB entries 5j0b (A chain) and 6d49. The initial 3D models of chCD33 were built by swapping residues: N20K, F21I, W22R, A65P, I67V, R69G (in 6d49), L78P, P96L. The side chain of His45 was modeled in two conformations (compare fig. 4): 'down' (as in PDB entry 6d49) and 'up' (as present in PDB entries 5ihb or 5j06 chains A). Additionally, systems were built that contain five molecules of Neu5GcαOMe or Neu5AcαOMe distributed in the simulation box which allowed to simulate binding events. In total, 27 MD trajectories were sampled for huCD33 and 20 for chCD33, most of them covering a microsecond timescale (compare supplementary fig. S4, Supplementary Material online).

Sialoglycan Microarray

The sialoglycan microarray experimental method was adopted from the literature reported earlier (Meng et al. 2018; Lu et al. 2019). Sialoglycans were printed in quadruplets on NHS-functionalized glass slides (PolyAn 3D-NHS; catalog# PO-10400401) and then blocked, washed, and dried. The slides were rehydrated using Ovalbumin-PBS (1%, w/v) and then incubated with CD33 proteins followed by probing with secondary Cy3-conjugated goat antihuman IgG. The slides were scanned at wavelength 532 nm and heatmap was plotted.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We are very thankful to Dr Sanjay Ram (University of Massachusetts, Worcester) and Dr Victor Nizet (University of California, San Diego) for the bacterial strains that have greatly benefited the project. This work was supported by National Institute of Health grant R01GM32373 and Cure Alzheimer's Fund grant (to A.V.).

Data Availability

The data for the resampling analysis are available on Code Ocean (DOI:10.24433/CO.9371212.v1; https://codeocean.com/capsule/2131051/tree).

References

Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl M-C, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, *et al.* 2000. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet.* **26**:76–80.

Angata T, Kerr SC, Greaves DR, Varki NM, Crocker PR, Varki A. 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, Varki A. 2000. Cloning, characterization, and phylogenetic analysis of siglec-9, a new member of the CD33-related group of siglecs. Evidence for co-evolution with sialic acid synthesis pathways. *J Biol Chem.* **275**:22127–22135.

Apicella MA, Mandrell RE, Shero M, Wilson ME, Griffiss JM, Brooks GF, Lammel C, Breen JF, Rice PA. 1990. Modification by sialic acid of *Neisseria gonorrhoeae* lipooligosaccharide epitope expression in human urethral exudates: an immunoelectron microscopic analysis. *J Infect Dis.* **162**:506–512.

Bornhöfft KF, Goldammer T, Rebl A, Galuska SP. 2018. Siglecs: a journey through the evolution of sialic acid-binding immunoglobulin-type lectins. *Dev Comp Immunol.* **86**:219–231.

Bradshaw EM, Chibnik LB, Keenan BT, Ottoboni L, Raj T, Tang A, Rosenkrantz LL, Imboywa S, Lee M, Von Korff A, *et al.* 2013. CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. *Nat Neurosci.* **16**:848–850.

Brinkman-Van der Linden EC, Angata T, Reynolds SA, Powell LD, Hedrick SM, Varki A. 2003. CD33/Siglec-3 binding specificity, expression pattern, and consequences of gene deletion in mice. *Mol Cell Biol*. **23**:4199–4206.

Byars SG, Voskarides K. 2020. Antagonistic pleiotropy in human disease. *J Mol Evol*. **88**:12–25.

Cant MA, Croft DP. 2019. Life-history evolution: grandmothering in space and time. *Curr Biol*. **29**:R215–R218.

Carlin AF, Lewis AL, Varki A, Nizet V. 2007. Group B streptococcal capsular sialic acids interact with siglecs (immunoglobulin-like lectins) on human leukocytes. *J Bacteriol*. **189**:1231–1237.

Carrasquillo MM, Belbin O, Hunter TA, Ma L, Bisceglio GD, Zou F, Crook JE, Pankratz VS, Sando SB, Aasly JO, *et al.* 2011. Replication of EPHA1 and CD33 associations with late-onset Alzheimer's disease: a multicentre case-control study. *Mol Neurodegener*. **6**:54.

Caugant DA, Brynildsrud OB. 2020. Neisseria meningitidis: using genomics to understand diversity, evolution and pathogenesis. *Nat Rev Microbiol*. **18**:84–96.

Charnov EL, Turner TF, Winemiller KO. 2001. Reproductive constraints and the evolution of life histories with indeterminate growth. *Proc Natl Acad Sci U S A*. **98**:9460–9464.

Colonna M, Butovsky O. 2017. Microglia function in the central nervous system during health and neurodegeneration. *Annu Rev Immunol*. **35**:441–468.

Consortium GP, Auton A, Brooks LD, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. 2015. A global reference for human genetic variation. *Nature* **526**:68–74.

de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, Hernandez-Rodriguez J, Dupanloup I, Lao O, Hallast P, *et al.* 2016. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* **354**:477–481.

Edler MK, Munger EL, Meindl RS, Hopkins WD, Ely JJ, Erwin JM, Mufson EJ, Hof PR, Sherwood CC, Raghanti MA. 2020. Neuron loss associated with age but not Alzheimer's disease pathology in the chimpanzee brain. *Philos Trans R Soc Lond B Biol Sci*. **375**:20190619.

Edler MK, Sherwood CC, Meindl RS, Hopkins WD, Ely JJ, Erwin JM, Mufson EJ, Hof PR, Raghanti MA. 2017. Aged chimpanzees exhibit pathologic hallmarks of Alzheimer's disease. *Neurobiol Aging*. **59**:107–120.

Edwards JL, Apicella MA. 2004. The molecular mechanisms used by *Neisseria gonorrhoeae* to initiate infection differ between men and women. *Clin Microbiol Rev*. **17**:965–981.

Eikelenboom P, van Gool WA. 2004. Neuroinflammatory perspectives on the two faces of Alzheimer's disease. *J Neural Transm (Vienna)*. **111**:281–294.

Eikelenboom P, Veerhuis R, Scheper W, Rozemuller AJ, van Gool WA, Hoozemans JJ. 2006. The significance of neuroinflammation in understanding Alzheimer's disease. *J Neural Transm (Vienna)*. **113**:1685–1695.

Finch CE, Sapolsky RM. 1999. The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms. *Neurobiol Aging*. **20**: 407–428.

Fong JJ, Tsai CM, Saha S, Nizet V, Varki A, Bui JD. 2018. Siglec-7 engagement by GBS β-protein suppresses pyroptotic cell death of natural killer cells. *Proc Natl Acad Sci U S A*. **115**:10410–10415.

Freeman SD, Kelm S, Barber EK, Crocker PR. 1995. Characterization of CD33 as a new member of the sialoadhesin family of cellular interaction molecules. *Blood*. **85**:2005–2012.

Gonzalez-Gil A, Porell RN, Fernandes SM, Maenpaa E, Li TA, Li T, Wong PC, Aoki K, Tiemeyer M, Yu ZJ, *et al.* 2022. Human brain sialoglycan ligand for CD33, a microglial inhibitory Siglec implicated in Alzheimer's disease. *J. Biol Chem*. **298**(6):101960.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, *et al.* 2010. A draft sequence of the Neandertal genome. *Science* **328**:710–722.

Griciuc A, Federico AN, Natasan J, Forte AM, McGinty D, Nguyen H, Volak A, LeRoy S, Gandhi S, Lerner EP, *et al.* 2020. Gene therapy for Alzheimer's disease targeting CD33 reduces amyloid beta accumulation and neuroinflammation. *Hum Mol Genet*. **29**: 2920–2935.

Griciuc A, Serrano-Pozo A, Parrado AR, Lesinski AN, Asselin CN, Mullin K, Hooli B, Choi SH, Hyman BT, Tanzi RE. 2013. Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron* **78**:631–643.

Hawkes K. 2004. Human longevity: the grandmother effect. *Nature* **428**:128–129.

Hawkes K. 2010. Colloquium paper: how grandmother effects plus individual variation in frailty shape fertility and mortality: guidance from human–chimpanzee comparisons. *Proc Natl Acad Sci U S A*. **107**(Suppl 2):8977–8984.

Hawkes K. 2016. Genomic evidence for the evolution of human postmenopausal longevity. *Proc Natl Acad Sci U S A*. **113**:17–18.

Hawkes K, O'Connell JF, Jones NG, Alvarez H, Charnov EL. 1998. Grandmothering, menopause, and the evolution of human life histories. *Proc Natl Acad Sci U S A*. **95**:1336–1339.

Helgason A, Pálsson S, Thorleifsson G, Emilsson V, Gunnarsdottir S, Adeyemo A, Chen Y, Chen G, Reynisdottir I, *et al.* 2007. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat Genet*. **39**:218–225.

Hernandez-Caselles T, Martinez-Esparza M, Perez-Oliva AB, Quintanilla-Cecconi AM, Garcia-Alonso A, Alvarez-Lopez DM, Garcia-Penarrubia P. 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.

Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskvina V, *et al.* 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet*. **43**:429–435.

Iadecola C. 2013. The pathobiology of vascular dementia. *Neuron* **80**: 844–866.

Johnstone RA, Cant MA. 2010. The evolution of menopause in cetaceans and humans: the role of demography. *Proc Biol Sci*. **277**: 3765–3771.

Khan N, de Manuel M, Peyregne S, Do R, Prufer K, Marques-Bonet T, Varki N, Gagneux P, Varki A. 2020. Multiple genomic events altering hominin SIGLEC biology and innate immunity predated the common ancestor of humans and archaic hominins. *Genome Biol Evol*. **12**:1040–1050.

Khan N, Kim SK, Gagneux P, Dugan LL, Varki A. 2020. Maximum reproductive lifespan correlates with CD33rSIGLEC gene number: implications for NADPH oxidase-derived reactive oxygen species in aging. *FASEB J*. **34**:1928–1938.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. **33**:1870–1874.

Lamba JK, Pounds S, Cao X, Downing JR, Campana D, Ribeiro RC, Pui CH, Rubnitz JE. 2009. Coding polymorphisms in CD33 and response to gemtuzumab ozogamicin in pediatric patients with AML: a pilot study. *Leukemia* **23**:402–404.

Landig CS, Hazel A, Kellman BP, Fong JJ, Schwarz F, Agarwal S, Varki N, Massari P, Lewis NE, Ram S, *et al.* 2019. Evolution of the exclusively human pathogen *Neisseria gonorrhoeae*: human-specific engagement of immunoregulatory Siglecs. *Evol Appl*. **12**:337–349.

Läubli H, Varki A. 2020. Sialic acid-binding immunoglobulin-like lectins (Siglecs) detect self-associated molecular patterns to regulate immune responses. *Cell Mol Life Sci*. **77**:593–605.

