Specific inactivation of two immunomodulatory *SIGLEC* genes during human evolution

Xiaoxia Wang[a,b,1], Nivedita Mitra[a,b,1], Ismael Secundino[a,b,c,1], Kalyan Banda[a,b], Pedro Cruz[d], Vered Padler-Karavani[a,b], Andrea Verhagen[a,b], Chris Reid[a,b], Martina Lari[e], Ermanno Rizzi[f], Carlotta Balsamo[e], Giorgio Corti[f], Gianluca De Bellis[f], Laura Longo[g], NISC Comparative Sequencing Program[d,2], William Beggs[h], David Caramelli[e], Sarah A. Tishkoff[h], Toshiyuki Hayakawa[i], Eric D. Green[d], James C. Mullikin[d], Victor Nizet[a,b,c], Jack Bui[a], and Ajit Varki[a,b,3]

[a]Departments of Medicine, Cellular and Molecular Medicine, Pathology, and Pediatrics, [b]Glycobiology Research and Training Center, and [c]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093; [d]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; [e]Department of Evolutionary Biology, University of Florence, 50122 Florence, Italy; [f]Institute for Biomedical Technologies (ITB), National Research Council (CNR), 20090 Milan, Italy; [g]Department of Environmental Science, University of Siena, 53100 Siena, Italy; [h]Departments of Genetics and Biology, University of Pennsylvania, Philadelphia, PA 19104; and [i]Primate Research Institute, Kyoto University, Inuyama 484-8506, Japan

Edited by Francisco J. Ayala, University of California, Irvine, CA, and approved April 25, 2012 (received for review November 28, 2011)

Sialic acid-recognizing Ig-like lectins (Siglecs) are signaling receptors that modulate immune responses, and are targeted for interactions by certain pathogens. We describe two primate Siglecs that were rendered nonfunctional by single genetic events during hominin evolution after our common ancestor with the chimpanzee. *SIGLEC13* was deleted by an *Alu*-mediated recombination event, and a single base pair deletion disrupted the ORF of *SIGLEC17*. Siglec-13 is expressed on chimpanzee monocytes, innate immune cells that react to bacteria. The human *SIGLEC17P* pseudogene mRNA is still expressed at high levels in human natural killer cells, which bridge innate and adaptive immune responses. As both resulting pseudogenes are homozygous in all human populations, we resurrected the originally encoded proteins and examined their functions. Chimpanzee Siglec-13 and the resurrected human Siglec-17 recruit a signaling adapter and bind sialic acids. Expression of either Siglec in innate immune cells alters inflammatory cytokine secretion in response to Toll-like receptor-4 stimulation. Both Siglecs can also be engaged by two potentially lethal sialylated bacterial pathogens of newborns and infants, agents with a potential impact on reproductive fitness. Neanderthal and Denisovan genomes show human-like sequences at both loci, corroborating estimates that the initial pseudogenization events occurred in the common ancestral population of these hominins. Both loci also show limited polymorphic diversity, suggesting selection forces predating the origin of modern humans. Taken together, these data suggest that genetic elimination of Siglec-13 and/or Siglec-17 represents signatures of infectious and/or other inflammatory selective processes contributing to population restrictions during hominin origins.

Sialic acids (Sias) are monosaccharides typically found at the outermost ends of complex glycan chains that decorate all vertebrate cell surfaces (1, 2). Sias are essential for embryonic development (3) and mediate important intrinsic organismal functions (1, 2). However, given their location and density, Sia-bearing glycans are also targets for recognition by many pathogen-binding proteins and toxins (1, 4, 5). Adding complexity to these opposing evolutionary selection forces, many important bacterial pathogens have evolved convergent mechanisms for molecular mimicry of host Sias (4, 6, 7). For all these reasons, both Sias and Sia-recognizing proteins are rapidly evolving in some taxa. Current data suggest that humans are an extreme example, with Sia-related genes representing a "hotspot" in human evolution (5). Of less than 70 human genes known to be involved in Sia biology, more than 10 have been documented to exhibit human-specific changes relative to the chimpanzee, our closest evolutionary cousins (5). Most of these human-specific genetic changes are in *Sia-recognizing Ig-like lectin* (*SIGLEC*) genes (5).

SIGLEC genes encode a family of transmembrane receptors that bind Sia-containing ligands via their amino-terminal extracellular Ig-like domains and modulate cellular responses via cytosolic signaling motifs (7–10). The CD33-related subset of Siglecs (CD33rSiglecs) is rapidly evolving within vertebrates (7–10). The most likely reason is that CD33rSiglecs are prominently expressed on innate immune cells, and modulate responses to pathogens. In this regard, Siglec-3 and Siglecs-5 to -11 in humans seem to recognize sialylated ligands as "self-associated molecular patterns" (11), limiting unwanted reactivity against other cells in the same organism (7–10). However, certain immune-modulating bacterial pathogens carry out molecular mimicry of sialylated CD33rSiglec ligands (12), dampening host innate immune cell responses and facilitating infection (13). In one instance, a human bacterial pathogen evolved a more stable protein–protein interaction with an inhibitory CD33rSiglec to suppress innate immunity (14).

Evolution has also generated CD33rSiglecs with opposing activatory potential (7, 15, 16), transmitting positive signals to immune cells via recruitment of the immunoreceptor tyrosine-based activation motif (ITAM)-containing DAP12 adaptor protein (17). Some activatory Siglecs pair with inhibitory ones (15, 16), supporting the notion that they represent a host evolutionary response to pathogen mimicry and engagement of inhibitory CD33rSiglecs (7). However, low avidity engagement of activatory Siglecs can mediate paradoxical inhibitory responses (ITAMi) (18, 19).

Two genomic loci encoding ITAM-containing primate Siglecs (*SIGLEC14* and *SIGLEC16*) are polymorphic, with their common alleles being nonfunctional (15, 16, 20). These polymorphisms exist in African populations, indicating that they likely originated before the migration of modern humans out of Africa about 60,000–70,000 y ago. Thus, functional and pseudogene alleles of *SIGLEC* genes can be maintained in populations over long periods. This finding likely reflects ongoing selection forces involving the need to maintain innate immune self-recognition and control damaging inflammatory responses, all against a backdrop of potential pathogen subversion of these mechanisms (7, 8, 20). An evolutionary balancing act is also supported by the high frequency of human-specific pathogens that carry out molecular mimicry of Sias through convergent evolutionary mechanisms (4, 6).

Author contributions: X.W. and A.Varki designed research; X.W., N.M., I.S., K.B., V.P.-K., A.Verhagen, C.R., M.L., E.R., C.B., G.C., G.D.B., L.L., N.I.S.C.C.S.P., W.B., D.C., S.A.T., E.D.G., J.C.M., V.N., and J.B. performed research; X.W. and N.M. contributed new reagents/analytic tools; X.W., P.C., W.B., S.A.T., and T.H. analyzed data; and X.W. and A.Varki wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]X.W., N.M., and I.S. contributed equally to this work.

[2]A complete list of the National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Staff can be found in the *SI Appendix*.

[3]To whom correspondence should be addressed. E-mail: a1varki@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1119459109/-/DCSupplemental.

EVOLUTION

Current genetic, archeological, skeletal, and radioactive dating evidence indicates that all modern humans are derived from a population with an effective size of 10,000 or fewer that originated in Africa ~100,000–200,000 y ago (21–25) and later spread across the planet, replacing other hominin species and having limited interbreeding with our closest extinct cousins, the Neanderthals (26) and Denisovans (27). Even smaller estimates have been made of the effective population size of Neanderthals (28). The reasons for the small effective population sizes are unknown, and host–pathogen interactions have not been previously considered as contributors. Here, we report two genomic inactivation events in *SIGLEC* genes that became fixed before the emergence of modern humans in Africa. Our evidence indicates that these genes may have been inactivated in the hominin lineage, because they can be engaged by pathogens associated with life-threatening invasive infections in newborns and infants. It is also possible that the immunomodulatory capacity of these genes was costly in contexts beyond bacterial engagement of Siglecs, such as toxic inflammatory effects on ancestral hominin immune cells.

Results and Discussion

Single *Alu*-Mediated Deletion Event Inactivated *SIGLEC13* in the Hominin Lineage. Analysis of genomic BAC clones indicated that the primate *SIGLEC13* gene was missing from the human genome but present in chimpanzees and baboons (29). To understand events accounting for this apparently human-specific deletion, we analyzed the genomic region encompassing the *SIGLEC13* locus in the currently available sequence builds of the human, chimpanzee, baboon, and rhesus genomes. RepeatMasker software identified several repetitive elements in this region in all of these species. These elements include *Alu* elements, which are primate-specific short interspersed elements (30). In the chimpanzee genome, five *Alu* elements are located in an ~10-kb genomic region containing the *SIGLEC13* locus (Fig. 1*A*). One *Alu* element belonging to the *Alu*Jb family is located upstream of the *SIGLEC13* locus, one *AluSx* and one *AluSz* elements are within the locus, and two *AluSc* elements are found downstream. These elements are also found in orthologous regions in the baboon and rhesus monkey genomes (Fig. 1*A*). In the human genome, we found just one composite *Alu* element (*Alu*Jb/Sc) occupying the region of ~7 kb that contains the *SIGLEC13* locus in chimpanzee, rhesus monkey, and baboon (Fig. 1*A*). The ~7-kb region deleted in the human genome is sandwiched between two ancestral *Alu*Jb and *Alu*Sc elements (Fig. 1*A*). Thus, a single *Alu*-mediated recombination event was the likely mechanism for human-specific deletion of *SIGLEC13*, leaving a single fused *Alu* element in the human genome.

Human-Specific Mutational Events Functionally Altered and then, Pseudogenized Primate *SIGLEC17*. We previously described the human *SIGLECP3* locus with an inactivating deletion in the predicted ORF (29). We now note that the rest of the predicted ORF remains intact. We found more than one cDNA clone predicting a full-length transcript derived from *SIGLECP3* (for example, BC041072 in *SI Appendix*, Fig. S1). In addition to the single nucleotide deletion, the remnant human *SIGLECP3* coding region harbors a human-unique missense mutation of the codon encoding an Arg residue that would have been involved in Sia recognition when the ORF was intact (*SI Appendix*, Fig. S2*A*). Taken together with evidence for Sia-binding properties of the resurrected protein when the Arg codon is restored (see below) and evidence for an intact marmoset ortholog (*SI Appendix*, Fig. S2*A*), we redesignate the original *SIGLECP3* locus as primate *SIGLEC17* and the corresponding human pseudogene as *SIGLEC17P*.

Phylogenetic sequence comparisons showed that the 1-bp deletion in *SIGLEC17P* is human-specific (Fig. 1*B*). A BLAST query using the human sequence identified an orthologous gene with a predicted intact ORF encoding Siglec-17 in the marmoset genome, with 93% DNA sequence similarity and 89% predicted protein

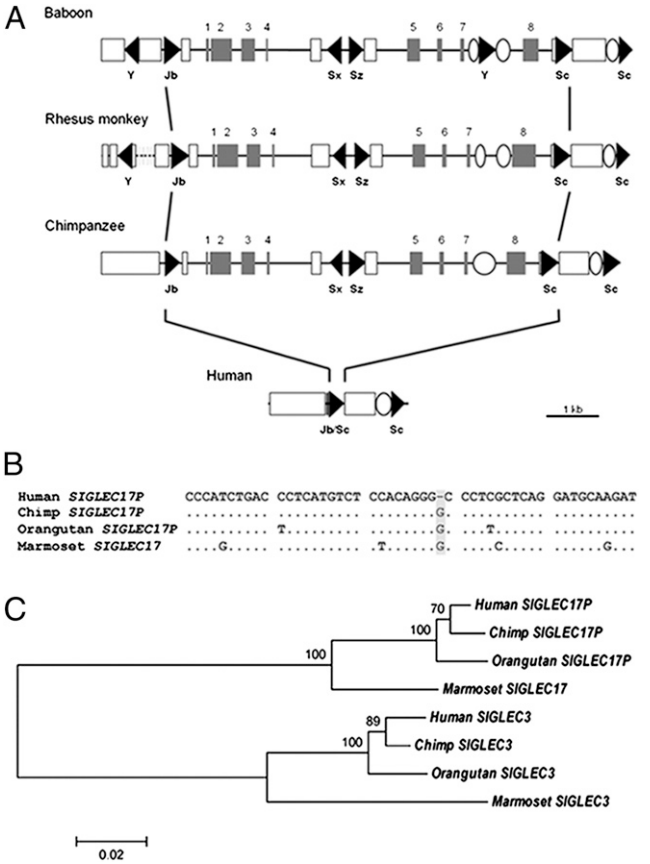


Fig. 1. Inactivation of two *SIGLEC* genes during hominin evolution. (*A*) Comparison of genomic structure surrounding the *SIGLEC13* locus among humans and other primates. Coding regions are represented by shaded boxes, and Alu elements are represented by triangles. Names of Alu subfamilies are shown. Open boxes and ellipses indicate LINE (long interspersed element) and LTR (long terminal repeat) elements, respectively. The dotted line indicates a sequence gap. (*B*) DNA sequence alignment using Clustal W in MEGA4 shows the human-specific loss of G in *SIGLEC17P* (highlighted in gray). (*C*) Reconstructed neighbor-joining (NJ) tree of *SIGLEC17P* and *SIGLEC3* among primates. Bootstrap values of 1,000 replicates are shown on internal branches. MEGA4 was used for NJ tree reconstruction and bootstrap analysis.

sequence similarity (*SI Appendix*, Fig. S2*A*). Analysis of the RT-PCR–derived orthologous segment from chimpanzee peripheral blood mononuclear cell mRNA confirmed that, although the 1-bp deletion is human-specific, independent pseudogenization events have occurred in the chimpanzee as well as orangutan genomes. The homologous region containing the *SIGLEC17P* locus seems to be completely deleted in the rhesus and baboon genomes (29). Thus, although well-conserved in primate evolution from New World monkeys to ancestral hominins, the *SIGLEC17* gene has also undergone independent deletion or pseudogenization events in multiple primate taxa, with a distinct event in the hominin lineage.