Liu Z, Dai X, Tao W, Liu H, Li H, Yang C, Zhang J, Li X, Chen Y, Ma C, *et al.* 2018. APOE influences working memory in non-demented elderly through an interaction with SPON1 rs2618516. *Hum Brain Mapp*. **39**:2859–2867.

Lu N, Ye J, Cheng J, Liu C-C, Yao W, Yan J, Khan N, Yi W, Varki A, Cao H. 2019. Redox-controlled site-specific α2-6-sialylation. *J Am Chem Soc*. **141**:4547–4552.

Malik M, Simpson JF, Parikh I, Wilfred BR, Fardo DW, Nelson PT, Estus S. 2013. CD33 Alzheimer's risk-altering polymorphism, CD33 expression, and exon 2 splicing. *J Neurosci*. **33**: 13320–13325.

Mendez FL, Poznik GD, Castellano S, Bustamante CD. 2016. The divergence of neandertal and modern human Y chromosomes. *Am J Hum Genet*. **98**:728–734.

Meng C, Sasmal A, Zhang Y, Gao T, Liu CC, Khan N, Varki A, Wang F, Cao H. 2018. Chemoenzymatic assembly of mammalian O-mannose glycans. *Angew Chem Int Ed Engl*. **57**:9003–9007.

Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, *et al.* 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**:222–226.

Moon JM, Aronoff DM, Capra JA, Abbot P, Rokas A. 2018. Examination of signatures of recent positive selection on genes involved in human sialic acid biology. *G3 (Bethesda)* **8**:1315–1325.

Munshi MN. 2017. Cognitive dysfunction in older adults with diabetes: what a clinician needs to know. *Diabetes Care* **40**:461–467.

Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, Gallins PJ, Buxbaum JD, Jarvik GP, *et al.* 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet*. **43**:436–441.

Nakajima T, Wooding S, Sakagami T, Emi M, Tokunaga K, Tamiya G, Ishigami T, Umemura S, Munkhbat B, Jin F, *et al.* 2004. Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. *Am J Hum Genet*. **74**:898–916.

Oriá RB, Patrick PD, Zhang H, Lorntz B, de Castro Costa CM, Brito GA, Barrett LJ, Lima AA, Guerrant RL. 2005. APOE4 protects the cognitive development in children with heavy diarrhea burdens in Northeast Brazil. *Pediatr Res*. **57**:310–316.

Padler-Karavani V, Hurtado-Ziola N, Chang YC, Sonnenburg JL, Ronagh A, Yu H, Verhagen A, Nizet V, Chen X, Varki N, *et al.* 2014. Rapid evolution of binding specificities and expression patterns of inhibitory CD33-related Siglecs in primates. *FASEB J*. **28**:1280–1293.

Parsons NJ, Patel PV, Tan EL, Andrade JR, Nairn CA, Goldner M, Cole JA, Smith H. 1988. Cytidine 5′-monophospho-N-acetyl neuraminic acid and a low molecular weight factor from human blood cells induce lipopolysaccharide alteration in gonococci when conferring resistance to killing by human serum. *Microb Pathog*. **5**:303–309.

Peila R, Rodriguez BL, Launer LJ, Honolulu-Asia AS. 2002. Type 2 diabetes, APOE gene, and the risk for dementia and related pathologies: the Honolulu-Asia Aging Study. *Diabetes* **51**:1256–1262.

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, *et al.* 2013. Great ape genetic diversity and population history. *Nature* **499**:471–475.

Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, *et al.* 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**:43–49.

Raj T, Ryan KJ, Replogle JM, Chibnik LB, Rosenkrantz L, Tang A, Rothamel K, Stranger BE, Bennett DA, Evans DA, *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.

Rathore N, Ramani SR, Pantua H, Payandeh J, Bhangale T, Wuster A, Kapoor M, Sun Y, Kapadia SB, Gonzalez L, *et al.* 2018. Paired immunoglobulin-like type 2 receptor alpha G78R variant alters ligand binding and confers protection to Alzheimer's disease. *PLoS Genet*. **14**:e1007427.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, *et al.* 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**:1053–1060.

Reiman EM, Arboleda-Velasquez JF, Quiroz YT, Huentelman MJ, Beach TG, Caselli RJ, Chen Y, Su Y, Myers AJ, Hardy J, *et al.* 2020. Exceptionally low likelihood of Alzheimer's dementia in APOE2 homozygotes from a 5,000-person neuropathological study. *Nat Commun*. **11**:667.

Saha S, Siddiqui SS, Khan N, Verhagen A, Jiang W, Springer S, Ghosh P, Varki A. 2019. Controversies about the subcellular localization and mechanisms of action of the Alzheimer's disease-protective CD33 splice variant. *Acta Neuropathol*. **138**:671–672.

Salminen A, Kaarniranta K. 2009. Siglec receptors and hiding plaques in Alzheimer's disease. *J Mol Med*. **87**:697–701.

Schwarz F, Springer SA, Altheide TK, Varki NM, Gagneux P, Varki A. 2016. Human-specific derived alleles of CD33 and other genes protect against postreproductive cognitive decline. *Proc Natl Acad Sci U S A*. **113**:74–79.

Seifert HS. 2019. Location, location, location-commensalism, damage and evolution of the pathogenic neisseria. *J Mol Biol*. **431**: 3010–3014.

Siddiqui SS, Springer SA, Verhagen A, Sundaramurthy V, Alisson-Silva F, Jiang W, Ghosh P, Varki A. 2017. The Alzheimer's disease-protective CD33 splice variant mediates adaptive loss of function via diversion to an intracellular pool. *J Biol Chem*. **292**:15312–15320.

Simen AA, Bordner KA, Martin MP, Moy LA, Barry LC. 2011. Cognitive dysfunction with aging and the role of inflammation. *Ther Adv Chronic Dis*. **2**:175–195.

Szumanska G, Vorbrodt AW, Mandybur TI, Wisniewski HM. 1987. Lectin histochemistry of plaques and tangles in Alzheimer's disease. *Acta Neuropathol*. **73**:1–11.

Tangestani Fard M, Stough C. 2019. A review and hypothesized model of the mechanisms that underpin the relationship between inflammation and cognition in the elderly. *Front Aging Neurosci*. **11**:56.

Thompson EE, Kuttab-Boulos H, Witonsky D, Yang L, Roe BA, Di Rienzo A. 2004. CYP3A variation and the evolution of salt-sensitivity variants. *Am J Hum Genet*. **75**:1059–1069.

Tortorici MA, Walls AC, Lang Y, Wang C, Li Z, Koerhuis D, Boons G-J, Bosch B-J, Rey FA, de Groot RJ, *et al.* 2019. Structural basis for human coronavirus attachment to sialic acid receptors. *Nat Struct Mol Biol*. **26**:481–489.

Tsai TY, Huang MT, Sung PS, Peng C-Y, Tao M-H, Yang H-I, Chang W-C, Yang A-S, Yu C-M, Lin Y-P, *et al.* 2021. SIGLEC-3 (CD33) serves as an immune checkpoint receptor for HBV infection. *J Clin Invest*. **131**:141965.

Vander Molen J, Frisse LM, Fullerton SM, Qian Y, Del Bosque-Plata L, Hudson RR, Di Rienzo A. 2005. Population genetics of CAPN10 and GPR35: implications for the evolution of type 2 diabetes variants. *Am J Hum Genet*. **76**:548–560.

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 lecins. *Glycobiology* **16**:1R–27R.

Varki A, Gagneux P. 2012. Multifarious roles of sialic acids in immunity. *Ann N Y Acad Sci*. **1253**:16–36.

Wen G, Wessel J, Zhou W, Rao F, Stridsberg M, Mahata SK, Gent PM, Das M, Cooper RS, Rao F, *et al.* 2007. An ancestral variant of Secretogranin II confers regulation by PHOX2 transcription factors and association with hypertension. *Hum Mol Genet*. **16**: 1752–1764.

Williams GC. 1957. Pleiotropy, natural selection, and the evolution of senescence. *Evolution* **11**:398–411.

Wißfeld J, Nozaki I, Mathews M, Raschka T, Ebeling C, Hornung V, Brüstle O, Neumann H. 2021. Deletion of Alzheimer's disease-associated CD33 results in an inflammatory human microglia phenotype. *Glia* **69**:1393–1412.

Xu W, Tan L, Yu JT. 2015. The role of PICALM in Alzheimer's disease. *Mol Neurobiol*. **52**:399–413.

Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, Frandsen P, Chen Y, Yngvadottir B, Cooper DN, *et al.* 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**:242–245.

Zhao L. 2019. CD33 in Alzheimer's disease – biology, pathogenesis, and therapeutics: a mini-review. *Gerontology* **65**:323–331.

Submission for Article (Discoveries Section)

Supplemental information for

Evolution of Human-specific Alleles Protecting Cognitive Function of Grandmothers

Sudeshna Saha[1], Naazneen Khan[1], Troy Comi[2], Andrea Verhagen[1], Aniruddha Sasmal[1],
Sandra Diaz[1], Hai Yu[3], Xi Chen[3], Joshua M. Akey[2], Martin Frank[4],
Pascal Gagneux[1*], and Ajit Varki[1*]

[1.] Departments of Medicine, Pathology, Anthropology and Cellular and Molecular Medicine, Center for Academic Research and Training in Anthropogeny and Glycobiology Research and Training Center, University of California San Diego, San Diego, California 92093, USA

[2.] Department of Genetics, Princeton University, Princeton, New Jersey 08544, USA

[3.] Department of Chemistry, University of California Davis, Davis, California 95616, USA

[4.] Biognos AB, Gothenburg, SE-402 74, Sweden

*Address correspondence to Ajit Varki: a1varki@ucsd.edu or Pascal Gagneux: pgagneux@ucsd.edu.

Detailed Materials and Methods

Bacterial culture and cell lines. The bacterial strains used were *Neisseria gonorrhoeae* F62Δ*lgtD* (generous gift from Sanjay Ram, University of Massachusetts Worcester), Group B *Streptococcus* (GBS) strains COH1wt, COH1Δ*neuA*, A909wt and A909Δ*neuA* (generous gifts from Victor Nizet, University of California San Diego). *Neisseria* were grown overnight on chocolate II agar plate and GBS on Todd Hewitt agar plate at 37 °C and 5% CO_2 from the respective frozen glycerol stocks. Prior to the assay, GBS was grown in Todd Hewitt broth at 37 °C and 5 % CO_2 without shaking. The *E. coli* K1 strain was grown in LB. For the CD33 protein purification, HEK293A cells were grown in DMEM media (Invitrogen) containing 10% FCS at 37 °C and 5 % CO_2.

Sialylation of *Neisseria*. Following overnight growth on chocolate agar plate, the bacteria were grown in GC broth supplemented with IsoVitaleX at 37 °C, 5% CO_2 and shaking at 200 rpm in presence or absence of 30 µM CMP-Neu5Ac (Nacalai USA. Inc.) until OD600 equivalent to 0.4–0.5.

Bacterial staining. Following appropriate growth, the bacteria were washed with pre-warmed HBSS and stained with 2 µM SYTO13 (Thermo Scientific) for 30 min at 37 °C and shaking at 200 rpm in dark. After incubation, the stained bacteria were washed with HBSS and resuspended to a final concentration of OD600 = 1/ml in HBSS for the binding assay.

Generation of CD33 mutant proteins. A genomic fragment (1228 bp) of human or Chimpanzee CD33(M), including the first 4 exons (2 Ig domains) was fused with pcDNA3.1(-) containing a C-terminal FLAG (EK) sequence followed by a hIgG1-Fc genomic fragment (hinge + 2 Ig-like domains) and described elsewhere (Angata and Varki, 2000; Angata et al., 2002). Sixteen mutant variants were made from either construct above using New England Biolabs Q5 site directed mutagenesis Kit according to the manufacturer's instructions (Supplemental Table 3). Mutagenesis primers listed were designed using NEBaseChanger software.

Truncated CD33(CD33m) _EK_Fc Construction: U937 cells were cultured in RPMI 1640 supplemented with 10% FCS. Total mRNA was isolated using Qiagens Oligotex Direct mRNA Mini Kit according to the manufacturer's instructions. CD33m was amplified by PCR using SuperScript III One-Step RT-PCR (Invitrogen) and Gene-specific primers 5'-TTATATGCTAGCGCCACCATGCCGCTGCTGCTACTGCTGC-3', NheI site underlined and 5'-GCGCGCGATATCATGAACCACTCCTGCTCTGGTCTCTTG-3', EcoRV site underlined. PCR products were run on 2% agarose gel and the 396 bp bands corresponding to CD33(m) were excised and cut with NheI/EcoRV restriction enzymes. Digested bands were sub-cloned into pcDNA3.1(-) containing a C-terminal FLAG (EK) sequence followed by a hIgG1-Fc genomic fragment (hinge + 2 Ig-like domains).

Purification of CD33 mutants. Transfection supernatants were collected and spun down at 500 g for 5 mins to remove cellular debris. The pH of each supernatant was adjusted to pH 8.0 for optimal binding of protein A-Sepharose beads to hIgG Fc fusion protein. Protein A-Sepharose 4 Fast Flow suspension (GE Healthcare) was washed with Tris-Buffered Saline (TBS) pH 8.0, and a 1:500 ratio of beads:media added to each supernatant. Each tube was subsequently incubated for 24 hrs on a roller in the cold-room. After 24 hours supernatants plus beads were transferred to disposable columns until all liquid has run thru. Beads were washed 3x with TBS pH 8.0 before being eluted directly in 0.3 ml of 1 M Tris-HCl pH 8.0 using 0.1 M Glycine Buffer pH 2.8. Each eluate was put into an Amicon Ultra-15 filter unit with MWCO 30 K for each full length CD33-EK_Fc variant and MWCO 10 K for huCD33m-EK_Fc. Tubes were centrifuged at 4,000 g for 20 mins. Run-through was discarded and the columns washed 3x with TBS pH 8.0. After the last wash, each retentate was recovered from the column and stored at -80^{o}C.

Binding assay with the bacteria. Bacterial binding with the CD33 proteins were done with the recombinant Fc-chimeric proteins of CD33. Briefly, protein A coated black 96-well plate (Pierce, Thermo Scientific) was washed thrice with TBS containing 0.05% Tween 20 (TBS-T) and coated with 200 ng/well of the respective CD33 protein diluted in 200 mM Tris-HCl pH 8.0, 150 mM NaCl and 1% BSA at 4 ℃ overnight. Following incubation, the coated plate was washed with 200 mM Tris pH 8.0, 150 mM NaCl to eliminate the unbound proteins. Stained bacteria equivalent to OD600 = 0.1 was added to each well of the plate and allowed to interact with the proteins for 30 min at 37 ℃ and 5% CO_2 without shaking. Following incubation, the plate was washed with TBS-T to eliminate any unbound bacteria and the residual fluorescence was measured upon excitation at 488 nm and emission at 530 nm. The data were analyzed using the excel and Prism software.

Evolutionary analysis and Detection of positive selection. The protein coding sequences of CD33 were aligned using CLUSTAL W program implemented in MEGA7 and then back translated to obtain a codon alignment. The phylogenetic tree of CD33 protein coding sequences were reconstructed with neighbor-joining method which was implemented in MEGA7 (Figure 1), 1000 bootstrap replicates (Kumar et al., 2016). The unrooted neighbor joining tree was used for the subsequent analysis.

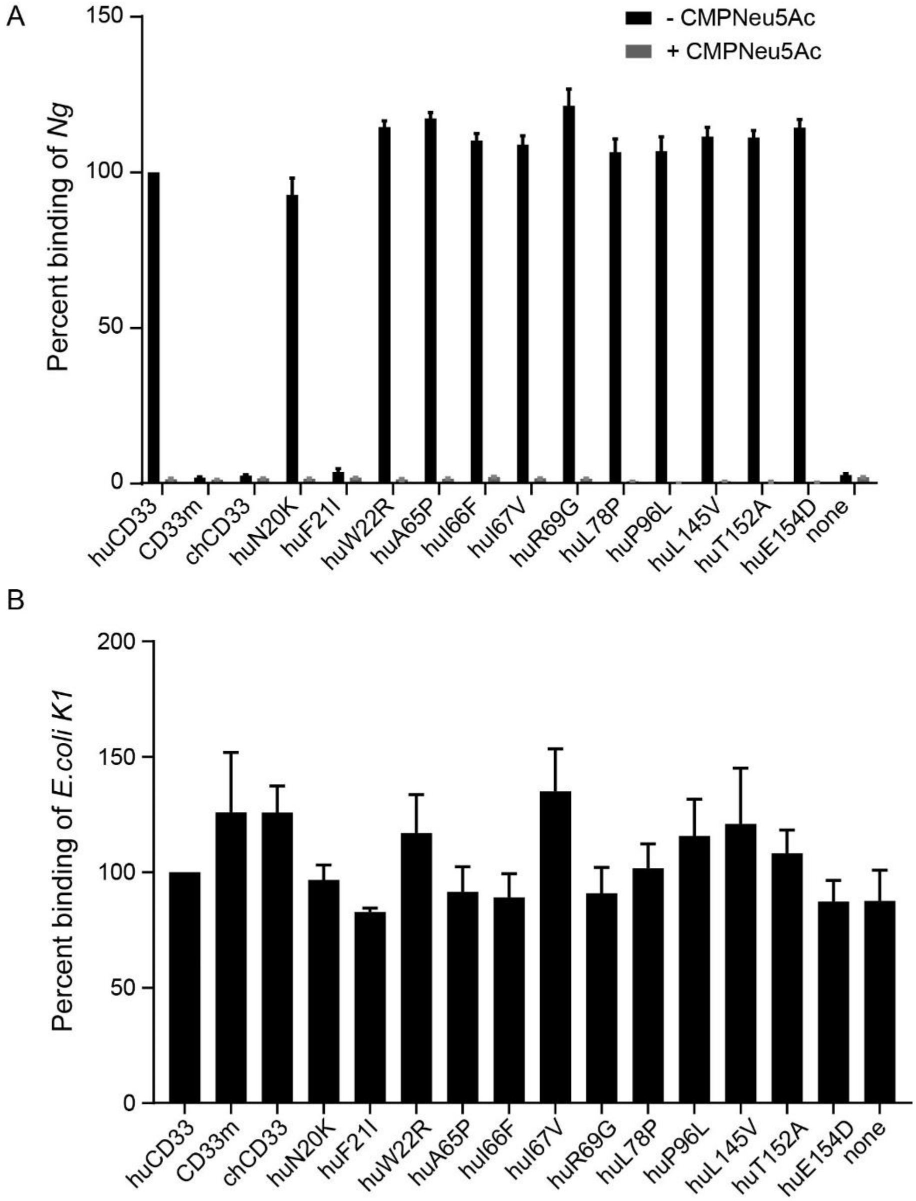
VCF files were accessed from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ for 1000 genomes project, http://cdna.eva.mpg.de/neandertal/altai/AltaiNeandertal/VCF/ for Altai Neanderthal and http://cdna.eva.mpg.de/denisova/VCF/hg19_1000g/ for Denisovan. Quality filters were obtained from https://bioinf.eva.mpg.de/altai_minimal_filters/ for Altai and Denisovan. Individuals in 1000 genomes datasets were assigned to populations using http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130606_sample_info/20130606_sample_info.txt. First, all vcf files were filtered by intersecting with the quality bed files using bedtools intersect (v2.28.0). The filtered vcf files were then combined, per chromosome, to match their position, reference and alternative allele using a custom python script. VCF information was retained along with per-population allele frequencies and archaic genotypes. Next, ancestral alleles were obtained from ensembl (https://rest.ensembl.org/variation/homo_sapiens) by querying each SNP id and appending to the joint vcf entries. The joint vcf files were used as input for further processing in a jupyter notebook to perform resampling analysis. Each SNP of interest is used to select collections of SNPs with matching global allele frequencies, +/- 0.004, or +/- 2 observed haplotypes. A single draw consists of selecting one SNP from each collection to produce a simulated observation and the number of SNPS with derived archaic haplotypes are recorded. After 10,000 such draws, the faction of draws with fewer or equal numbers of derived SNPs is used to produce a p-value estimate. The process is repeated 10,000 times to produce a histogram and provide a confidence estimate on the reported p values (+/- SEM). Methods to replicate the analysis can be found on Code Ocean (DOI:10.24433/CO.9371212.v1).

Non-synonymous/ synonymous substitution ratios (ω = dN/dS, or Ka/Ks) have become a useful means for quantifying the impact of natural selection on molecular evolution. In general, the ratio ω = dN/dS is less than one if the gene is undergoing purifying selection, equal to one if the gene is evolving neutrally, and greater than one if positive selection has accelerated the fixation of non-synonymous substitutions that resulted in amino acid changes. The pair-wise computation of Ka/Ks between V-set exon of each species were performed using the program DnaSp v.0 6.0. The initial unrooted tree fed to the program in the format of Newick was: ((Chimpanzee5:0.00000000, Bonobo:0.00000000):0.00222522, Gorilla:0.00979959, Human:0.01351877).