Additional BLAST analyses showed that *SIGLEC17* is most closely related to *SIGLEC3* (encoding Siglec-3/CD33), with the two loci evolving as paralogs in primates (Fig. 1*C*). The predicted V-set and C2-set domains of the resurrected human *SIGLEC17* and human *SIGLEC3* genes share 66% DNA sequence similarity (*SI Appendix*, Fig. S2*B*). The human pseudogene *SIGLECP6* also has homology to human *SIGLEC17P*. A phylogenetic tree of primate *SIGLEC3*, *SIGLECP6*, and *SIGLEC17P* gave no evidence for a gene conversion during primate evolution among these loci.

Inactivation Events of *SIGLEC13* and *SIGLEC17* Are Human-Universal. The *Alu*-mediated *SIGLEC13* deletion and the *SIGLEC17* frame-

shift mutation were homozygous in 28 HapMap human samples representing a worldwide population (11 Yoruba Africans, 9 Japanese and Chinese, and 8 North Europeans). These events were not found in genomes from 19 chimpanzees (samples provided by Pascal Gagneux, University of California, San Diego, CA) or 6 orangutans (samples from the Coriell repository), ruling out an ancestral hominid polymorphism. Notably, except for the universal frame-shift mutation, the rest of the *SIGLEC17* ORF was intact in all 28 humans. The HapMap sample results left open the possibility that rare functional alleles might persist in some African populations. However, both mutations were found to be in a homozygous state in ~230 ethnically and geographically diverse Africans comprising many major African population groups (*SI Appendix*, Table S1) (31). This finding is in contrast to *SIGLEC14* and *SIGLEC16*, which show polymorphic segregating pseudogenization in all human populations (15, 16).

As these pseudogenization events of *SIGLEC13* and *SIGLEC17* are both human-unique and -universal compared with other primates, we decided to explore the functional implications by studying the chimpanzee ortholog of *SIGLEC13* and the functionally resurrected form of human *SIGLEC17*.

Siglec-13 Is Selectively Expressed on Chimpanzee Monocytes. A monoclonal antibody against the recombinant soluble extracellular domains of chimpanzee Siglec-13 (*SI Appendix*) was used for flow cytometry analysis of chimpanzee peripheral blood leukocytes, showing expression predominantly on monocytes (Fig. 2*A*). As expected from the human *SIGLEC13* deletion, there was no specific antibody binding to human leukocytes. Thus, earlier hominin ancestors likely expressed Siglec-13 on monocytes, a cell type participating in innate immune defenses against bacterial pathogens.

Human *SIGLEC17P* Message Is Selectively Expressed in Natural Killer Cells. Because the *SIGLEC17* ORF is interrupted by independent events in the human and chimpanzee genomes, we could not study chimpanzee cells to understand the original function of this protein. Gene Atlas analysis showed prominent expression of *SIGLEC17P* message almost exclusively in natural killer (NK) cells (Fig. 2*B*). This finding was confirmed by quantitative RT-PCR in human peripheral blood mononuclear cells, with the strongest signal in isolated NK cells (Fig. 2*C*). Thus, hominin ancestors likely expressed Siglec-17 on NK cells, a cell type that can activate macrophages to mount antibacterial functions and bridge the innate and adaptive immune responses.

Chimpanzee Siglec-13 and Functionally Resurrected Human Siglec-17 Bind Sias. Recombinant soluble chimpanzee Siglec-13-Fc and human Siglec-17-Fc with a functionally restored ORF were prepared by fusing sequences encoding their first two extracellular Ig-like domains with coding regions for the Fc portion of human IgG1. Both Siglec-Fcs bound to a sialoglycan microarray, and binding was lost when the essential arginine residue was absent (*SI Appendix*, Fig. S3). These data show that these genes originally encoded functional Siglecs in hominin ancestors.

Both Proteins Interact with the DAP12 Signaling Adaptor Protein. The ORFs of chimpanzee *SIGLEC13* and resurrected human *SIGLEC17* each encode a transmembrane protein with a single positively charged residue within the membrane-spanning region. Other immune cell proteins with this feature (including human Siglec-14 and -16) can associate with DNAX-activating protein of 12 kDa (DAP12) (17), which has a corresponding negatively charged residue within the membrane-spanning region (15, 16). Transient transfection of a FLAG-tagged DAP12 cDNA into 293T cells along with *SIGLEC13* or *SIGLEC17* cDNAs was required for optimal surface expression of both molecules (Fig. 3 *A* and *B*). Consistent with this finding, a DAP12–FLAG-tagged protein could coimmunoprecipitate with Siglec-13 or -17 (Fig. 3 *C* and *D*). Mutation of the single positively charged residue within the transmembrane domain of each Siglec to a neutral alanine residue diminished the expression of both proteins. Thus, the positively charged residue is involved in surface expression of both *SIGLEC13* and *SIGLEC17* genes (Fig. 3 *A* and *B*).

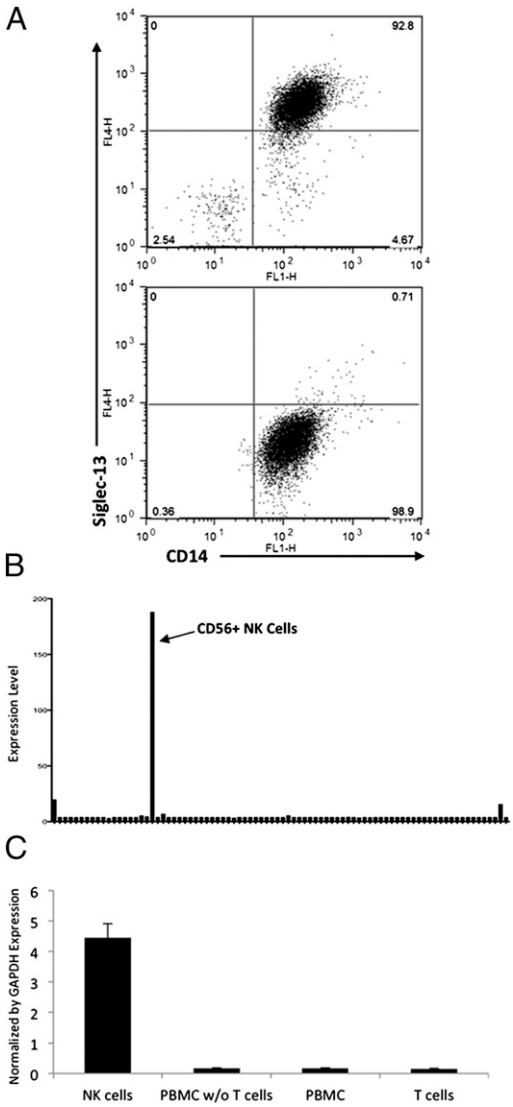


Fig. 2. Cell-type specific expression of Siglec-13 and -17. (*A*) Chimpanzee and human peripheral blood mononuclear cells (PBMCs) were labeled with monoclonal anti–Siglec-13 mouse IgG. Colabeling for CD14 (a marker for PBMCs) showed specific expression in chimpanzee (*Upper*), but not in human (*Lower*) monocytes. (*B*) Transcript expression profile of human *SIGLEC17P* in 84 human tissues and cell lines acquired from BioGPS with probe gnf1h07492_at. The list of tissues and cell types studied can be found in the *SI Appendix*. (*C*) Quantitative RT-PCR for *SIGLEC17P* expression in human NK cells. Human T cells, whole-blood PBMCs, and whole-blood PBMCs depleted of T cells were used as controls. Results were normalized to GAPDH expression.

Functional Analysis of Chimpanzee Siglec-13 or Resurrected Human Siglec-17. To examine signaling events mediated by chimpanzee Siglec-13 and resurrected Siglec-17, we used RAW246.7 macrophages that intrinsically expresses DAP12 (32). Semistably transfected RAW264.7 cells with Siglec-13 or -17 cDNAs in pcDNA3.1 were acquired after 3 wk of selection in 1.5 mg/mL G418. Expression of the Siglecs in the transfected cells was confirmed by flow cytometry or RT-PCR (*SI Appendix*, Fig. S4). As shown in *SI Appendix*, Fig. S5 *A* and *B*, semistably transfected cells with Siglec-13 or -17 showed increased intracellular TNF levels after low-dose LPS stimulation [Toll-like receptor 4 (TLR4) activation] compared with mock-transfected cells. With Siglec-13 transfection, untreated

cells sometimes displayed constitutive TNF production, which was further boosted by LPS (*SI Appendix*, Fig. S5*A*). Thus, Siglec-13 and -17 likely mediated signaling through DAP12 in ancestral monocytes and NK cells, modulating cytokine secretion.

Specific Interactions of Siglec-13 and -17 with Sialylated Bacterial Pathogens. Some other CD33-related Siglecs are known to recognize certain pathogenic bacteria (13, 14) [e.g., Group B *Streptococcus* (GBS), a Gram-positive bacterial pathogen that expresses Sias and causes invasive infections in human newborn infants (33)]. GBS engagement of Siglecs can involve the Sias (13) and/or a specific cell surface-anchored protein, the β-protein (14). Because Siglec-13 and -17 could respond to a bacterial product (LPS) (*SI Appendix*, Fig. S5), we hypothesized that they would also interact with certain important sialylated human pathogens. Indeed, the extracellular domain of Siglec-13 bound to sialylated GBS A909 but not the nonpathogenic Gram-positive bacteria *Lactococcus lactis* (Fig. 4*A*). Extracellular domains of both Siglec-13 and -17 also bound to the sialylated Gram-negative pathogen *Escherichia coli* K1 (another leading cause of sepsis and meningitis in human newborns) but not nonpathogenic *E. coli* K-12 (Fig. 4*A*). Interestingly, although some Sia-dependent binding is observed, these interactions prominently involved trypsin-sensitive protein–protein interactions (Fig. 4 and *SI Appendix*, Fig. S6). With GBS, analysis of an isogenic bacterial mutant identified the likely binding partner for Siglec-13 as the GBS β-protein (Fig. 4*C*), an interaction previously shown to suppress human leukocyte responses through Siglec-5 (14).

Reduced Intracellular TNF in Siglec-13–Transfected RAW 264.7 Cells in Response to the Bacterial Infection. Semistably transfected RAW264.7 cells with Siglec-13 cDNA in pcDNA3.1 were acquired as mentioned above. Cells were infected with *E. coli* K1 or GBS A909 for 1 h at low multiplicity of infection (MOI; 0.6 and 0.1, respectively), and the level of intracellular TNF was then measured by an APC (Allophycocyanin) rat anti-mouse TNF antibody for all of the infected cells. Interestingly, compared with the mock control, the Siglec-13–transfected cells showed reduced intracellular TNF (Fig. 5 *A* and *B*). In this regard, it is known from work on other DAP12-interacting proteins that engagement of corresponding receptors can mediate either activating responses (through classical ITAM-Syk kinase signaling) or paradoxical inhibitory responses [ITAMi through Src kinase and Src homology phosphatase (SHP-1)] (18, 19). Thus, pathogenic bacteria might have been taking advantage of the cell surface-expressed Siglec-13 in ancestral hominins to dampen host innate immune cell responses and facilitate infection. Similar studies with Siglec-17 were not possible, because semistable expression in these cells resulted in markedly retarded growth.

Attempting to further recreate ancestral interactions of cells expressing Siglec-13 and -17 with human pathogenic bacteria, we tried to generate stable macrophage cell lines expressing them. However, compared with control-transfected cells, those expressing Siglec-13 or -17 grew very slowly in culture, and we could not generate long-term stable lines in either mouse RAW264.7 or human THP-1 macrophage cell lines. Similar difficulties were encountered with making stable transfectants of Siglec-17 in the human NK cell line NK-92. Apparently, sustained expression of these molecules is toxic to these cell types. Given the expression of Siglec-13 on chimpanzee monocytes, we assume that ancestral hominin macrophages were able to tolerate expression of this Siglec. Regardless, it is reasonable to suggest that Siglec-13–positive monocytes and/or Siglec-17–positive NK cells on ancestral innate immune cells may have influenced interactions with pathogenic bacteria and/or have had toxic effects.

Timing of *SIGLEC* Pseudogenization Events During Hominin Evolution. The two hominin pseudogenization events occurred after a common ancestor with the chimpanzee, but predated the common origin of modern humans. Fresh analysis of genomic sequence data from a newly obtained well-preserved Neanderthal sample from Monti Lessini (MLS3) (*SI Appendix*, see ancient DNA analysis) showed that the pseudogenized allele of *SIGLEC17*

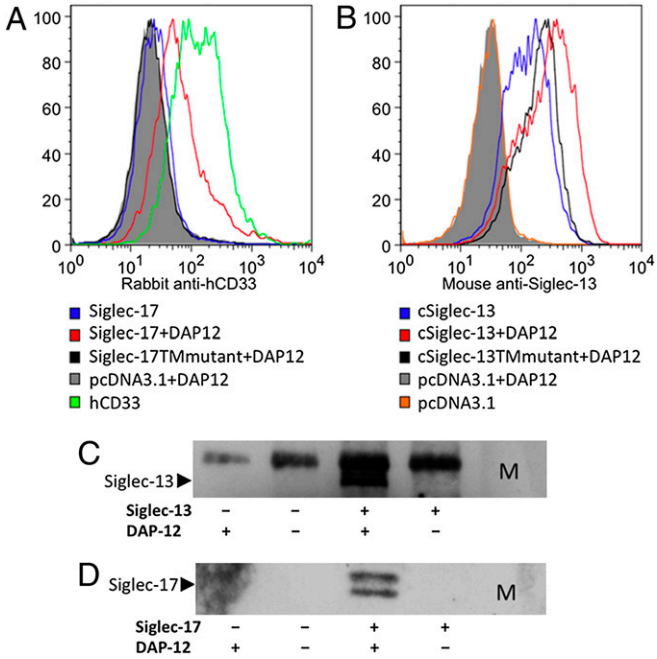


Fig. 3. Importance of DAP12 for optimal surface expression of Siglec-13 and -17. (*A*) 293T cells were transiently transfected with a Siglec-17 cDNA in pcDNA3.1 with or without cotransfection with FLAG-tagged DAP12. Human CD33-transfected cells were used positive controls for detection by a rabbit anti-human CD33 antibody, which partially cross-reacts with human Siglec-17. Cells cotransfected with pcDNA3.1 and DAP12 were used as a negative control. Fluorescence was measured after staining with rabbit anti-human CD33 and then Alexa Fluor 647 donkey anti-rabbit IgG. The Siglec-17TMmutant in pcDNA3.1 was made from Siglec-17–pcDNA3.1 by introducing a K253A mutation. Cotransfected pIRES2-EGFP (Clontech) was used to gate positively transfected cells. (*B*) The 293T cells were transiently transfected with a Siglec-13 cDNA in pcDNA3.1 with or without cotransfection with FLAG-tagged DAP12. Controls were as in *A*. Fluorescence was measured after staining with mouse anti–Siglec-13 and then Alexa Fluor 647 goat anti-mouse IgG. The cSiglec-13TMmutant in pcDNA3.1 was made from cSiglec-13–pcDNA3.1 by introducing a K352A mutation. Cotransfected pIRES2-EGFP (Clontech) was used to gate positively transfected cells. (*C* and *D*) 293T cells transiently transfected with cDNAs for chimpanzee Siglec-13 (*C*) or human Siglec-17 (*D*) with/without DAP12 were lysed. M2 agarose beads were used to pull down FLAG tagged DAP12. Mouse anti-Siglec-13 or rabbit anti-human CD33 (which cross reacts with Siglc-17) were used in Western blots followed by HRP conjugated secondary antibodies, as shown in *C* and *D*, respectively. The upper band on the coimmunoprecipitation is nonspecific because of the use of M2 beads carrying a mouse antibody (*C*). M indicates All Blue protein standard (BIO-RAD).

was already present before the population divergence of Neanderthals and modern humans (between 270,000 and 440,000 y ago) (26). The same genotype was noted in published Neanderthal genomic sequences. Whether *SIGLEC13* was also deleted from the Neanderthal genome could not be determined with certainty. However, we found no evidence for the chimpanzee version of the gene in the Neanderthal sample analyzed here (*SI Appendix*) or the published Neanderthal genome sequence. The recently published Denisovan genome sequence (27) also showed evidence of the modern human versions of *SIGLEC17P* and the possible deletion of *SIGLEC13*.

Taken together, the data indicate that both the *SIGLEC13* deletion allele and the modern human *SIGLEC17P* allele were already present in the common ancestral populations of Neanderthals, Denisovans, and humans. However, other similar *SIGLEC* genes (*SIGLEC14* and *SIGLEC16*) show polymorphic pseudogenization in modern humans, with moderate frequency persistence of the intact functional allele in all human populations (15, 16). Initial pseudogenization of *SIGLEC16* was estimated to occur at least 3 Mya (20). Given the small

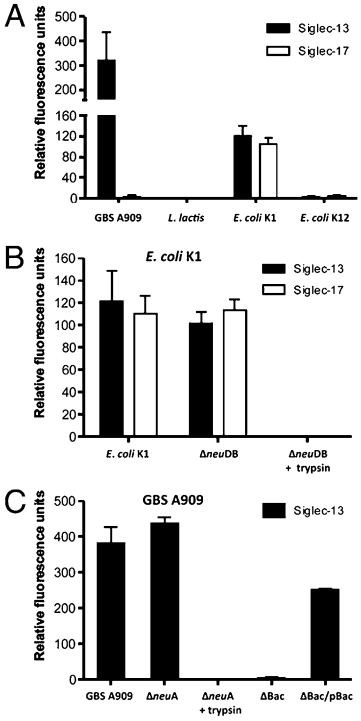


Fig. 4. Chimpanzee Siglec-13 or resurrected Siglec-17 interacts selectively with bacterial pathogens. (*A*) Chimpanzee Siglec-13-Fc or resurrected human Siglec-17-Fc (with Arg) chimeras were immobilized to ELISA wells by protein A, and binding of FITC-labeled sialylated strains GBS A909 (serotype Ia) or *E. coli* K1 RS218 (Str^R) was studied. Negative controls were *L. lactis* and non-encapsulated laboratory *E. coli* K-12 strain DH5α. (*B*) Binding of *E. coli* K1 strain RS218 (Str^R), isogenic Sia-deficient *E. coli* K1 Δ*neu*DB, or *E. coli* K1 Δ*neu*DB pretreated with trypsin was studied as in *A*. (*C*) Binding of GBS A909 serotype Ia, isogenic Sia-deficient GBS Δ*neu*A, or GBSΔ*neu*A pretreated with trypsin was studied as in *A*. StrainsΔBac (lacking β-protein) or the plasmid complemented mutant ΔBac + pBac were also studied. All values are means from three independent experiments ± SD.

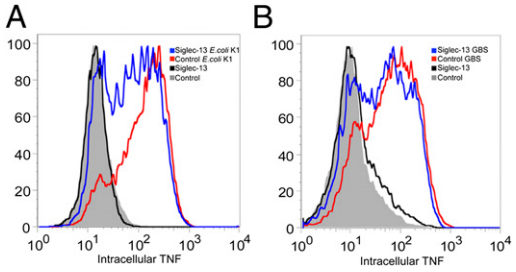


Fig. 5. Reduced intracellular TNF of Siglec-13–transfected RAW264.7 cells in response to bacterial pathogen infection; 500,000 cells semistably transfected with Siglec-13 or vector only and selected with G418 for 3 wk were seeded in a 12-well plate. The next day, 2 h before infection, cells were washed three times with HBSS, and regular culture medium without G418 added. Cells were infected with bacterial pathogens at an appropriate MOI for 1 h at 37 °C. The BD Fixation and Permeabilization Solution Kit (555028) with APC rat anti-mouse TNF-α was used to detect the intracellular TNF level using the recommended protocol. Cells without bacterial infection were used as controls. Cells transfected with pcDNA3.1 vector were used as a control. (*A*) Cells infected by *E. coli* K1 RS218 (Str^R) at MOI = 0.6. (*B*) Cells infected by GBS A909 (serotype Ia) at MOI = 0.1.

population of Neanderthals and Denisovans analyzed and the limited quality of these data, we cannot be certain regarding the timing of fixation of the *SIGLEC13* and *SIGLEC17* pseudogenes in hominin populations. Nevertheless, we hypothesize that expression of these genes became detrimental to survival under the selective pressure of pathogenic bacteria that were able to bind to the Siglecs and subvert their homeostatic immune functions. Another (not mutually exclusive) selection pressure for Siglec elimination could have been toxic overactivation of the immune system. To seek evidence for these hypotheses, we looked for residual signatures of selection surrounding these loci.

Genomic Evidence for Selection at the *SIGLEC13* and *SIGLEC17* Loci. When positive directional selection involves specific genes or pseudogenes, the genomic regions encompassing such loci can show limited variation in the time period immediately after (34, 35) depending on the type of selection and whether the allele has reached fixation. This kind of signature of selection will be eroded over time by additional random mutations and/or recombination. The deepest time at which such signatures can still be confidently detected is thought to be 5,000–10,000 generations or about 100,000–200,000 y (34, 35). This depth of time happens to be similar to the depth of time proposed for the origin of modern humans (21, 22, 25). Thus, it is impossible to conclusively prove a classical "selective sweep" before the origin of modern humans ~200,000 y ago. However, we decided to look for any residual evidence of selection surrounding these two loci compared with data for adjacent genes.

We, indeed, found low π- (nucleotide diversity) and θ- (nucleotide polymorphism) values for *SIGLEC17P* and the *SIGLEC13* flanking regions relative to adjacent loci such as *SIGLEC8* and *SIGLEC10*. Using GENECONV, we did not find any evidence that either locus was involved in gene conversions that may explain our observations. Furthermore, the estimated *H* values of Fay and Wu of human *SIGLEC13* flanking region and *SIGLEC17* are significantly negative, and the DH and DHEW tests (36, 37) show significant results for *SIGLEC17P* (*SI Appendix*, Table S4). However, not all of the statistics significantly rejected neutral evolution of two loci in our coalescent analysis (*SI Appendix*, Table S4). Notably, relatively low polymorphisms were also observed in *SIGLEC7*, which is ~10 kb away from *SIGLEC17P*. Like *SIGLEC17P* (*SI Appendix*, Table S4), *SIGLEC7* also showed significant deviation from neutrality in some statistical tests. However, *SIGLEC9*, another adjacent locus at a comparable distance, showed neither low polymorphisms (*SI Appendix*, Table S2) nor statistical significance in any of the tests. Thus, *SIGLEC7* could not be the locus of selection. In addition, an HKA (Hudson–Kreitman–Aguadé) test (38) between *SIGLEC17P* and *SIGLEC7* loci indicated the significantly lower polymorphism in *SIGLEC17P* compared with *SIGLEC7* with the chimpanzee sequence as an outgroup ($P < 0.05$), supporting the notion that *SIGLEC17P* might be the center of the proposed ancient selective event.

Interestingly, the resurrected Siglec-17 also shows a low nonsynonymous/synonymous rate ratio (dN/dS) that is similar to most other currently functional Siglecs, suggesting that it was subject to purifying selection before its pseudogenization (*SI Appendix*, Fig. S7). Notably, with the exception of the human-universal single base pair deletion, the *SIGLEC17* ORF remains conserved in all 28 HapMap humans studied, suggesting inadequate time for accumulation of other random mutations. We, of course, cannot rule out the possibility that this pseudogene is still undergoing purifying selection because of some other unknown function (39) (e.g., as a small RNAi-altering gene expression (40).

Although no single test can be conclusive at this depth of evolutionary time, our collective data indicate a residual signature of ancient selection forces acting on both of these loci, which must have predated the common origin of modern humans. However, such signatures should have faded in 100,000–200,000 y (34, 35). Thus, it is reasonable to speculate that positive selection on these pseudogenization events may have been involved in population bottlenecks close to the origin of modern humans.

Dating the Selection on the Inactivation of *SIGLEC13* and *SIGLEC17*. The hominin-specific events inactivating *SIGLEC13* and *SIGLEC17*

likely occurred first in the common ancestral populations of modern humans, Neanderthal, and Denisovans. In keeping with this finding, coalescence analysis estimated the time of the most recent common ancestor of identified haplotypes at 800,000–900,000 for the *SIGLEC13* deletion locus and *SIGLEC17P* (*SI Appendix*). At first glance, this timing may seem at odds with aforementioned signals of selection at both loci, which should have been completely erased over such long periods of time. However, we can suggest a scenario consistent with all of the data—that some active alleles of these genes persisted in the common ancestral population of modern humans, Neanderthal, and Denisovans until selection eliminated them at some point close to the common origin of modern humans. To seek additional evidence for this possibility, we calculated dates for selection acting on these events (*SI Appendix*). Approximate times were calculated as ∼105,000 y for *SIGLEC17P* and ∼46,000 y for *SIGLEC13*. These numbers are approximations, and the method does not allow error estimations. Because both pseudogenization events are universal to all modern human populations, they must actually date back to at least the common origin of modern humans ∼100,000–200,000 y ago. Regardless of exact timing, these data provide support for selection acting on these pseudogenes close to the common origin of modern humans.

Bacterial Pathogens as Selective Agents in Hominin Evolution? Independent of exactly when fixation of these pseudogenes occurred in ancestral hominins, the question arises as to what selective forces were involved. Alteration of innate immune defense against invasive human neonatal pathogens such as GBS and *E. coli* K1 would exert a powerful selection pressure on reproductive success, and other prevalent microbial pathogens could also have subverted Siglec signaling to promote infection. Thus, in addition to other extant theories about the origin of modern humans, interactions of infectious agents with the innate immune system of early humans should be examined as a source of potential selection. After all, infectious agents are already widely recognized as selective agents to explain human polymorphisms (e.g., the role of malaria in selecting for the sickle cell hemoglobin trait). Alternative (but not mutually exclusive) hypotheses include Siglec-mediated overactivation of immune cells and/or changes in commensal–host interactions, which could have compromised reproductive success or the ability to care for the young. Additional studies are needed to further investigate the theoretical role of bacterial selection pressure during human origins. Strong selection by pathogens could result in severe population restrictions and bottlenecks, such as seen in hominin origins. If so, we may need to consider the possibility of an "infectious origin" of modern humans.

Materials and Methods

DNA samples and population groups are described in *Results*. PCR, RT-PCR, sequencing details, and details of flow cytometry for detection of Siglecs or intracellular TNF expression are in *SI Appendix*. Cells transfected with expression constructs for Siglec-13 or -17 with or without DAP12–FLAG were analyzed by coimmunoprecipitation and Western blotting (*SI Appendix*). Details of array fabrication and binding assays, transfection of Siglecs in mouse Raw264.7 cells, molecular and cellular assays, and human population genetic analysis are in *SI Appendix*. Interaction of Siglec-Fcs with bacteria was studied as described (33) with minor modifications (*SI Appendix*).