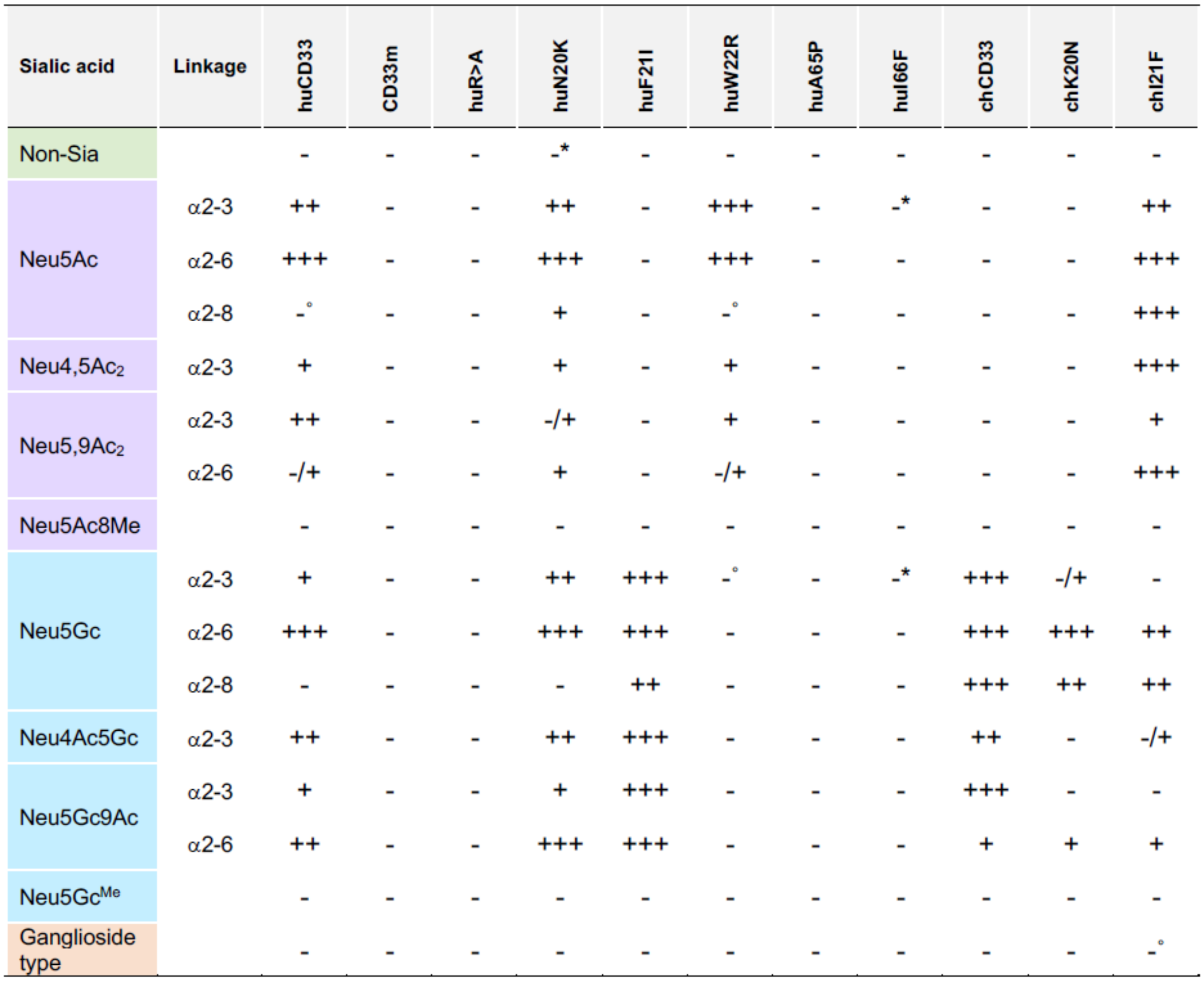
Molecular Simulation. Starting structures of the V-type domain (residues 18-142) of CD33 were built based on PDB entries 5j0b (A chain) and 6d49 using the graphical interface of YASARA (Krieger and Vriend, 2014). The two structures differ significantly with respect to the conformation of the C-C’ loop (residues 63-71, compare Supplemental Figure S3). A single mutation (G69R) was introduced into 5j0b to build CD33(human). The initial 3D models of chCD33 were built by swapping residues: N20K, F21I, W22R, A65P, I67V, R69G (in 6d49), L78P, P96L. An N-glycan core (M3) was attached to Asn100. The side chain of His45 was modeled in two conformations (compare Figure 4): ‘down’ (as in PDB entry 6d49) and ‘up’ (as present in PDB entries 5ihb or 5j06 chains A). The systems were solvated in 0.9% NaCl solution (0.15 M) and simulations were performed at 310 K using periodic boundary conditions. The box size was rescaled dynamically to maintain a water density of 0.996 g/ml. Additionally systems were built that contain five molecules of Neu5GcαOMe or Neu5AcαOMe distributed in the simulation box which allowed to simulate binding events. Simulations were performed using YASARA with GPU acceleration (Krieger and Vriend, 2015). In total 27 MD trajectories were sampled for huCD33 and 20 for chCD33, most of them covering a microsecond timescale (compare Supplemental Figure S4). Conformational Analysis Tools (CAT, http://www.md-simulations.de/CAT/) was used for analysis of trajectory data, general data processing and generation of scientific plots. VMD (Humphrey et al., 1996) was used to generate molecular graphics.

Sialoglycan microarray. The sialoglycan microarray experimental method was adopted from the literature reported earlier (Meng et al., 2018; Lu et al., 2019). Chemoenzymatically synthesized sialoglycans were quantitated utilizing DMB-HPLC method (Ji et al., 2021) and 10 mM aqueous stock solutions were prepared. Next, the glycans were diluted to 100 µM in 300 mM Na-phosphate buffer (pH 8.4) and printed in quadruplets on NHS-functionalized glass slides (PolyAn 3D-NHS; catalog# PO-10400401) using an ArrayIt SpotBot® Extreme instrument. The slides were blocked using 0.05M ethanolamine solution in 0.1 M Tris-HCl (pH 9.0), washed with warm Milli-Q water and dried. Printed slides were fitted in a multi-well microarray hybridization cassette (ArrayIt, CA) and rehydrated using 400 µl of ovalbumin (1% w/v, PBS) for one hour in a humid chamber with gentle shaking. The solution was discarded followed by the addition of 400 µl solution of the CD33 protein (30 µg/ml in PBS with 1% w/v ovalbumin) in the individual well. The slides were incubated for 2 h at ambient temperature with gentle shaking followed by washing with PBS-Tween (0.1% v/v) and PBS. The wells were then treated with Cy3-conjugated goat anti-human IgG (1:500 dilution in PBS), incubated for 1h in a dark humid chamber with gentle shaking. After washing and drying, the slides were scanned using a Genepix 4000B scanner (Molecular Devices

Corp., Union City, CA) at wavelength 532 nm. Data analysis was performed using the Genepix Pro 7.3 software (Molecular Devices Corp., Union City, CA).

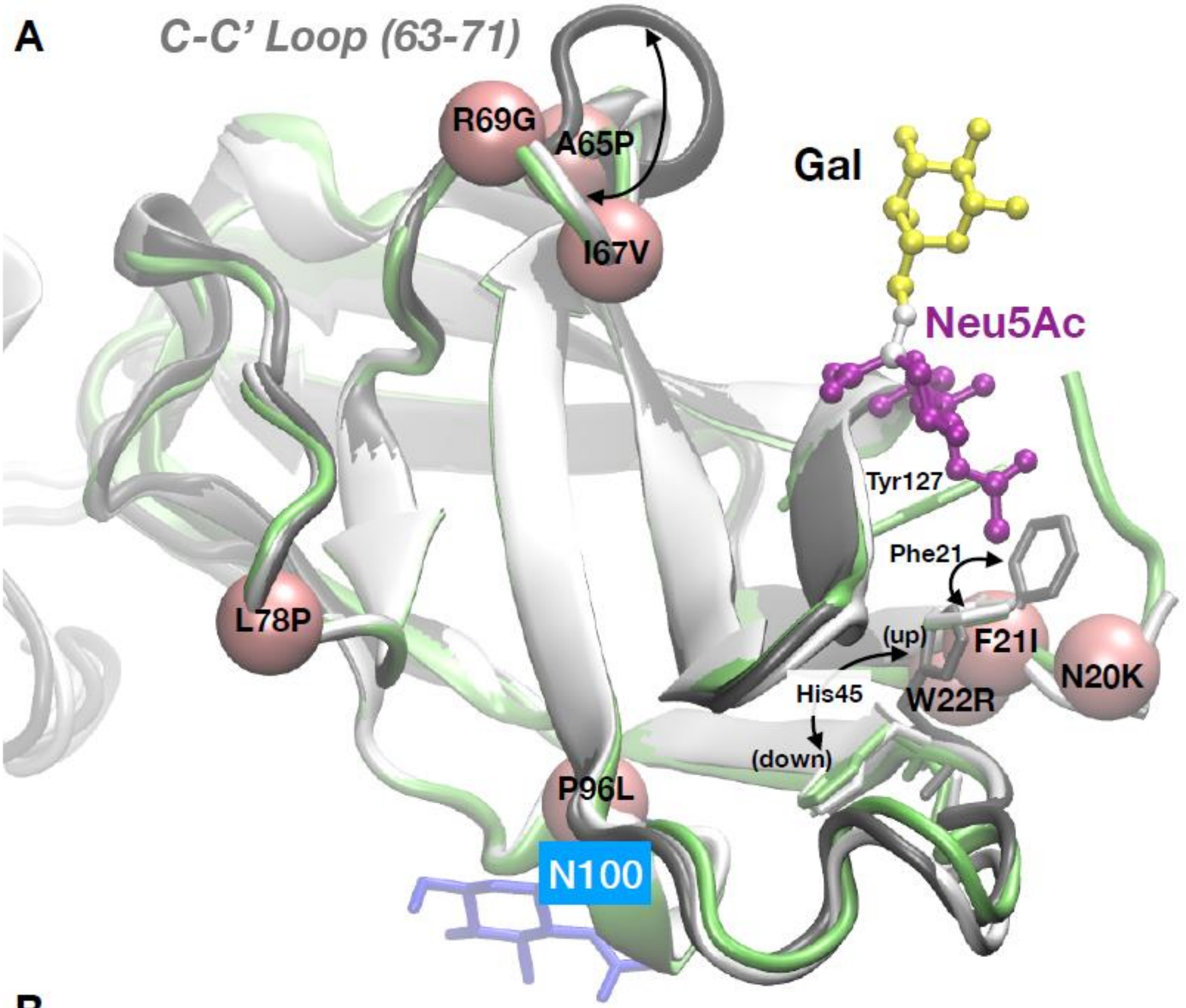


Supplemental Figure S1: While *E. coli* does not bind CD33, human CD33 binding by *Neisseria* is Sia-dependent. (A) Binding of Ng with wildtype or mutant CD33 proteins was determined in the same manner as in Figure 2A. The bacteria for the assay were either grown in presence (+) or absence (-) of exogenous CMP-Neu5Ac as indicated in the legend. All the binding was normalized to wildtype human CD33 binding. Cumulative data from 2 independent experiments, each done in triplet is presented. **(B)** Binding of *E. coli* K1 was determined using the different CD33 proteins. None of the proteins showed any increased binding to the bacteria relative to no protein (control) containing blank well, indicating that there is no binding of the bacteria with the protein.