ACKNOWLEDGMENTS. We thank Pascal Gagneux for valuable comments. This work was funded by grants listed in the *SI Appendix*.

1. Schauer R (2009) Sialic acids as regulators of molecular and cellular interactions. *Curr Opin Struct Biol* 19:507–514.
2. Chen X, Varki A (2010) Advances in the biology and chemistry of sialic acids. *ACS Chem Biol* 5:163–176.
3. Schwarzkopf M, et al. (2002) Sialylation is essential for early development in mice. *Proc Natl Acad Sci USA* 99:5267–5270.
4. Severi E, Hood DW, Thomas GH (2007) Sialic acid utilization by bacterial pathogens. *Microbiology* 153:2817–2822.
5. Varki A (2010) Colloquium paper: Uniquely human evolution of sialic acid genetics and biology. *Proc Natl Acad Sci USA* 107(Suppl 2):8939–8946.
6. Lewis AL, et al. (2009) Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. *Proc Natl Acad Sci USA* 106:13552–13557.
7. Cao H, Crocker PR (2011) Evolution of CD33-related siglecs: Regulating host immune functions and escaping pathogen exploitation? *Immunology* 132:18–26.
8. Varki A, Angata T (2006) Siglecs—the major subfamily of I-type lectins. *Glycobiology* 16:1R–27R.
9. von Gunten S, Bochner BS (2008) Basic and clinical immunology of Siglecs. *Ann N Y Acad Sci* 1143:61–82.
10. Lopez PH, Schnaar RL (2009) Gangliosides in cell recognition and membrane protein regulation. *Curr Opin Struct Biol* 19:549–557.
11. 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.
12. Jones C, Virji M, Crocker PR (2003) Recognition of sialylated meningococcal lipopolysaccharide by siglecs expressed on myeloid cells leads to enhanced bacterial uptake. *Mol Microbiol* 49:1213–1225.
13. Carlin AF, et al. (2009) Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. *Blood* 113:3333–3336.
14. Carlin AF, et al. (2009) Group B Streptococcus suppression of phagocyte functions by protein-mediated engagement of human Siglec-5. *J Exp Med* 206:1691–1699.
15. Angata T, Hayakawa T, Yamanaka M, Varki A, Nakamura M (2006) Discovery of Siglec-14, a novel sialic acid receptor undergoing concerted evolution with Siglec-5 in primates. *FASEB J* 20:1964–1973.
16. Cao H, et al. (2008) SIGLEC16 encodes a DAP12-associated receptor expressed in macrophages that evolved from its inhibitory counterpart SIGLEC11 and has functional and non-functional alleles in humans. *Eur J Immunol* 38:2303–2315.
17. Lanier LL (2009) DAP10- and DAP12-associated receptors in innate immunity. *Immunol Rev* 227:150–160.
18. Turnbull IR, Colonna M (2007) Activating and inhibitory functions of DAP12. *Nat Rev Immunol* 7:155–161.
19. Pfirsch-Maisonnas S, et al. (2011) Inhibitory ITAM signaling traps activating receptors with the phosphatase SHP-1 to form polarized "inhibisome" clusters. *Sci Signal* 4: ra24.
20. Wang X, et al. (2012) Evolution of Siglec-11 and Siglec-16 genes in hominins. *Mol Biol Evol*, 10.1093/molbev/mss077.
21. Campbell MC, Tishkoff SA (2008) African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 9:403–433.
22. Jacobs Z, et al. (2008) Ages for the Middle Stone Age of southern Africa: Implications for human behavior and dispersal. *Science* 322:733–735.
23. Gunz P, et al. (2009) Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. *Proc Natl Acad Sci USA* 106: 6094–6098.
24. White TD (2009) Human origins and evolution: Cold Spring Harbor, deja vu. *Cold Spring Harb Symp Quant Biol* 74:335–344.
25. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
26. Green RE, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328: 710–722.
27. Reich D, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053–1060.
28. Briggs AW, et al. (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325:318–321.
29. Angata T, Margulies EH, Green ED, Varki A (2004) Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. *Proc Natl Acad Sci USA* 101:13251–13256.
30. Schmid CW (1998) Does SINE evolution preclude Alu function? *Nucleic Acids Res* 26: 4541–4550.
31. Tishkoff SA, et al. (2009) The genetic structure and history of Africans and African Americans. *Science* 324:1035–1044.
32. Humphrey MB, et al. (2006) TREM2, a DAP12-associated receptor, regulates osteoclast differentiation and function. *J Bone Miner Res* 21:237–245.
33. 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.
34. Przeworski M (2003) Estimating the time since the fixation of a beneficial allele. *Genetics* 164:1667–1676.
35. Sabeti PC, et al. (2006) Positive natural selection in the human lineage. *Science* 312: 1614–1620.
36. Zeng K, Fu YX, Shi S, Wu CI (2006) Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* 174:1431–1439.
37. Zeng K, Shi S, Wu CI (2007) Compound tests for the detection of hitchhiking under positive selection. *Mol Biol Evol* 24:1898–1908.
38. Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159.
39. Balakirev ES, Ayala FJ (2003) Pseudogenes: Are they "junk" or functional DNA? *Annu Rev Genet* 37:123–151.
40. Wen YZ, et al. (2011) Pseudogene-derived small interference RNAs regulate gene expression in African *Trypanosoma brucei*. *Proc Natl Acad Sci USA* 108:8345–8350.

Wang et al. *PNAS* - SI Appendix

DETAILS OF ACKNOWLEDGMENTS (GRANT FUNDING)

This work was funded by grant 1P01HL107150 from the National Institutes of Health and by the Mathers Foundation of New York. Ancient DNA sequencing was supported by MIUR grant n 2008TEB8s_002 and FIRB grant n RBFR08U07M (to DC and ER). I.S. was initially supported through a postdoctoral fellowship from the University of California Institute for Mexico and the United States (UC-MEXUS) program.

Supplement to Author List

NISC Comparative Sequencing Program: Betty Benjamin, Robert Blakesley, Gerry Bouffard, Shelise Brooks, Grace Chu, Holly Coleman, Mila Dekhtyar, Michael Gregory, Xiaobin Guan, Jyoti Gupta, Joel Han, April Hargrove, Shi-ling Ho, Taccara Johnson, Richelle Legaspi, Sean Lovett, Quino Maduro, Cathy Masiello, Baishali Maskeri, Jenny McDowell, Casandra Montemayor, Betsy Novotny, Morgan Park, Nancy Riebow, Karen Schandler, Brian Schmidt, Christina Sison, Mal Stantripop, James Thomas, Meg Vemulapalli, and Alice Young.

Supplement to Material and Methods

Genomic DNA samples, PCR and sequencing. Genomic DNAs of 28 human individuals were acquired from NHGRI including 11 Africans, 9 Japanese and Chinese, and 8 Northern Europeans. The presence of human *SIGLEC17P* and potential *SIGLEC17* alleles was checked in 228 African individuals by PCR with primers SP3F: 5'-GTTTGAGGTTCCTCTTCTGTG-3'and SP3R: 5'-TGAGCCTGACGTGCTTTATTC-3' followed by one-direction DNA sequencing. The missing of *SIGLEC13* in humans was checked in 230 African individuals (Table S1) by PCR following general protocol using two pairs of primers. Primer S13F1: 5'-

TGGGGTCTGGATCCCACGGTAAAGGG-3' and S13R1: 5'-GATGCCACTGCACTGTCACTAGAACTC-3' can only amplify the missing allele (2748bp) since the WT allele was too long to be amplified. Primer S13Int1: 5'-GATTACCCAGGAGGCGGAATCGACCC-3' was designed from the Siglec-13 deletion region and gave a PCR product of 1156bp when paired with S13R1 if the WT allele is present, but gave none when missing allele is present. All PCR reactions were done using 100 ng genomic DNA. The Expand Long Template PCR System (Roche) along with buffer 3 was used. The reaction conditions are as follows; 1 cycle (94°C 2 minutes); 10 cycles (94°C 10 seconds, 62°C 30 seconds (-1°C/cycle), 68°C 10 minutes); 20 cycles (94°C 15 seconds, 52°C 30 seconds, 68°C 10 minutes (+20 seconds/cycle)); 1 cycle (72°C 7 minutes). SNPs data of human *SIGLEC17P* were acquired from NHGRI. Six regions surrounding the missing spot of human *SIGLEC13* were amplified using PCR SuperMix high fidelity (Invitrogen) following the recommended protocol. RepeatMasker (http://www.repeatmasker.org/) was used to avoid the repetitive elements when designing primers. Each PCR product is 800-1000 bp long. PCR primers are shown in Table S3. PCR products were directly sequenced by Genewiz, Inc. DNA sequences were assembled in Sequencher 4.10.1 (Gene Codes Corporation).

Genomic sequences containing *SIGLEC13* locus. Chimpanzee genomic sequence containing *SIGLEC13* locus was extracted from the BAC sequence submitted by NHGRI (GenBank accession number AC132069). Baboon BAC sequence (GenBank accession number AC130272) was used to obtain the genomic sequence containing baboon *SIGLEC13* locus. The genomic sequence containing rhesus monkey *SIGLEC13* locus was obtained from the rhesus monkey genome NCBI build 1.2. RepeatMasker (http://www.repeatmasker.org/) was used to detect the repetitive elements.

Phylogenetic analysis. Human *SIGLEC17P* cDNA sequence based on BC041072 clone was used as the query to blast against chimpanzee, orangutan, rhesus macaque, and marmoset genomes. Intact ORF (Open Reading Frame) of marmoset *SIGLEC17* was predicted by GENSCAN (http://genes.mit.edu/GENSCAN.html) and evaluated by Wise2 (EMBL-EBI) through comparison of resurrected human Siglec-17 to marmoset genomic DNA sequence. DNA or protein sequences alignment was conducted by CLUSTALW in MEGA4 (1). Neighbor-Joining tree of DNA sequences of *SIGLEC17/*P and *SIGLEC3* was reconstructed in MEGA4. Bootstrap values of 1000 replicates were estimated for all of the internal branches.

Human and Chimpanzee PBMC collection. Chimpanzee blood samples were collected in EDTA-containing tubes at the Yerkes National Primate Center (Atlanta, GA) and shipped on ice overnight. With approval from the University of California Institutional Review Board, blood from healthy human volunteers was similarly collected into EDTA-containing tubes and then stored overnight on ice or start processing immediately. Peripheral Blood Mononuclear Cells (PBMCs) were isolated by centrifugation using the Ficoll-Paque PLUS (GE Healthcare) following established protocol (2)

Siglec-13 expression on chimpanzee monocytes. Chimpanzee and human PBMCs were collected as above. Cell surface expression of Siglec-13 was probed by Mouse monoclonal anti-Chimp-Siglec-13 antibody (see below) and then stained with Alexa Fluor 647 goat anti-mouse IgG (Invitrogen). Cells were then washed and labeled for CD14 (a marker for peripheral blood monocytes) using FITC Mouse anti human-CD14 IgG (BD Pharmagen). Antibody against human CD14 is able to react equally well against chimpanzee CD14.

qRT-PCR of *SIGLEC17P* transcript in human Natural Killer cells. Human PBMC was collected freshly as above. Natural killer cells and T cells were enriched from PBMC using

Human NK cell enrichment kit and CD3 positive selection kit (Easysep), respectively. Total RNA was purified from extracted NK cells, T cells, total PBMC, and total PBMC without T cells using RNeasy plus Mini kit following the recommended protocol (Qiagen). Half microgram of total RNA was used for cDNA synthesis using the QuantiTect Reverse Transcription kit (Qiagen). The cDNA was stored at -80°C until use. Real-time PCR was performed using C1000 thermal cycler (BIO-RAD) with QuantiTect SYBR green PCR master mix (Qiagen) according to the manufacturer's protocol. The PCR condition was 15min at 95°C followed by 40 cycles of 15sec at 95°C, 30sec at 55°C and 30sec at 70°C. Primer mix of Hs_SIGLECP3_1_SG (Qiagen) was used for qPCR. All qPCR reactions were performed in triplicate. GAPDH mRNA levels were used to normalized the relative expression levels of target mRNAs.

Preparation of human Siglec-17-pcDNA3.1, chimpanzee Siglec-13-pcDNA3.1, and human DAP12-FLAG-pcDNA3.1 constructs. Clone BC041072 (**IMAGE:** 5484659) was purchased from Invitrogen. The full-length potential coding region of *SIGLEC17P* was amplified by PCR from BC041072 clone using PfuUltra high-fidelity polymerase (Stratagene) following recommended protocol with primer SigP3fullF (5'-AGATTATCTAGAGCCACCATGCTGCCGCTGCTGCTGCCG-3'; Kozak sequence underlined, preceded by an *Xba*I site) and primer SigP3fullR (5'-TCGTGCGATATCTCACCGCCTTCTTCTTGTGGAAACTG-3'; *EcoRV* site underlined). PCR products were digested with *XbaI* and *EcoRV* and subcloned into *XbaI-EcoRV* sites of pcDNA3.1 (-) (Invitrogen). Resurrected Siglec-17-pcDNA3.1 construct was prepared by introducing a single nucleotide ΦG' into Siglec-17p-pcDNA3.1 construct using QuikChange II XL site-directed mutagenesis kit (Stratagen) following the manufacturer's instructions, using primer pair SigP3mutant#1 (5'- CTGCCCCTGCTGTGGGCAGGGGCCCTCGCTCAGGATGC -3';

Inserted ΦG‘ underlined) and SigP3mutant#2 (complementary to SigP3mutant#1). Siglec-17-pcDNA3.1 point mutant (W120R), which has Arginine (responsible for Sialic acid binding) restored was prepared by introducing point mutation into Siglec-17-pcDNA3.1 construct by QuikChange II XL site-directed mutagenesis kit, using primer pair SigP3Arg#1 (5’-GGTTCATACTTCTTTCGGGTGGCGAGAGGAAG -3’; Arg codon underlined) and SigP3Arg#2 (complementary to SigP3Arg#1). Chimpanzee Siglec-13-pcDNA3.1 was kindly provided by Takashi Angata. A human DAP12 clone in vector pCMV-SPORT6 was purchased from Open Biosystems (Hunsville, AL). DAP12 was amplified using 5’-CCCCTCGAGCCACCATGGGGGGACTTG AACCCTG-3’; (*XhoI* site underlined) and 5’-GGGGAATTCTTA *CTTGTCATCGTCGTCCTTGTAGTC* TTTGTAATACGGCCTCTGTGTG-3’ (*EcoRI* site underlined, complementary FLAG tag-coding sequence is italicized). The amplified PCR product was digested using *XhoI* and *EcoRI* and subcloned into pcDNA3.1(-).

The sequences of all constructs have been verified by DNA sequencing.

Preparation of human Siglec-17-Fc and chimpanzee Siglec-13-Fc expression constructs. The cDNA region of human *SIGLEC17* encoding the extracellular domain without signal peptide was amplified from Siglec-17-pcDNA3.1 construct or its W120R mutant construct by PCR using PfuUltra high-fidelity polymerase following recommended protocol with the forward primer 5’-AAGCTTCAGGATGCAAGATTCCGG-3’ (*HindIII* site underlined) and the backward primer 5’-TCTAGAGGAGACACTGAGCTG-3’ (*XbaI* site underlined). PCR products were digested with *XbaI* and *HindIII* and subcloned into *XbaI*-*HindIII* sites of Signal pIgplus MCS vector (Lab storage), which carries Siglec-3 signal peptide. Upon mammalian cell transfection, the resulting construct expressed a recombinant soluble Siglec-17-Fc protein with or without arginine restored

(a fusion protein of the extracellular two Ig-like domains of human Siglec-17 and human IgG Fc fragment).

Chimpanzee Siglec13-EK-Fc/pcDNA3.1 was kindly provided by Takashi Angata. This construct expressed a recombinant soluble Chimpanzee Siglec-13 extracellular domain including the signal peptide and two Ig-like domains and human IgG Fc fragment. The expression construct for Siglec13-EK-Fc/pcDNA3.1 point mutant (R120K) was prepared by introducing point mutation into Siglec-13-Fc construct using primer pair CSig13ArgMutant#1 (5'-CAGTGGTTCGTACTTCTTTAAGGTGGAGGAAACAATGATG -3'; Lys codon underlined) and CSig13ArgMutant#2 (complementary to CSig13ArgMutant#1). Upon mammalian cell transfection, the resulting constructs expressed recombinant soluble Siglec-13-Fc proteins with or without the arginine mutated (a fusion protein of the extracellular two Ig-like domains of chimpanzee Siglec-13 and human IgG Fc fragment). The fusion proteins were prepared by transient transfection of Chinese hamster ovary TAg cells with Siglec-Fc constructs following the established protocol (3). Siglec-Fc proteins were purified from culture supernatant by adsorption to protein A-Sepharose (GE Healthcare).

Purification of Mouse monoclonal anti-Siglec-13. The two Ig-like domains were cleaved from the Siglec-13-Fc fusion protein using enterokinase and separated from the Fc domain using IgG Sepharose (Amersham), which preferentially bound to the Fc region. The purified two-domain peptide of Siglec-13 was used to immunize mice. The immunization, fusion and subcloning were conducted by Promab Biotechnologies following the recommended protocol. Multiple subcloning supernantants were tested for binding of Siglec-13 by ELISA and Immunohistochemistry. The specificity of the antibody was evaluated by its binding to human

Siglec-6, 9, 11, and 12. The supernantant of one of the two best clones which bound to Siglec-13 specifically with no reactivity to other Siglecs were used in our study.

293T Cell transfection and flow cytometry: Human 293T cells were transfected with Siglec-pcDNA3.1 or cotransfected with DAP12-FLAG-pcDNA3.1 (1:1 ratio) vector using Lipofectamine2000 following the recommended protocol (Invitrogen). Transfected cells with both pcDNA3.1 and DAP12 were used as a negative control in flow cytometry. Rabbit polyclonal anti-CD33 antibody (Novus biological) was used to probe the cell surface expression of Siglec-17, and fluorescence was measured after staining with Alexa Fluor 647 donkey anti-rabbit IgG (Invitrogen). The quality of flow cytometry results was controlled using 293T cells transfected with human CD33 (*SIGLEC3*). Mouse monoclonal anti-Chimp-Siglec-13 antibody was used to probe the cell surface expression of Siglec-13, and fluorescence was measured after staining with Alexa Fluor 647 goat anti-mouse IgG.

Co-immunoprecipitation of chimpanzee Siglec-13 or human Siglec-17 with human DAP12. 293T cells were plated in 6-well plates. Next day, Siglec-13-pcDNA3.1 or Siglec-17-pcDNA3.1 was transfected into these cells either individually or together with DAP12-FLAG-pcDNA3.1 using Lipofectamine 2000 (Invitrogen) following the manufacturer recommendation. 293T cells transfected with DAP12 only or together with pcDNA3.1 were used as negative controls. Twenty four hours after transfection, cells from each well was lysed in 300 μl of 10 mM Tris HCl pH 7.4, with 1% Triton X-100 and 150 mM NaCl and 1mM EDTA with Protease Inhibitor Cocktail Set III (EMD Biosciences, La Jolla, CA). 200ug of cell lysate was then incubated with anti-FLAG conjugated M2-agarose beads (Sigma). The immunoprecipitated materials were loaded on SDS-PAGE gel, blotted on nitrocellulose membrane (Bio-Rad) and probed with anti-Siglec-13 mouse monoclonal antibody or anti-CD33 rabbit polyclonal antibody (Novus

biological) followed by anti-mouse/rabbit-HRP (Jackson laboratories). The presence of DAP12 was verified by rabbit anti-FLAG (Sigma).

Sialoglycan-microarray fabrication. Arrays were fabricated by KAMTEK Inc. (Gaithersburg, MD) in a manner similar to that already reported by us (4). Epoxide-derivatized Corning slides were purchased from Thermo Fisher Scientific (Pittsburgh, PA). Glycoconjugates were distributed into six 384-well source plates (12-16 samples per plate) using 4 replicate wells per sample and 20 µl per well (Version 12). Each glycoconjugate was prepared at 100 µM in an optimized print buffer (300 mM phosphate buffer, pH 8.4). To monitor printing quality, replicate-wells of human IgG (Jackson ImmunoResearch, at 150 µg/ml) and Alexa -555 (25ng/µl) in print buffer were used for each printing run. The arrays were printed with four SMP5B (Array It). These are quill pins with bubble uptake channel and gives spot diameter of ~160µm (glycan spots ~ 70 µm). Printing was done using VersArray ChipWriter Pro (Virtek/BioRad). There is one super grid cluster on X-axis and two on Y-axis generating 8 sub-array blocks on each slide. Blotting was done (5 blots/dip; Blot Time: 0.05 secs) in order to have uniform sample distribution. Each grid (sub-array) has 16 spots/row, 20 columns with spot to spot spacing of 225 µm. The humidity level in the arraying chamber was maintained at about 66% during printing. Printed slides were left on arrayer deck over-night, allowing humidity to drop to ambient levels (40-45%). Next, slides were packed, vacuum-sealed and stored in a desiccant chamber at RT until used. Slides were printed in one batch of 45 slides.

Array binding assays. Slides were incubated for 1 hour in a staining dish with 50°C pre-warmed blocking solution (0.05 M Ethanolamine in 0.1 M Tris pH 9) to block the remaining reactive groups on the slide surface, then washed twice with 50°C pre-warmed dH_2O. Slides were centrifuged at 200×*g* for 3 min. Dry slides were fitted with ProPlate® Multi-Array slide

module (Invitrogen) to divide into the subarrays then blocked with 200 µl/sub-array blocking solution 2 (PBS/OVA, 1% w/v Ovalbumin in PBS pH 7.4) for 1 hour at room temperature (RT), with gentle shaking. Next, the blocking solution was aspirated and diluted samples (Siglec-Fc at 7.5 µg/ml in PBS/OVA, 200 µl/sub-array) were added to each slide and allowed to incubate with gentle shaking for 2 h at RT. Slides were washed three times with PBST (PBS, 1% Tween) then with PBS for 10 min/wash with shaking. Bound antibodies were detected by incubating with 200 µl/sub-array of the relevant fluorescent-labeled secondary antibody Cy3-anti-human IgG (1.5 µg/ml) diluted in PBS at RT for 1 hour. Slides were washed three times with PBST (PBS, 1% Tween) then with PBS 10 min/wash followed by removal from of ProPlate® Multi-Array slide module and immediately dipping slide in a staining dish with dH_2O for 10 min with shaking, then centrifuged at $200 \times g$ for 3 min. Dry slides were vacuum-sealed and stored in dark until scanning the following day.

Array slide processing. Slides were scanned at 10 μm resolution with a Genepix 4000B microarray scanner (Molecular Devices Corporation, Union City, CA) using gain 450. Image analysis was carried out with Genepix Pro 6.0 analysis software (Molecular Devices Corporation). Spots were defined as circular features with a variable radius as determined by the Genepix scanning software. Local background subtraction was performed.

Semi-stable Transfection of Siglecs in Raw264.7 cells. Mouse macrophage cell line RAW264.7 is purchased from ATCC. Cells were transfected with Siglec-13-pcDNA3.1, Siglec-17-pcDNA3.1, or pcDNA3.1 vector only using lipofectamine2000 (Invitrogen) following the recommended protocol. After 36 hours, 1.5 mg/ml G418 (Invitrogen) was added into the medium in order to select the positively transfected cells. Antibiotic containing medium was removed and replaced every 3-4 days. The surviving Siglec-transfected cells as well as the control cells

were used after three weeks selection for studies of LPS response and bacterial pathogen infection assay.

Intracellular TNF after exposure to a TLR ligand. Siglec-transfected or control transfected RAW264.7 cells were exposed to 0.1 ng/ml bacterial lipopolysaccharide (LPS, Invitrogen) for 2.5 hours at 37°C, with one microliter of BD GolgiStop protein transport inhibitor added. Cells were washed three times by staining buffer, permeabilized, and fixed by BD Cytofix/Cytoperm Fixation/Permeabilization kit. Intracellular TNF secretion was detected by APC rat anti-mouse TNF (BD) using flow cytometry.

Bacterial binding to Siglec-Fc molecules. The interaction of chimp Siglec-13-Fc or human resurrected Siglec-17-Fc with bacteria was determined using a previously described method with minor modifications (5). Immulon ELISA plates were coated with 0.025 mg/ml protein A (Sigma) in coating buffer (67 mM $NaHCO_3$, 33 mM Na_2CO_3, pH=9.6) overnight at 4ºC. Wells were washed and blocked with assay buffer (20 mM Tris pH=8.0, 150 mM NaCl, 1% BSA) for 1.5 h at 37ºC. Chimpanzee Siglec-13-Fc or resurrected human Siglec-17-Fc chimera (with or without Arginine) diluted in assay buffer were added to individual wells at 0.025 mg/ml for 2 h 37ºC. For *E. coli* K1 strain RS218 (Str^R), *E. coli* K1 Δ*neu*DB (Cm^R) (isogenic sialic acid-deficient) and *E. coli* K-12 strain DH5α, bacteria were grown to an OD_{600} of 0.6; for Group B *Streptococcus* (GBS) strain A909 (serotype Ia), GBS Δ*neu*A (isogenic sialic-acid deficient), GBS ΔBac (isogenic β-protein deficient) and *Lactococcus lactis*, bacteria were grown to an OD_{600} of 0.4. Bacteria were labeled with 0.1% FITC (Sigma) for 1 h 37ºC and then suspended at 1 x 10^7 cfu/ml in assay buffer and added to each well and centrifuged at 805 x g for 10 min. Bacteria were allowed to adhere for 15 min at 37ºC. The initial fluorescence was verified, wells were washed to remove unbound bacteria, and the residual fluorescence intensity (excitation, 485

nm; emission, 538 nm) was measured using a Spectra Max Gemini XS fluorescence plate reader (Molecular Devices). Trypsin treatment was done by incubating *E. coli* Δ*neu*DB or GBS Δ*neu*A in 0.5% trypsin (Sigma) + 10 mM EDTA in PBS for 30 min at 37ºC and washed 5 times with PBS as previously described (Carlin, A, et al 2009). Gram staining of trypsin-treated bacteria was conducted using crystal violet and intact bacterial cell walls were observed under microscopic examination.