Sialic acid	Linkage	huCD33	CD33m	huR>A	huN20K	huF21I	huW22R	huA65P	huI66F	chCD33	chK20N	chI21F
Non-Sia		-	-	-	-*	-	-	-	-	-	-	-
Neu5Ac	α2-3	++	-	-	++	-	+++	-	-*	-	-	++
	α2-6	+++	-	-	+++	-	+++	-	-	-	-	+++
	α2-8	-°	-	-	+	-	-°	-	-	-	-	+++
Neu4,5Ac$_2$	α2-3	+	-	-	+	-	+	-	-	-	-	+++
Neu5,9Ac$_2$	α2-3	++	-	-	-/+	-	+	-	-	-	-	+
	α2-6	-/+	-	-	+	-	-/+	-	-	-	-	+++
Neu5Ac8Me		-	-	-	-	-	-	-	-	-	-	-
Neu5Gc	α2-3	+	-	-	++	+++	-°	-	-*	+++	-/+	-
	α2-6	+++	-	-	+++	+++	-	-	-	+++	+++	++
	α2-8	-	-	-	-	++	-	-	-	+++	++	++
Neu4Ac5Gc	α2-3	++	-	-	++	+++	-	-	-	++	-	-/+
Neu5Gc9Ac	α2-3	+	-	-	+	+++	-	-	-	+++	-	-
	α2-6	++	-	-	+++	+++	-	-	-	+	+	+
Neu5Gc^{Me}		-	-	-	-	-	-	-	-	-	-	-
Ganglioside type		-	-	-	-	-	-	-	-	-	-	-°

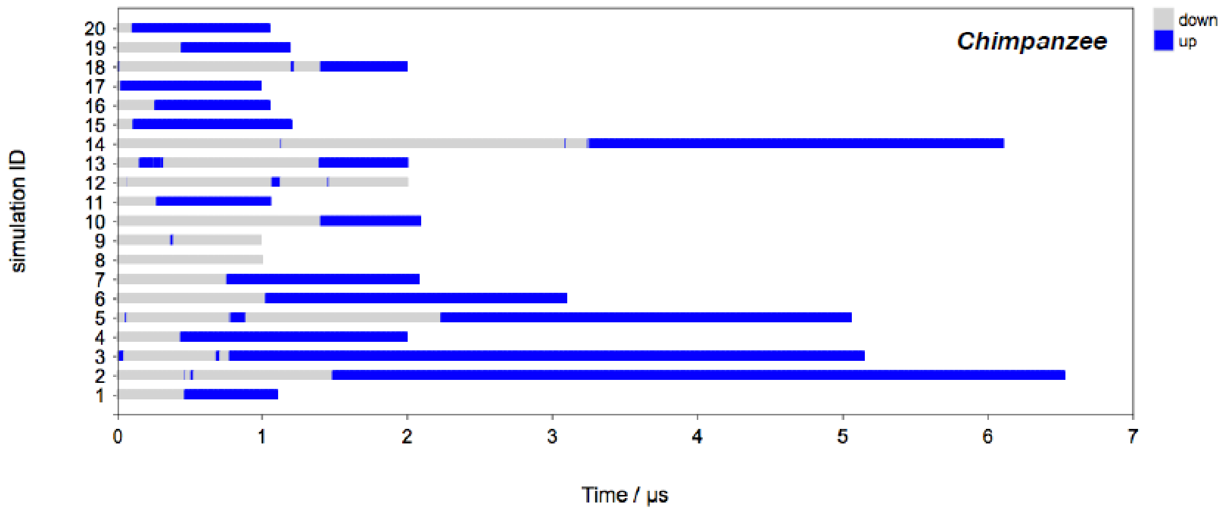
Supplemental Figure S2: Summarized result of CD33 sialoglycan binding. The results of the sialoglycan microarray binding of different wildtype and mutant CD33 proteins presented in Figure 3 are summarized here. A differential sialoglycan binding preference was observed when wildtype and mutant human/chimpanzee CD33 proteins were tested on the microarray. Binding is annotated with a positive (+) symbol and the strength of the binding is indicated by the number of the symbols. +++ indicates a very strong binding. Negative (-) symbol implies non-binding and -/+ indicates very faint interaction. In some cases, only a few sulfated glycans showed strong binding signal (indicated with asterisk). Degree (˚) symbols indicate binding with a very few numbers of glycans only. Linkage indicates the nature of the glycosidic bond of the terminal Sia to the underlying glycan.

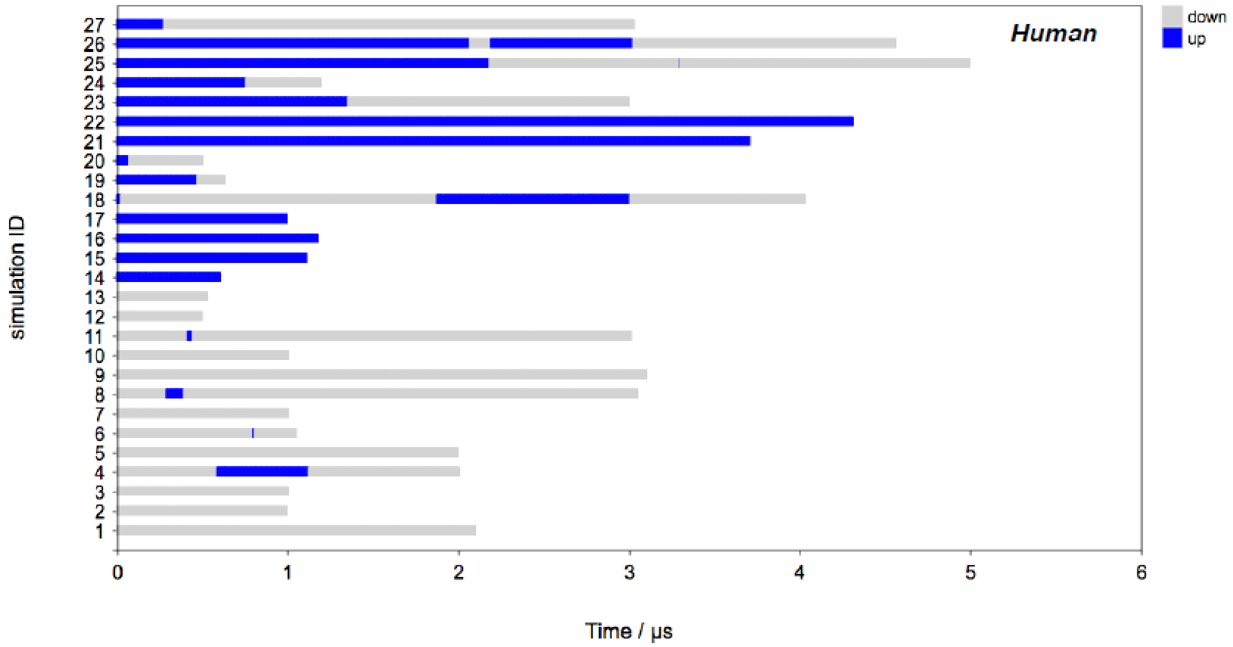


B

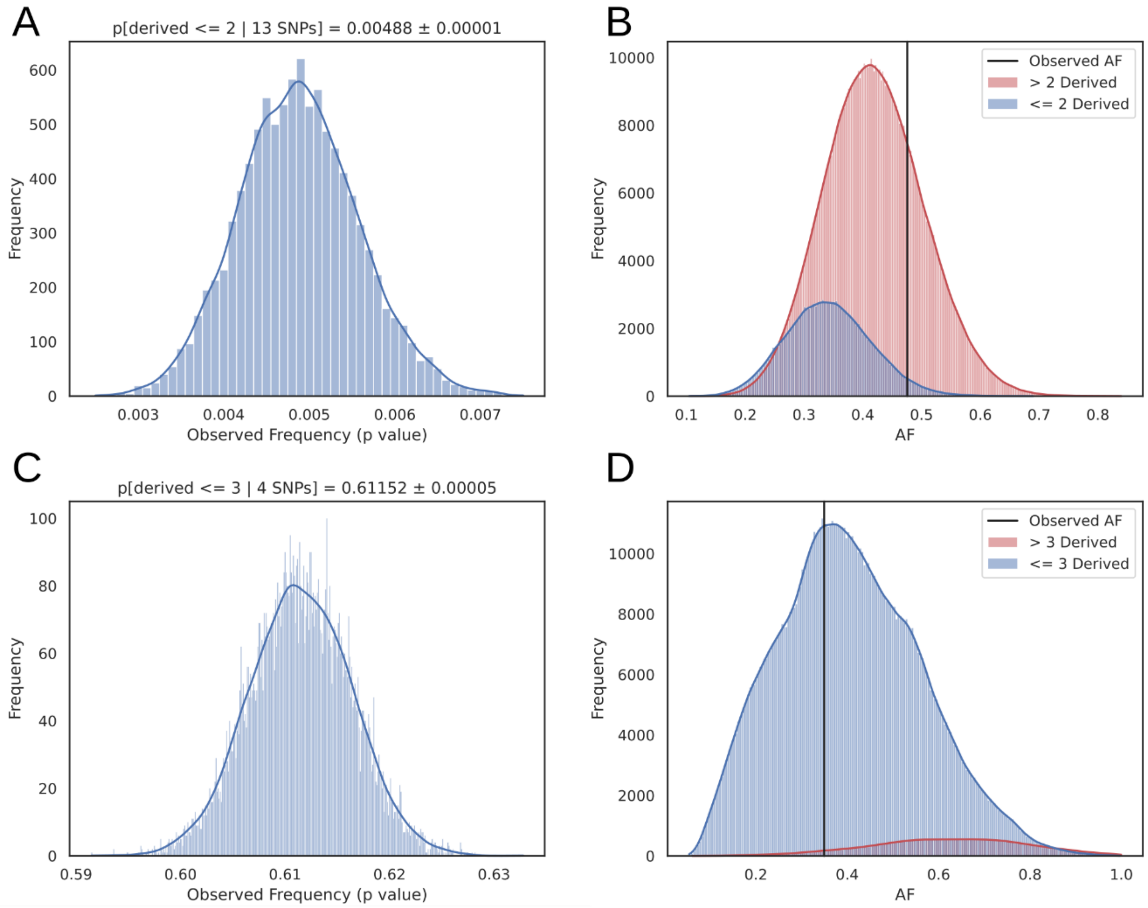
```
                                .......010.......020.......030.......040.......050.......060.......070.......080.......090.......100
>sp|P20138|CD33_HUMAN____  100: MPLLLLLPLLWAGALAMDPNFWLQVQESVTVQEGLCVLVPCTFFHPIPYYDKNSPVHGYWFREGAIISRDSPVATNKLDQEVQEETQGRFRLLGDPSRNN 100
>5j0b_A___________________ 099: -------------------GFWLQVQESVTVQEGLCVLVPCTFFHPIPYYDKNSPVHGYWFREGAIISGDSPVATNKLDQEVQEETQGRFRLLGDPSRNN 81
>6d49_A___________________ 099: ----------------GDPNFWLQVQESVTVQEGLCVLVPCTFFHPIPYYDKNSPVHGYWFREGAIISRDSPVATNKLDQEVQEETQGRFRLLGDPSRNN 84
>5j0b_chimp_A_____________ 095: -----------------DPKIRLQVQESVTVQEGLCVLVPCTFFHPIPYYDKNSPVHGYWFREGPIVSGDSPVATNKPDQEVQEETQGRFRLLGDLSRNN 83
>6d49_chimp_A_____________ 095: -----------------DPKIRLQVQESVTVQEGLCVLVPCTFFHPIPYYDKNSPVHGYWFREGPIVSGDSPVATNKPDQEVQEETQGRFRLLGDLSRNN 83

                                .......110.......120.......130.......140.......150.......160.......170.......180.......190.......200
>sp|P20138|CD33_HUMAN____  100: CSLSIVDARRRDNGSYFFRMERGSTKYSYKSPQLSVHVTDLTHRPKILIPGTLEPGHSKNLTCSVSWACEQGTPPIFSWLSAAPTSLGPRTTHSSVLIIT 200
>5j0b_A___________________ 099: CSLSIVDARRRDNGSYFFRMERGSTKYSYKSPQLSVHVTDLTHRPKILIPGTLEPGHSKNLTCSVSWACEQGTPPIFSWLSAAPTSLGPRTTHSSVLIIT 181
>6d49_A___________________ 099: CSLSIVDARRRDNGSYFFRMERGSTKYSYKSPQLSVHVTDLT---------------------------------------------------------- 126
>5j0b_chimp_A_____________ 095: CSLSIVDARRRDNGSYFFRMERGSTKYSYKSPQLSVHVTDLT---------------------------------------------------------- 125
>6d49_chimp_A_____________ 095: CSLSIVDARRRDNGSYFFRMERGSTKYSYKSPQLSVHVTDLT---------------------------------------------------------- 125
```