Human population genetic analysis. The DHEW test (6) of human *SIGLEC17* sequences was conducted using the program dh.jar, kindly provided by Kai Zeng, University of Edinburgh. The chimpanzee sequence was used as the outgroup in computing DH and DHEW. P-values in the DHEW test were estimated using 100,000 replications of coalescent simulation. The coalescent analysis of the Tajima's D and Fay and Wu'H were conducted in DNAsp5.10 (7). For human *SIGLEC17P*, $u = 74/(6*10^6*2) = 0.616*10^{-5}$ per year per locus. 74 divergence sites were found between human and chimpanzee *SIGLEC17P* of 9047bp. For human *SIGLEC13*, $u = 63/(6*10^6*2) = 5.25*10^{-6}$ per year per locus. 63 divergence sites were found between human and chimpanzee *SIGLEC13* of 5145 bp. The population recombination rate of human *SIGLEC17P* region used was estimated to be $R = 4Nr = 4 \times 10{,}000 \times (3.3 \times 10^{-8} \times 9047) = 12$, where 3.3×10^{-8} is the pedigree-based recombination rate per generation per nucleotide at *SIGLEC17P* locus (8). The population recombination rate of human *SIGLEC13* flanking regions was estimated to be $R = 4Nr = 4 \times 10{,}000 \times (3.9 \times 10^{-8} \times 5145) = 8$, where 3.9×10^{-8} is the pedigree-based recombination rate per generation per nucleotide at *SIGLEC13* locus (8). PHASE 2.1.1 was used to reconstruct the haplotypes in both *SIGLEC17P* and *SIGLEC13* loci. The average divergence between two major haplotype sublineages of *SIGLEC17* is 11 out of 9047 bp and 9 out of 5145 bp for *SIGLEC13*. Therefore, the TMRCA was estimated to be 11/(2u) and

9/(2u), respectively. The nucleotide polymorphism (Theta), nucleotide diversity (Pi), and HKA test were estimated in DNAsp 5.10. The Dn (the nonsysnonymous substitution per nonsynonymous site) and Ds (the synonymous substitution per synonymous site) were calculated in MEGA4 using the modified Nei-Gojobori method.

Dating the Selection on the Inactivation of *SIGLEC13* and *SIGLEC17*. We used a published method (9) to approximately predict the time back to the selection on the deletion of *SIGLEC13* and pseudogenization of *SIGLEC17*. We assumed that mutation rate = $D_{H-C}/(6*10^6*2)$ (D is the divergence between human and chimpanzee orthologous regions) and that t = S/(nu) (where S: is the number of segregating sites; n is: the number of haplotypes included; and, u is: the neutral mutation rate of the locus). Fifty-four haplotypes were analyzed for *SIGLEC17* and 48 haplotypes for *SIGLEC13*. The number of segregating sites observed at each locus is 35 and 9, respectively.

Supplement to Figure Legend 2

Human tissues or cell types studied in Figure 2B are as follows (left to right on the X-axis): 721_B_lymphoblasts, Adipocyte, Adrenal Cortex, Adrenalgland, Amygdala, Appendix, Atrioventricular Node, BDCA4+_Dentritic Cells, Bone marrow, Bronchial EpithelialCells, Cardiac Myocytes, Caudate nucleus, CD105+_Endothelial, CD14+_Monocytes, CD19+_B Cells(neg._sel.), CD33+_Myeloid, CD34+, CD4+_Tcells, CD56+_NKCells, CD71+_EarlyErythroid, CD8+_Tcells, Cerebellum, Cerebellum Peduncles, Ciliary Ganglion, Cingulate Cortex, Colon, Colorectal Adenocarcinoma, Dorsal Root Ganglion, Fetal brain, Fetal liver, Fetal lung, Fetal Thyroid, Globus Pallidus, Heart, Hypothalamus, Kidney, Leukemia_chronic Myelogenous K-562, Leukemia_promyelocytic-HL-60, Leukemia lymphoblastic (MOLT-4), Liver, Lung, Lymph node, Lymphoma_Burkitts (Daudi),

Lymphoma_Burkitts (Raji), Medulla Oblongata, Occipital Lobe, Olfactory Bulb, Ovary, Pancreas, Pancreatic Islet, Parietal Lobe, Pineal_day, Pineal_night, Pituitary, Placenta, Pons, Prefrontal Cortex, Prostate, Retina, Salivary gland, Skeletal Muscle, Skin, Small_intestine, Smooth Muscle, Spinal cord, Subthalamic Nucleus, Superior Cervical Ganglion, Temporal Lobe, Testis, Testis Germ Cell, Testis Intersitial, Testis Leydig Cell, Testis SeminiferousTubule, Thalamus, Thymus, Thyroid, Tongue, Tonsil, Trachea, Trigeminal Ganglion, Uterus, Uterus Corpus, Whole Blood, Whole Brain.

REFERENCES for SI Appendix

1. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
2. Soto PC, Stein LL, Hurtado-Ziola N, Hedrick SM, Varki A (2010) Relative over-reactivity of human versus chimpanzee lymphocytes: implications for the human diseases associated with immune activation. *J Immunol* 184:4185–4195.
3. 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.
4. Padler-Karavani V, Hurtado-Ziola N, Pu M, Yu H, Huang S, Muthana S, Chokhawala HA, Cao H, Secrest P, Friedmann-Morvinski D *et al*. (2011) Human xeno-autoantibodies against a non-human sialic acid serve as novel serum biomarkers and immunotherapeutics in cancer. *Cancer Res* 71:3352–3363.

5. 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* 89:1231–1237.

6. Zeng K, Shi S, Wu CI (2007) Compound tests for the detection of hitchhiking under positive selection. *Mol Biol Evol* 24:1898–1908.

7. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.

8. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G *et al*. (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31:241–247.

9. Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol* 2:e286.

Figure S1. Human cDNA Clone BC041072. The location of single base-pair deletion "G" in the predicted ORF is indicated with an arrow. The start and stop codons of the resurrected ORF are indicated in boxes.

```
CCAAGATCTC ATGCTCCTCC CCACAGCCCT CTTCTCTGCT CACACAGGAA   50
GCCCAGGAAG CCTCTGCCTC AGAGATGCTG CCGCTGCTGC TGCCGCTGCC  100
CCTGCTGTGG GCAGGGCCCT CGCTCAGGAT GCAAGATTCC GGCTGGAGAT  150
GCCAGAGTCC GTGACGGTGC AGGAGGGTCT GTGCATCTTT GTGCACTGTT  200
CGGTCTTCTA CCTCGAGTAT GGCTGGAAAG ATTCTACCCC TGCTTATGGC  250
CACTGGTTCC GGGAAGGGGT CAGTGTAGAC CAGGAGACTC CAGTGGCCAC  300
AAACAACTCA ACTCAAAAAG TGCAGAAGGA GACCCAGGGC CGATTCCACC  350
TCCTCGGTGA TCCCTCAAGG AACAACTGCT CCCTGAGCAT CAGAGACGCC  400
AGGAGGAGGG ACAACGGTTC ATACTTCTTT TGGGTGGCGA GAGGAAGAAC  450
AAAATTTAGT TACAAATATT CCCCGCTCTC TGTGTATGTG ACAGCCCTGA  500
CCCACAGGCC CGACATCCTC ATCCCGGAGT TCCTAAAGTC TGGCCATCCC  550
AGCAACCTGA CCTGCTCTGT GCCCTGGGTC TGTGAGCAGG GAACACCCCC  600
CATCTTCTCC TGGATGTCAG CTGCCCCCAC CTCCCTGGGC CCCAGGACCC  650
TCCACTCCTC AGTGCTCACG ATCATCCCAC GGCCTCAGGA CCACGGCACC  700
AACCTCATCT GTCAGGTGAC GTTCCCCGGA GCTGGTGTGA CCACGGAGAG  750
AACCATCCAG CTCAGTGTCT CCTGGAAATC AGGAACCGTG GAAGAGGTGG  800
TTGTTTTGGC CGTGGGGGTA GTGGCTGTGA AGATCCTGCT TCTCTGCCTT  850
TGCCTCATCA TCCTCAGGTC CTTGTCTCTT CACTCAATGT CAGTTTCCAC  900
AAGAAGAAGG CGGTGAGGGC AGTGGAGGTT GAGGAGAATG TATATGCTGT  950
CATGGGTTAA TCTCTCAGGC CTCCAGACTG TACTTCCAGA TGTCTCCTCA 1000
TCCAGTTCCT CCACAGTCTG AATGGCCATG TTTCTTCTTC ATTGCTGGAG 1050
AATGAAGTGC AAATGCCACT GCCTGGACTG AAGGCCTTTC ACGATCTGTC 1100
TTCTGCTGGA CTCTGCTCCT GATCCCCCTT CTCCTTGCAT CACCCGAAGT 1150
CTCCCTACAC CCACCAGGCC AAGCCCTCTG TGATTCTGAG ACTTTGCATG 1200
TGTAGTTACT TCTCCTGAAA TGGCCTTCCT CCCCATTCCT GCCAATCCAG 1250
GTCCTTATCA TCCTTCAGGT TGTCTTAAAT GTCATCCAGG TGTGTGTATT 1300
TTTATGTAAT CCTTGTATGA TATTAAGCGG AGATGTGGCA TTTGTTCATT 1350
AATTTGTAGA CATATTCAGT AACCATACTG AATACATATA ATGACTATGT 1400
GCCAGCATTT CCGTATGTGC AAGAAGTTCA TCAATAGATA TAGACTCAAA 1450
GAGCTCTGTC ATCAAGCTGT TGTTCTGAAG AGCAGAAGGA TACAAATAAA 1500
AAGAAATAAG TAAAATAAAA AAAAAAAAAA AAAAAAA
```

Figure S2 (A) The amino acid sequence alignment of the resurrected human *SIGLEC17* and marmoset *SIGLEC17.* The location of Arginine residue that is responsible for sialic acid binding is highlighted in gray. Clustal W in MEGA4 was used for sequence alignment.

```
Human SIGLEC17      M--LPLLLPL PLLWAGALAQ DARFRLEMPE SVTVQEGLCI FVHCSVFYLE YGWKDSTPAY GHWFREGVSV DQETPVATNN
Marmoset SIGLEC17   .LL.......  .........P.  ..G....V..  ..R......V  ......S.IQ  ..........  .Y........  ..D.......

Human SIGLEC17     STQKVQKETQ GRFHLLGDPS RNNCSLSIRD ARRRDNGSYF FWVARGRTKF SYKYSPLSVY VTALTHRPDI LIPEFLKSGH
Marmoset SIGLEC17 ..........  .........G  ..........  .....D....  .R.E...I..  .....Q...H  ......K...  ..........

Human SIGLEC17     PSNLTCSVPW VCEQGTPPIF SWMSAAPTSL GPRTLHSSVL TIIPRPQDHG TNLICQVTFP GAGVTTERTI QLSVSWKSGT
Marmoset SIGLEC17 ..........  A.........  ..........  ..........  M.T......R  ...T......  ....A...I.  ..........

Human SIGLEC17     VEEVVVLAVG VVAVKILLLC LCLIILRSLS LHSMSVSTRR RR*
Marmoset SIGLEC17 MA........  .A........  .......P..  ...V.I....  Q..
```

Figure S2 (B) Amino acid sequence alignment of human Siglec-3(CD33) and resurrected human Siglec-17. The two IgG domains (V-set and C2-set) are indicated. The mutation of Arginine residue in Siglec-17 is indicated by a circle. Clustal W in MEGA4 was used for sequence alignment.

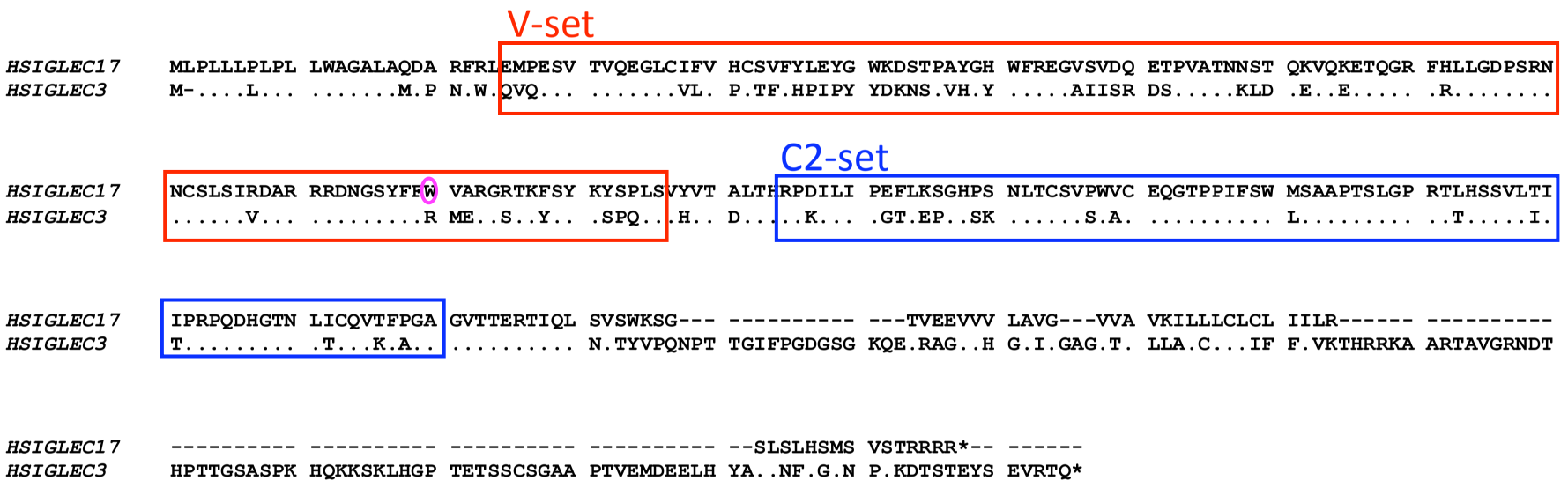
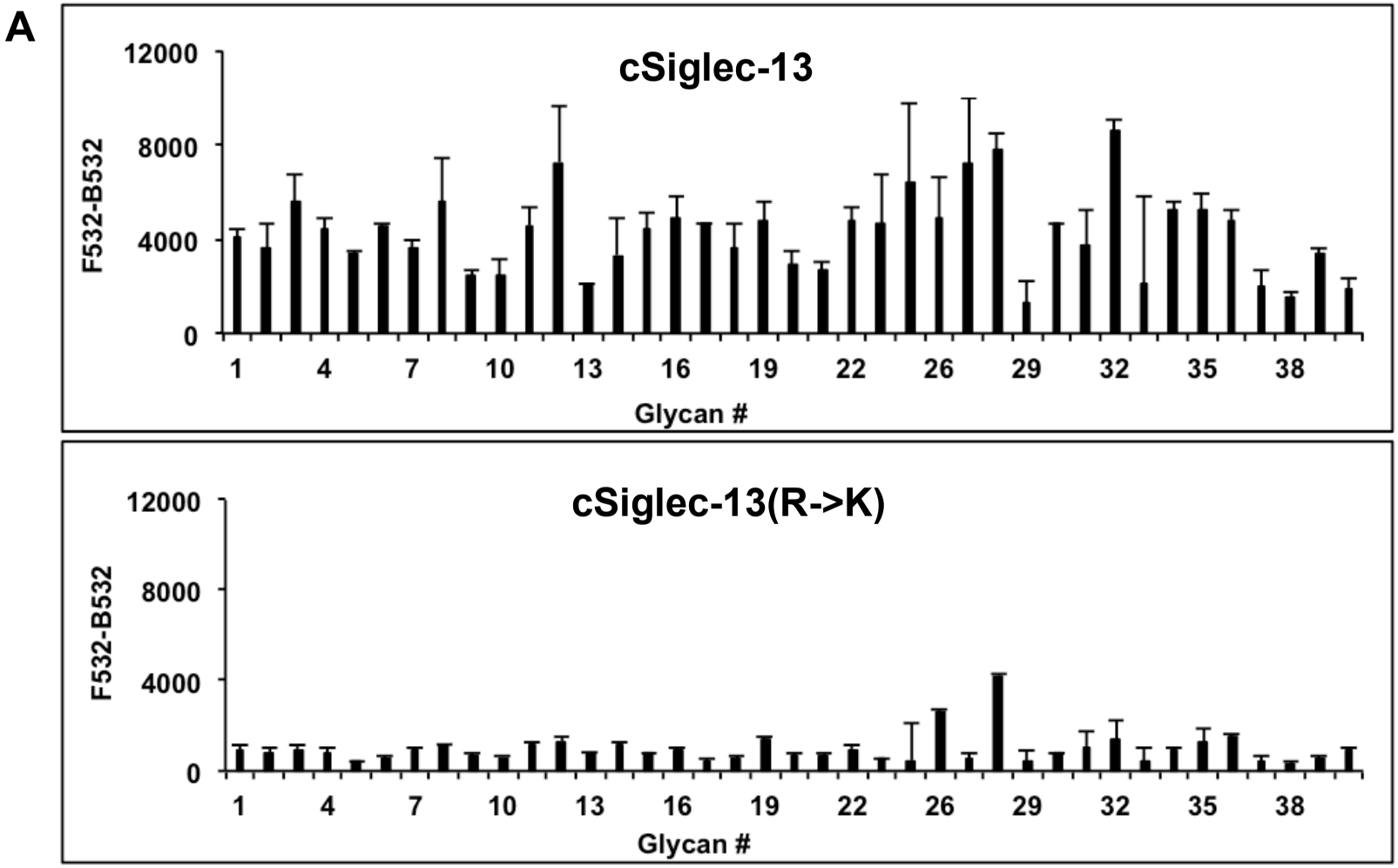


Figure S3. Binding analysis of Siglec-Fc to sialoglycan-microarray slides. Various sialoglycan-pairs (glycans numbers are as detailed in Padler-Karavani et al., 2011) with terminal Neu5Gc or Neu5Ac were spotted on Epoxy-coated slides, then developed using **(A)** chimpanzee Siglec-13 or its mutated version (Arg to Lys) and **(B)** the human Siglec-17 or its mutated version (Trp resurrected to Arg) at 7.5 µg/ml each and detected by Cy3-anti-human IgG (1.5 µg/ml). Data were analyzed with Excel, are representative of more than three independent experiments and show mean ± SD of 4 replicate spots.



B

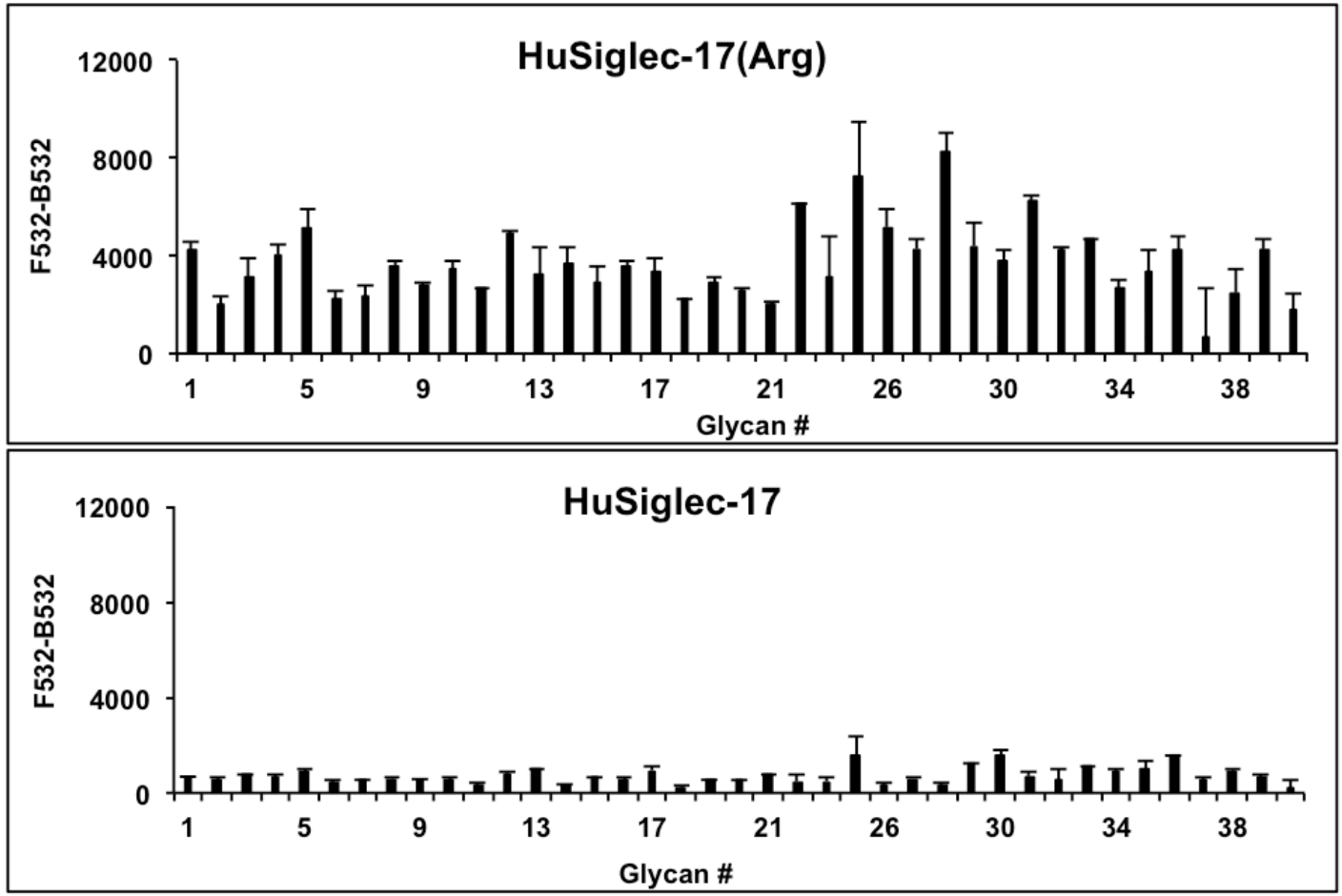

HuSiglec-17(Arg)
F532-B532
12000
8000
4000
0
1
5
9
13
17
21
26
30
34
38
Glycan #
HuSiglec-17
F532-B532
12000
8000
4000
0
1
5
9
13
17
21
26
30
34
38
Glycan #

Figure S4. Expression of Siglec-13 and Siglec-17 on transfected RAW264.7 cells. Cells transfected with pcDNA3.1 vector were used as the negative control. (**A)** Flow cytometry shows the expression of Siglec-13 in transfected RAW264.7 cells. Mouse BD Fc Block (Rat Anti-Mouse CD16/CD32; BD Pharmingen) was used to eliminate the non-specific binding. Mouse anti-Siglec-13 was used as primary antibody and Alexa Fluor 647 Goat anti-Mouse IgG (Invitrogen) as secondary antibody. (**B)** RT-PCR shows the presence of Siglec-17 transcript in transfected RAW264.7 cells. Primers used in RT-PCR reaction are S17F: TGTATGTGACAGCCCTGACCCAC and S17R: TCCTGATTTCCAGGAGACACTGAG. Total RNAs were extracted from the cultured cells with RNeasy Mini Kit (Qiagen, 74104). One-step PCR (Qiagen, 210210) was used for RT-PCR reaction. Commercial recommended protocols were followed. M: 1 KB plus DNA ladder (Invitrogen).