Supplemental Figure S3: Structure and dynamics of CD33. (A) Examples of x-ray structures of huCD33. PDB entries 5ihb (chain A: dark grey, chain B: white), 6d49 (lime). The dynamics of the C-C' loop and residues Phe21 and His45 are indicated. Positions of mutations present in chimpanzee are labeled on the pink spheres. **(B)** Amino acid sequences. 1: CD33 human (Uniprot), 2-3: PDB entries used for modeling. 3-4: sequences of the chCD33 models.





Supplemental Figure S4: MD trajectories of up/down states of Histidine at position 45 (His 45). Molecular dynamics of His45 side chain orientation. Individual MD trajectories of 'up'(blue)/'down'(grey) conformational states are shown. For chCD33, it can be observed reproducibly that simulations started with His45 in a 'down' conformational state undergo a transition to the 'up' conformational state on the microsecond timescale. Therefore, it may be concluded that chCD33 exists mainly with His45 in an 'up' orientation, which would be favorable for binding of Neu5Gc. For huCD33, both conformational states can exist for multiple µs, which explains why huCD33 can bind to Neu5Ac (preferably binds when His45 is 'down') and Neu5Gc (preferably binds when His45 is 'up').



Supplemental Figure S5: Resampling analysis of the 5.9 million SNPs from 1000 genomes variants. As an alternative to matching AF directly, the set of filtered SNPs were further restricted to those with a derived population frequency greater than 0.05 resulting in a universe of 5.9 million SNPs. We estimated the probability of observing at most two SNPs derived in either the Neanderthal or Denisovan reference genomes and a mean allele frequency as large as the empirical variants of interest (AF = 0.476). By randomly drawing SNPs, we found that the probability of observing 13 SNPs with such as high global allele frequency and lack of derived alleles in archaic genomes to be highly unusual (*p*-value = 0.00487 ± 0.00001) **(A).** The low frequency is driven by two factors, as shown in **(B).** Most of the SNPs sampled have more than two archaic-derived SNPs (red curve). Of those with fewer than two archaic-derived SNPs, the overall allele frequency is typically low compared to the target set. With other Siglec SNPs, resampling captures similar properties **(C** and **D)**, indicating the LOAD protective set does not represent a random sampling from the genome.

Gene	Associated disease	SNP ID	Description	References
CD33	LOAD	rs12459419, rs3865444	This study for details	Malik et al., 2013, Raj et al., 2014
APOE	LOAD, CVD	rs7412, rs429358	Encodes plasma protein APOE, is polymorphic in humans. Three alleles (E2, E3, E4) encode proteins with distinct affinity for lipoprotein particles. The ancestral E4 allele is associated with highest LOAD risk, and increased atherosclerosis and vascular dementia. The derived alleles E2 and E3 seems protective against LOAD, with the lowest risk is in homozygous E2 individuals.	Finch and Sapolsky, 1999, Fullerton et al., 2000
PICALM	AD	rs3851179 rs10792832	Encodes phosphatidylinositol-binding clathrin assembly protein (PICALM), considered to be one of numerous reproducible risk genes for LOAD.	Xu et al., 2015
SPON1	Dementia	rs2618516	Encodes the developmentally regulated protein F-spondin, reported to be a putative ligand for the amyloid precursor protein (APP).	Liu et al., 2018
TCFLC2	Diabetes	rs7903146	Associated with impaired insulin secretion and enhanced hepatic glucose production.	Helgason et al., 2007
ARID5B	AD	rs2588969	Gene encodes a member of AT-rich interaction domain (ARID) family of DNA binding proteins. The encoded protein forms a histone H3K9Me2 demethylase complex with PHD finger protein 2 and regulates the transcription of target genes involved in adipogenesis and liver development.	Carrasquillo et al., 2011
PILRA	AD	rs1859788	A cell surface inhibitory receptor that recognizes specific O-glycosylated proteins and expressed on various innate immune cell types including microglia	Rathore et al., 2018
CYP3A5	Salt retention and hypertension	rs776746	Cytochrome P450 (CYP) genes are abundant in animal, plant, and bacterial genomes and have evolved to metabolize a variety of diverse compounds.	Thompson et al., 2004
PPARG	Diabetes	rs1801282	A nuclear hormone receptor that regulates adipogenesis	Altshuler et al., 2000
BIN1	AD	rs7561528	Also known as amphiphysin 2, has recently been identified as the most important LOAD risk locus	Naj et al., 2011
SCG2	Hypertension	rs1017448	Secretogranin II (SCG2) associates with hypertension	Wen et al., 2007
CAPN10	Diabetes	rs2975760	CAPN10 encodes a member of the calpain-like cysteine protease family that regulates blood glucose levels.	Vander Molen et al., 2005
AGT	Sodium retention	rs699	Sodium homeostasis links with hypertension	Nakajima et al., 2004

LOAD: Late onset Alzheimer’s disease; AD: Alzheimer’s disease’; CVD: Cardiovascular disease

Supplemental Table S1: Genes affecting cognitive functions in post-reproductive age exhibiting disease-protective alleles uniquely in humans. The corresponding references for each of the genes are mentioned in the table.

Gene	SNP ID	Allele	Global frequency	African	East Asian	European	South Asian	American
CD33	rs12459419	C	0.789	0.949	0.814	0.69	0.84	0.52
		T	0.211	0.051	0.186	0.31	0.16	0.48
	rs3865444	C	0.789	0.949	0.814	0.69	0.84	0.52
		A	0.211	0.051	0.186	0.31	0.16	0.48
APOE	rs7412	C	0.925	0.897	0.9	0.937	0.96	0.96
		T	0.075	0.103	0.1	0.063	0.04	0.04
	rs429358	**T**	0.849	0.732	0.914	0.845	0.91	0.9
		C	0.151	0.268	0.086	0.155	0.09	0.1
PICALM	rs3851179	**T**	0.351	0.105	0.407	0.371	0.39	0.39
		C	0.685	0.895	0.593	0.629	0.61	0.61
	rs10792832	**A**	0.313	0.094	0.409	0.372	0.4	0.39
		G	0.685	0.895	0.593	0.628	0.6	0.61
SPON1	rs2618516	**T**	0.341	0.259	0.302	0.382	0.52	0.24
		C	0.659	0.741	0.698	0.618	0.48	0.76
TCFLC2	rs7903146	**C**	0.772	0.74	0.977	0.683	0.7	0.77
		T	0.228	0.26	0.0023	0.317	0.3	0.23
ARID5B	rs2588969	**C**	0.532	0.472	0.482	0.641	0.63	0.42
		A	0.468	0.528	0.518	0.359	0.37	0.58
PILRA	rs1859788	**A**	0.341	0.102	0.612	0.321	0.29	0.5
		G	0.659	0.898	0.388	0.679	0.71	0.5
CYP3A5	rs776746	T	0.379	0.82	0.287	0.05	0.33	0.2
		C	0.621	0.18	0.713	0.95	0.67	0.8
PPARG	rs1801282	C	0.93	0.995	0.974	0.88	0.88	0.88
		G	0.07	0.005	0.026	0.12	0.12	0.12
BIN1	rs7561528	**G**	0.8	0.809	0.881	0.683	0.87	0.74
		A	0.2	0.191	0.119	0.317	0.13	0.26
SGC2	rs1017448	**C**	0.879	0.635	0.963	0.979	0.97	0.95
		T	0.121	0.365	0.037	0.021	0.03	0.05
CAPN10	rs2975760	**T**	0.882	0.971	0.907	0.841	0.79	0.87
		C	0.118	0.029	0.093	0.159	0.21	0.13
AGT	rs699	**A**	0.295	0.097	0.147	0.588	0.36	0.36
		G	0.705	0.903	0.853	0.412	0.64	0.64

Supplemental Table S2: Analysis of Gene variants directly or indirectly affecting cognitive function with their human population frequency. The global frequency of the SNPs identified in Supplemental Table S1 was studied across different populations as indicated in the top of the columns.

Amino acid position	Human CD33(M)_EK_Fc Variant	Chimpanzee CD33(M)_EK_Fc Variant	Mutagenesis Primer Pairs_Forward/Reverse_5’ > 3’
20	N20K	-	TGGATCCAAAaTTCTGGCTGCAAGTGCAGG TAGCCAGGGCCCCTGCCC
21	F21I	-	GGATCCAAATaTCTGGCTGCAAGTGCAG ATAGCCAGGGCCCCTGCC
22	W22R	-	TCCAAATTTCcGGCTGCAAGTGCAGG TCCATAGCCAGGGCCCCT
65	A65P	-	CCGGGAAGGAcCCATTATATC AACCAGTAACCATGAACTG
66	I66F	-	GGAAGGAGCCtTTATATCCAGG CGGAACCAGTAACCATGAAC
67	I67V	-	AGGAGCCATTgTATCCAGGGAC TCCCGGAACCAGTAACCA
69	R69G	-	CATTATATCCgGGGACTCTCCAGTG GCTCCTTCCCGGAACCAG
78	L78P	-	ACAAACAAGCcAGATCAAGAAGTACAGGAG GGCCACTGGAGAGTCCCT
96	P96L	-	CTTGGGGATCtCAGTAGGAACAAC GAGGCGGAATCTGCCCTG
148	L148V	-	GCCCAAAATCgTCATCCCTGG CTGTGGGTCAAGTCTGTC
152	T152A	-	CATCCCTGGCgCTCTAGAACC AGGATTTTGGGCCTGTGG
154	E154D	-	GCACTCTAGAtCCCGGCCACT CAGGGATGAGGATTTTGGG
21	-	I21F	GGATCCAAAAtTCCGGCTGCAAGTG ATAGCCAGGGCCCCTGTG
20	-	K20N	TGGATCCAAAtATCCGGCTGCAAGTGC TAGCCAGGGCCCCTGTGG

Supplemental Table S3: List of the mutagenesis primers used in the study to generate the CD33 mutants. Lowercase letters correspond to base change.

Supplemental File S1: List of the glycans used for the sialoglycan microarray. The complete list of the chemoenzymatically synthesized glycans used to determine the binding profile of different CD33 proteins are presented. The binding intensity of the different proteins (indicated on the top of the columns) towards the corresponding glycan are shown in the heatmap (same heatmap as in Figure 3). The red indicates maximum, and blue indicates minimum binding. R = propylamine linker present in the underlying glycan structure. Gal = galactose, GalNAc = *N*-acetylgalactosamine, Glc = glucose, GlcNAc = *N*-acetyl glucosamine, Fuc =L-fucose. The linkage between the monosaccharides is indicated as α- or β- with numbers.

References

1. Angata T, Varki A. 2000. Cloning, characterization, and phylogenetic analysis of siglec-9, a new member of the CD33-related group of siglecs. Evidence for co-evolution with sialic acid synthesis pathways. J Biol Chem. 275:22127-22135.
2. Angata T, Kerr SC, Greaves DR, Varki NM, Crocker PR, Varki A. 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.
3. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 33:1870-1874.
4. Krieger E, Vriend G. 2014. YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics. 30:2981-2982.
5. Krieger E, Vriend G. 2015. New ways to boost molecular dynamics simulations. J Comput Chem. 36:996-1007.
6. Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. J Mol Graph. 14:33-8, 27.
7. Meng C, Sasmal A, Zhang Y, Gao T, Liu CC, Khan N, Varki A, Wang F, Cao H. 2018. Chemoenzymatic Assembly of Mammalian O-Mannose Glycans. Angew Chem Int Ed Engl. 57:9003-9007.
8. Lu N, Ye J, Cheng J et al. 2019. Redox-Controlled Site-Specific α2-6-Sialylation. J Am Chem Soc. 141:4547-4552.
9. Ji Y, Sasmal A, Li W et al. 2021. Reversible O-Acetyl Migration within the Sialic Acid Side

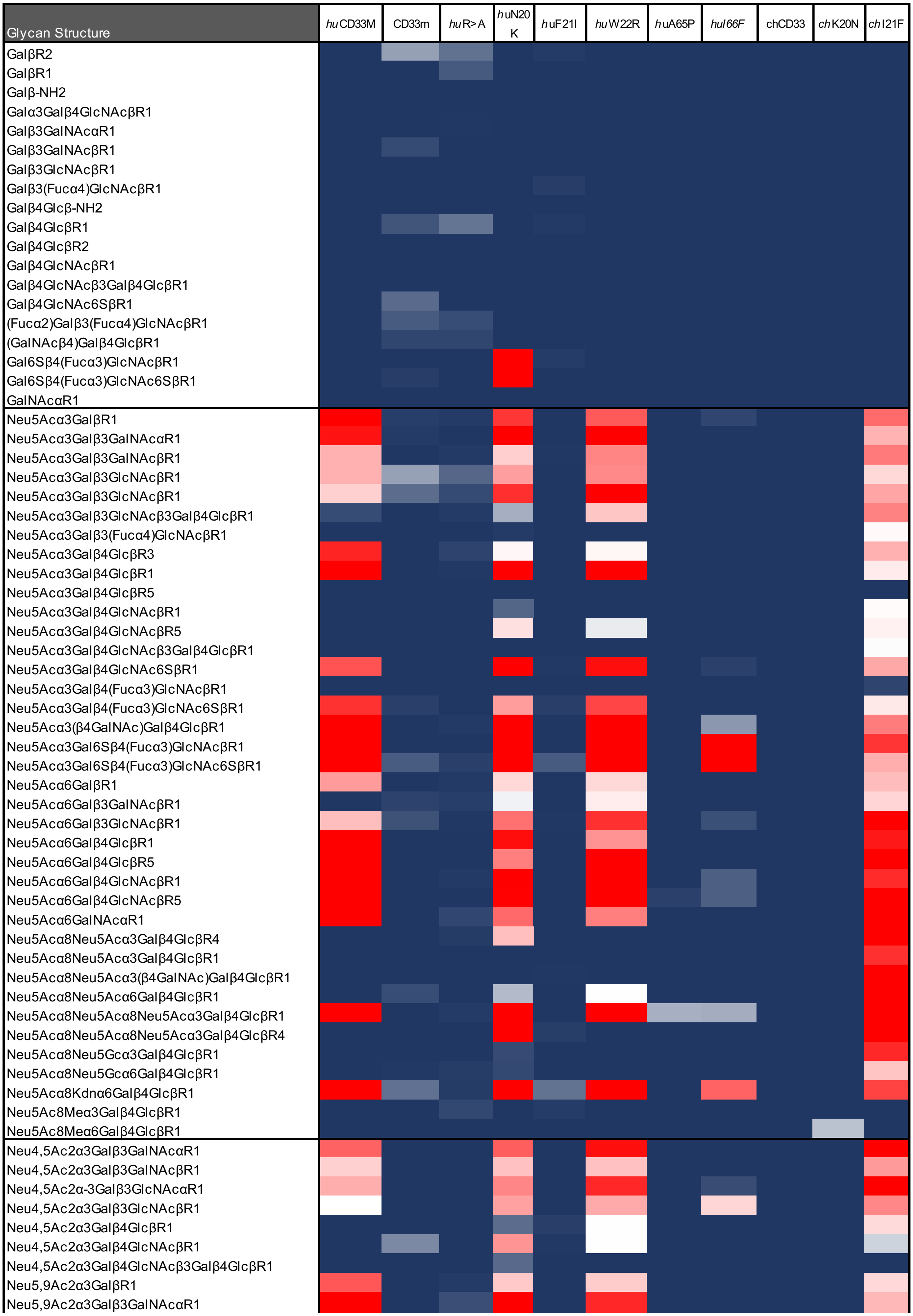
Chain and Its Influence on Protein Recognition. ACS Chem Biol. 16:1951-1960.

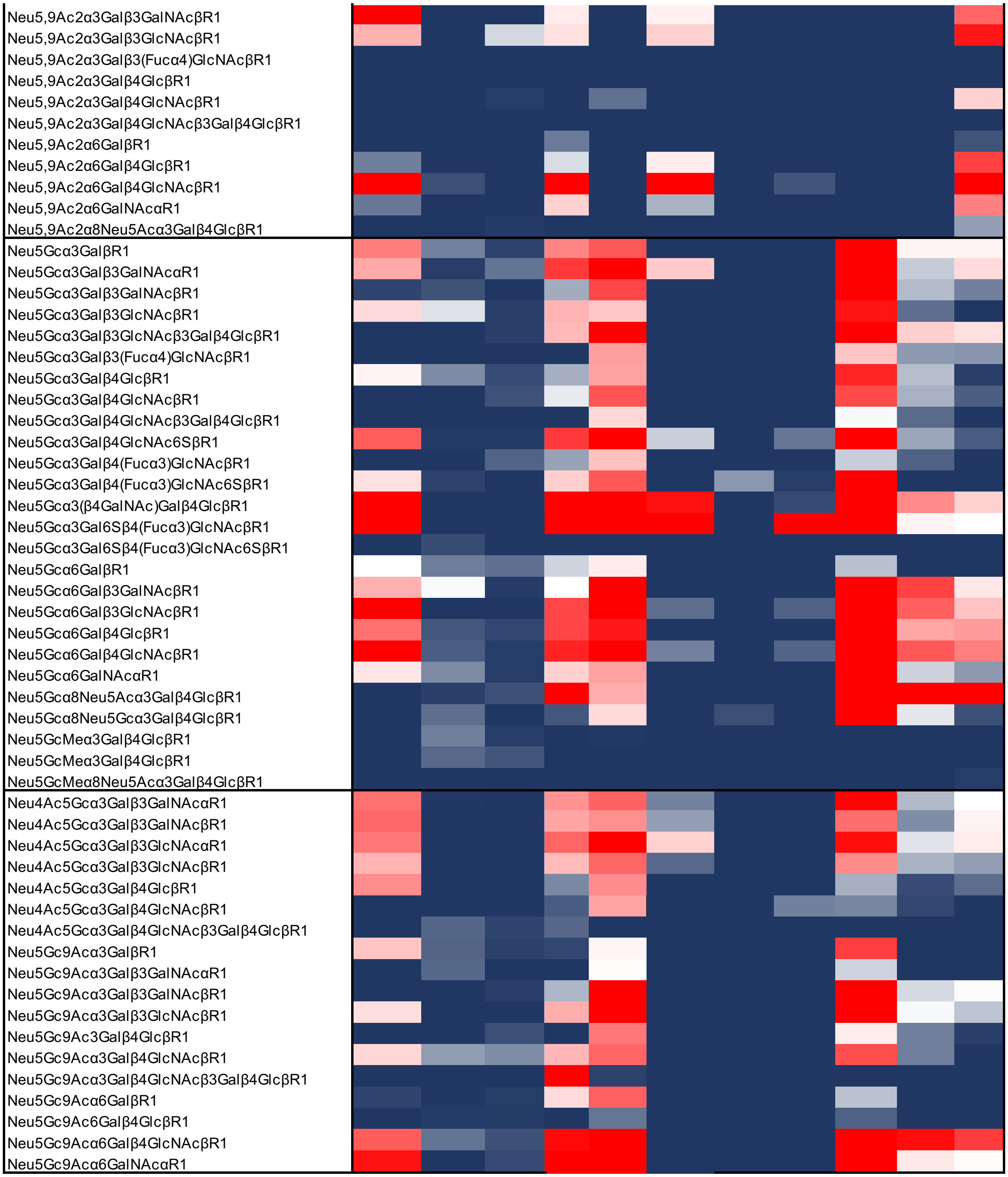
10. Malik M, Simpson JF, Parikh I, Wilfred BR, Fardo DW, Nelson PT, Estus S. 2013. CD33 Alzheimer's risk-altering polymorphism, CD33 expression, and exon 2 splicing. J Neurosci. 33:13320-13325.
11. Raj T, Ryan KJ, Replogle JM et al. 2014. CD33: increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. Hum Mol Genet.
12. Fullerton SM, Clark AG, Weiss KM et al. 2000. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. Am J Hum Genet. 67:881-900.
13. Finch CE, Sapolsky RM. 1999. The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms. Neurobiol Aging. 20:407-428.
14. Xu W, Tan L, Yu JT. 2015. The Role of PICALM in Alzheimer's Disease. Mol Neurobiol. 52:399-413.
15. Liu Z, Dai X, Tao W et al. 2018. APOE influences working memory in non-demented elderly through an interaction with SPON1 rs2618516. Hum Brain Mapp. 39:2859-2867.
16. Helgason A, Pálsson S, Thorleifsson G et al. 2007. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. Nat Genet. 39:218-225.
17. Carrasquillo MM, Belbin O, Hunter TA et al. 2011. Replication of EPHA1 and CD33 associations with late-onset Alzheimer's disease: a multi-centre case-control study. Mol Neurodegener. 6:54.
18. Rathore N, Ramani SR, Pantua H et al. 2018. Paired Immunoglobulin-like Type 2 Receptor Alpha G78R variant alters ligand binding and confers protection to Alzheimer's disease. PLoS Genet. 14:e1007427.
19. Thompson EE, Kuttab-Boulos H, Witonsky D, Yang L, Roe BA, Di Rienzo A. 2004. CYP3A variation and the evolution of salt-sensitivity variants. Am J Hum Genet. 75:1059-1069.
20. Altshuler D, Hirschhorn JN, Klannemark M et al. 2000. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet. 26:76-80.
21. Naj AC, Jun G, Beecham GW et al. 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet. 43:436-441.
22. Wen G, Wessel J, Zhou W et al. 2007. An ancestral variant of Secretogranin II confers

regulation by PHOX2 transcription factors and association with hypertension. Hum Mol Genet. 16:1752-1764.

23. Vander Molen J, Frisse LM, Fullerton SM, Qian Y, Del Bosque-Plata L, Hudson RR, Di Rienzo A. 2005. Population genetics of CAPN10 and GPR35: implications for the evolution of type 2 diabetes variants. Am J Hum Genet. 76:548-560.

24. Nakajima T, Wooding S, Sakagami T et al. 2004. Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. Am J Hum Genet. 74:898-916.


Glycan Structure
*hu*CD33M
CD33m
*hu*R>A
*hu*N20K
*hu*F21I
*hu*W22R
*hu*A65P
huI66F
chCD33
*ch*K20N
*ch*I21F
GalβR2
GalβR1
Galβ-NH2
Galα3Galβ4GlcNAcβR1
Galβ3GalNAcαR1
Galβ3GalNAcβR1
Galβ3GlcNAcβR1
Galβ3(Fucα4)GlcNAcβR1
Galβ4Glcβ-NH2
Galβ4GlcβR1
Galβ4GlcβR2
Galβ4GlcNAcβR1
Galβ4GlcNAcβ3Galβ4GlcβR1
Galβ4GlcNAc6SβR1
(Fucα2)Galβ3(Fucα4)GlcNAcβR1
(GalNAcβ4)Galβ4GlcβR1
Gal6Sβ4(Fucα3)GlcNAcβR1
Gal6Sβ4(Fucα3)GlcNAc6SβR1
GalNAcαR1
Neu5Acα3GalβR1
Neu5Acα3Galβ3GalNAcαR1
Neu5Acα3Galβ3GalNAcβR1
Neu5Acα3Galβ3GlcNAcβR1
Neu5Acα3Galβ3GlcNAcβR1
Neu5Acα3Galβ3GlcNAcβ3Galβ4GlcβR1
Neu5Acα3Galβ3(Fucα4)GlcNAcβR1
Neu5Acα3Galβ4GlcβR3
Neu5Acα3Galβ4GlcβR1
Neu5Acα3Galβ4GlcβR5
Neu5Acα3Galβ4GlcNAcβR1
Neu5Acα3Galβ4GlcNAcβR5
Neu5Acα3Galβ4GlcNAcβ3Galβ4GlcβR1
Neu5Acα3Galβ4GlcNAc6SβR1
Neu5Acα3Galβ4(Fucα3)GlcNAcβR1
Neu5Acα3Galβ4(Fucα3)GlcNAc6SβR1
Neu5Acα3(β4GalNAc)Galβ4GlcβR1
Neu5Acα3Gal6Sβ4(Fucα3)GlcNAcβR1
Neu5Acα3Gal6Sβ4(Fucα3)GlcNAc6SβR1
Neu5Acα6GalβR1
Neu5Acα6Galβ3GalNAcβR1
Neu5Acα6Galβ3GlcNAcβR1
Neu5Acα6Galβ4GlcβR1
Neu5Acα6Galβ4GlcβR5
Neu5Acα6Galβ4GlcNAcβR1
Neu5Acα6Galβ4GlcNAcβR5
Neu5Acα6GalNAcαR1
Neu5Acα8Neu5Acα3Galβ4GlcβR4
Neu5Acα8Neu5Acα3Galβ4GlcβR1
Neu5Acα8Neu5Acα3(β4GalNAc)Galβ4GlcβR1
Neu5Acα8Neu5Acα6Galβ4GlcβR1
Neu5Acα8Neu5Acα8Neu5Acα3Galβ4GlcβR1
Neu5Acα8Neu5Acα8Neu5Acα3Galβ4GlcβR4
Neu5Acα8Neu5Gcα3Galβ4GlcβR1
Neu5Acα8Neu5Gcα6Galβ4GlcβR1
Neu5Acα8Kdnα6Galβ4GlcβR1
Neu5Ac8Meα3Galβ4GlcβR1
Neu5Ac8Meα6Galβ4GlcβR1
Neu4,5Ac2α3Galβ3GalNAcαR1
Neu4,5Ac2α3Galβ3GalNAcβR1
Neu4,5Ac2α-3Galβ3GlcNAcαR1
Neu4,5Ac2α3Galβ3GlcNAcβR1
Neu4,5Ac2α3Galβ4GlcβR1
Neu4,5Ac2α3Galβ4GlcNAcβR1
Neu4,5Ac2α3Galβ4GlcNAcβ3Galβ4GlcβR1
Neu5,9Ac2α3GalβR1
Neu5,9Ac2α3Galβ3GalNAcαR1


Neu5,9Ac2α3Galβ3GalNAcβR1
Neu5,9Ac2α3Galβ3GlcNAcβR1
Neu5,9Ac2α3Galβ3(Fucα4)GlcNAcβR1
Neu5,9Ac2α3Galβ4GlcβR1
Neu5,9Ac2α3Galβ4GlcNAcβR1
Neu5,9Ac2α3Galβ4GlcNAcβ3Galβ4GlcβR1
Neu5,9Ac2α6GalβR1
Neu5,9Ac2α6Galβ4GlcβR1
Neu5,9Ac2α6Galβ4GlcNAcβR1
Neu5,9Ac2α6GalNAcαR1
Neu5,9Ac2α8Neu5Acα3Galβ4GlcβR1
Neu5Gcα3GalβR1
Neu5Gcα3Galβ3GalNAcαR1
Neu5Gcα3Galβ3GalNAcβR1
Neu5Gcα3Galβ3GlcNAcβR1
Neu5Gcα3Galβ3GlcNAcβ3Galβ4GlcβR1
Neu5Gcα3Galβ3(Fucα4)GlcNAcβR1
Neu5Gcα3Galβ4GlcβR1
Neu5Gcα3Galβ4GlcNAcβR1
Neu5Gcα3Galβ4GlcNAcβ3Galβ4GlcβR1
Neu5Gcα3Galβ4GlcNAc6SβR1
Neu5Gcα3Galβ4(Fucα3)GlcNAcβR1
Neu5Gcα3Galβ4(Fucα3)GlcNAc6SβR1
Neu5Gcα3(β4GalNAc)Galβ4GlcβR1
Neu5Gcα3Gal6Sβ4(Fucα3)GlcNAcβR1
Neu5Gcα3Gal6Sβ4(Fucα3)GlcNAc6SβR1
Neu5Gcα6GalβR1
Neu5Gcα6Galβ3GalNAcβR1
Neu5Gcα6Galβ3GlcNAcβR1
Neu5Gcα6Galβ4GlcβR1
Neu5Gcα6Galβ4GlcNAcβR1
Neu5Gcα6GalNAcαR1
Neu5Gcα8Neu5Acα3Galβ4GlcβR1
Neu5Gcα8Neu5Gcα3Galβ4GlcβR1
Neu5GcMeα3Galβ4GlcβR1
Neu5GcMeα3Galβ4GlcβR1
Neu5GcMeα8Neu5Acα3Galβ4GlcβR1
Neu4Ac5Gcα3Galβ3GalNAcαR1
Neu4Ac5Gcα3Galβ3GalNAcβR1
Neu4Ac5Gcα3Galβ3GlcNAcαR1
Neu4Ac5Gcα3Galβ3GlcNAcβR1
Neu4Ac5Gcα3Galβ4GlcβR1
Neu4Ac5Gcα3Galβ4GlcNAcβR1
Neu4Ac5Gcα3Galβ4GlcNAcβ3Galβ4GlcβR1
Neu5Gc9Acα3GalβR1
Neu5Gc9Acα3Galβ3GalNAcαR1
Neu5Gc9Acα3Galβ3GalNAcβR1
Neu5Gc9Acα3Galβ3GlcNAcβR1
Neu5Gc9Ac3Galβ4GlcβR1
Neu5Gc9Acα3Galβ4GlcNAcβR1
Neu5Gc9Acα3Galβ4GlcNAcβ3Galβ4GlcβR1
Neu5Gc9Acα6GalβR1
Neu5Gc9Ac6Galβ4GlcβR1
Neu5Gc9Acα6Galβ4GlcNAcβR1
Neu5Gc9Acα6GalNAcαR1