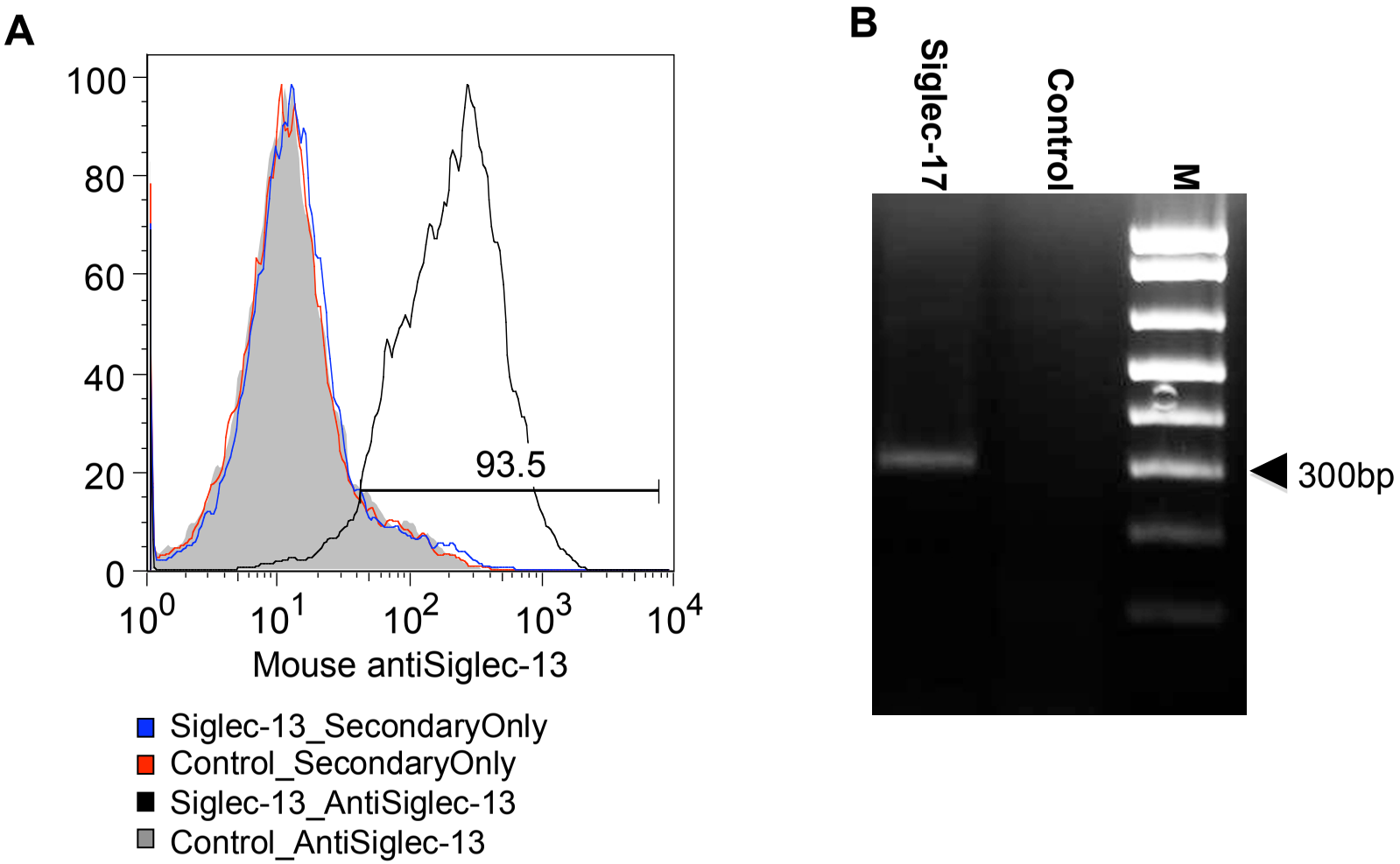
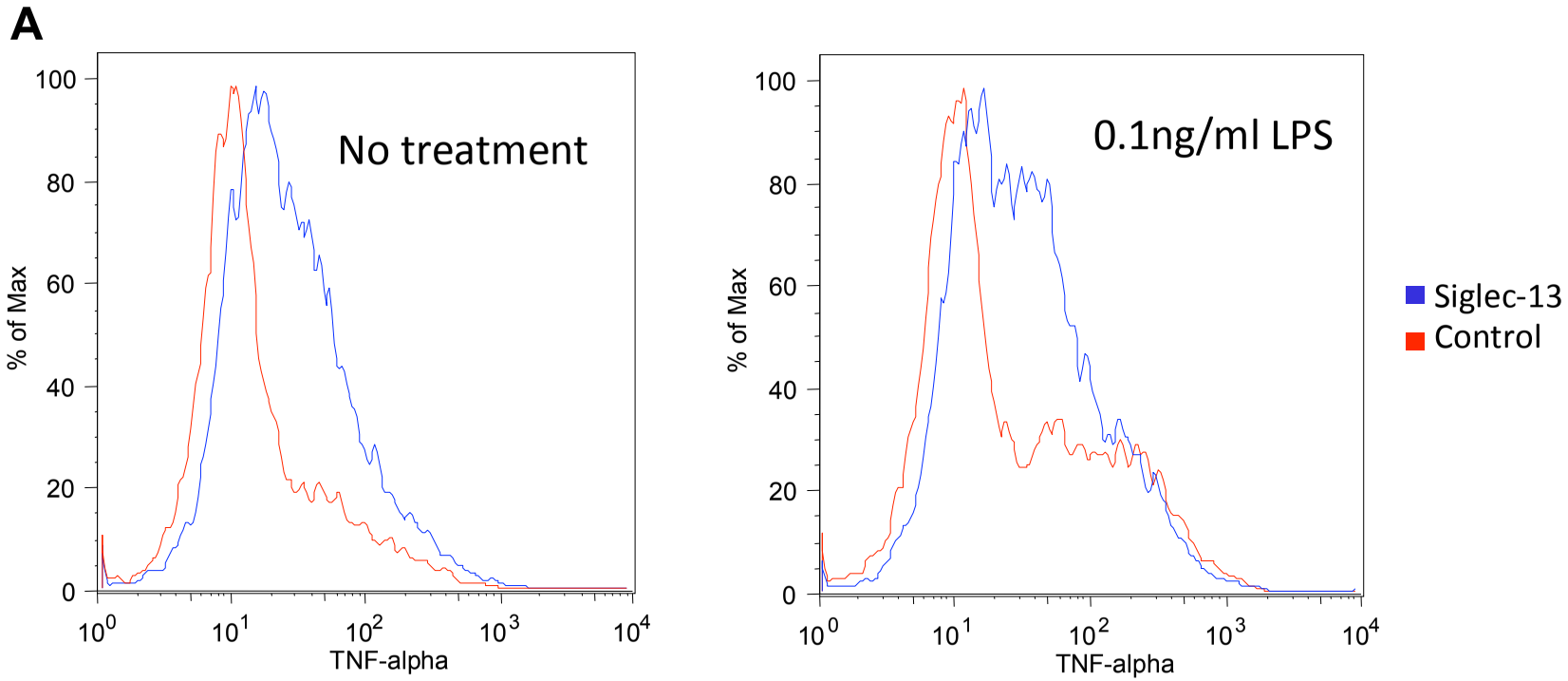
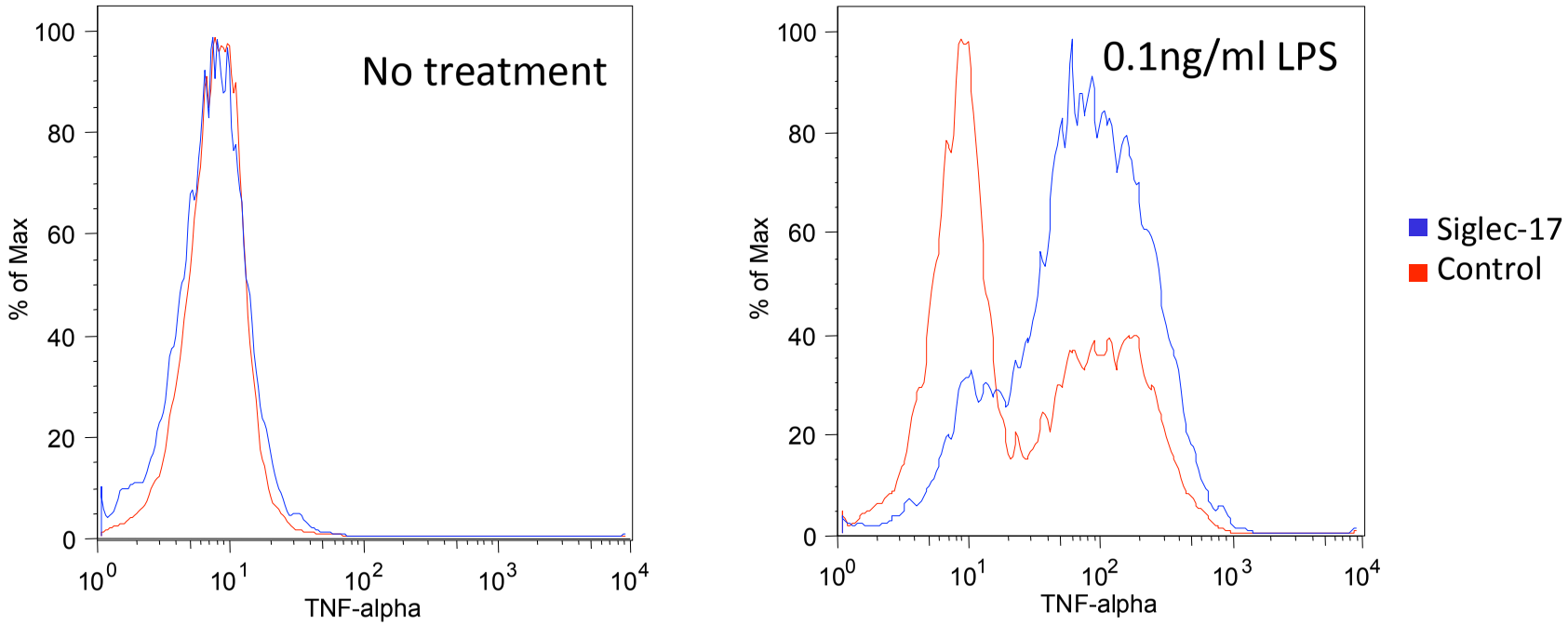


Figure S5. Increased intracellular TNF level of Siglec transfected RAW264.7 cells in response to LPS treatment. **(A)** Murine macrophage RAW264.7 cells were semi-stably transfected with cDNA encoding chimp Siglec-13 or a vector control and stimulated with LPS (0.1 ng/ml) for 2.5 hours at 37°C. Intracellular TNF was detected as described in SI. **(B)** Increased intracellular TNF in human Siglec-17 transfected cells. RAW264.7 cells were semi-stably transfected with cDNA encoding the human resurrected Siglec-17 or a vector control and stimulated with LPS (0.1 ng/ml) for 2.5 hours at 37°C. Intracellular TNF was detected as described in SI.



B



Ancient DNA Analysis
See Separate File

Figure S6. Siglec interaction with bacterial pathogens. Chimpanzee Siglec-13-Fc (Siglec-13$^{Arg->Lys}$) or resurrected human Siglec-17-Fc (Siglec-17^{Trp}) chimeras were immobilized to ELISA wells via protein-A, and binding of FITC-labeled sialylated strains GBS A909 (serotype Ia) or *E. coli* K1 RS218 (Str^{R}) was studied. Negative control was human IgG. Human Siglec-10-Fc and Siglec-9-Fc were also used as negative controls for GBS A909 and *E. coli* K1 binding, respectively. All values are means from two independent experiments ± standard deviation. Siglec-13 binding to GBS A909 is partially dependent on sialic acid interaction (See also Fig. 4A). The binding of Siglec-13 or Siglec-17 to *E. coli* K1 is sialic acid independent.

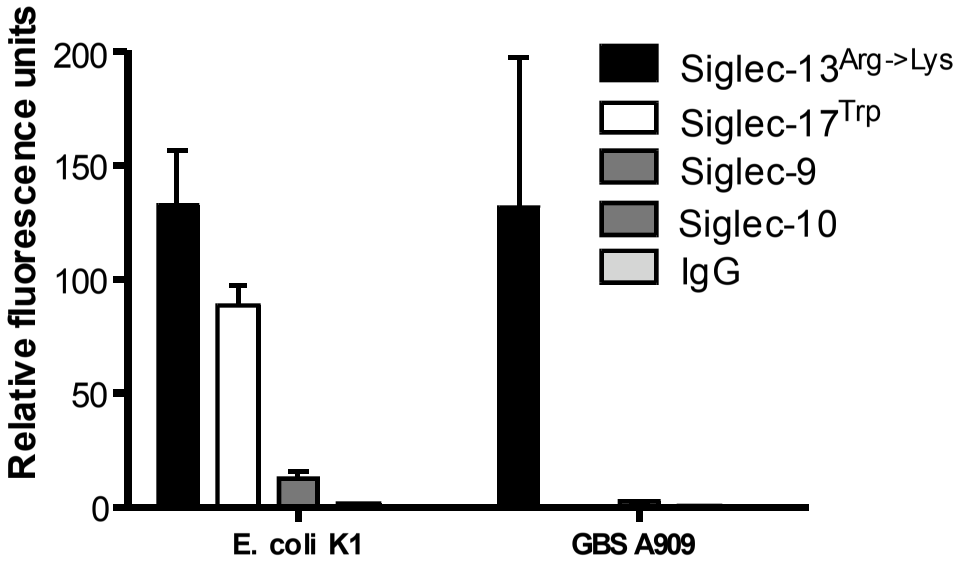


Figure S7. Dn (Nonsynonymous substitution per nonsynonymous site) and Ds (Synonymous substitution per synonymous site) estimated from pairwise comparison of *SIGLEC3*, *SIGLEC5*, *SIGLEC6*, *SIGLEC8*, *SIGLEC10*, *SIGLEC12* and *SIGLEC11* sequences (solid squares) and resurrected SIGLEC17 sequence (blank circle) among human, chimpanzee, and orangutan. Dots above the diagonal line indicates Ds>Dn.

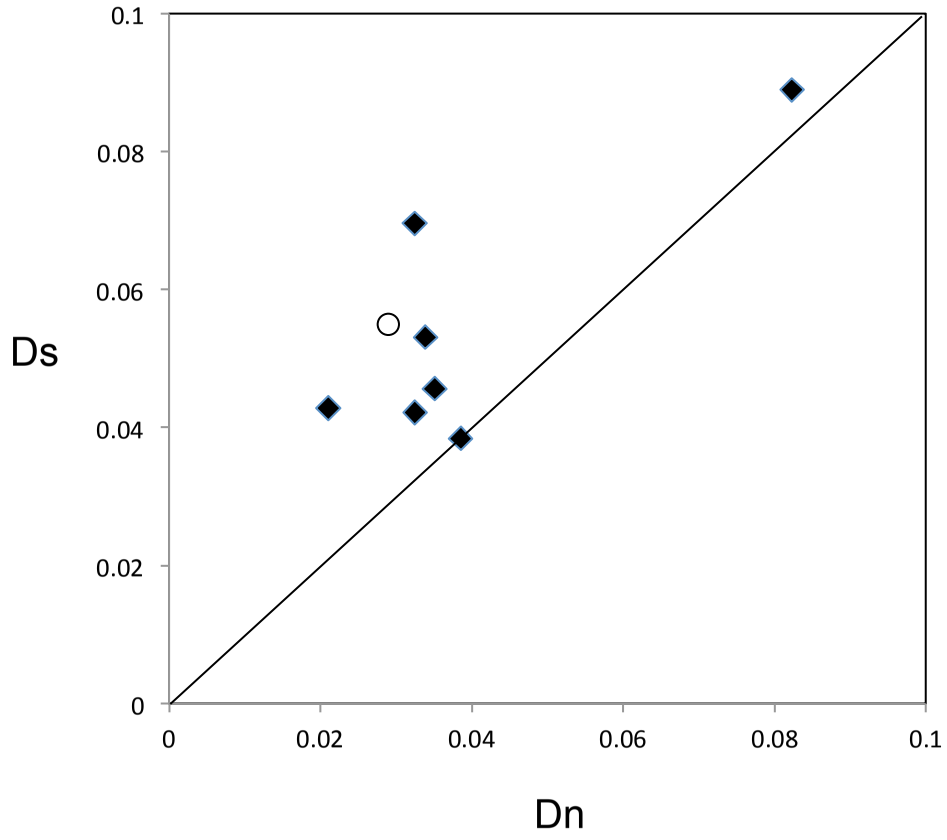


Table S1. Ethnically and Geographically Diverse African populations checked for genotyping of *SIGLEC17P* and *SIGLEC13*.

African Ethnicity	*SIGLEC17P*	*SIGLEC13*
Bakola Pygmy	17	8
Baniamer	6	8
Boni	8	18
Borana	18	18
Bulala	14	16
Datog	15	17
Fulani	17	12
Hadandawa	6	10
Hadza	19	15
Iraqw	15	16
Lemande	18	16
Luo	11	19
Mada	15	10
Sandawe	18	18
Sengwer	16	16
Yoruba	15	13
Total	228	230

Table S2. Intraspecific variation of human *SIGLEC17P*, *SIGLEC13* flanking regions and their adjacent Siglec genes. DNA sequencing of PCR product was run for 24-28 Hapmap human individuals of diverse geographic origins. Whole gene/pseudogene sequences and 3Kb upstream TSS (Transcription Start Site) sequences of *SIGLEC7-10*, and *SIGLEC17P* are NISC sequences. Six dispersed regions surrounding the *SIGLEC13* deletion within 8.5kb interval were sequenced at UCSD. The size of each region and the number of SNPs acquired are shown in table. Nucleotide diversity per site π and Watterson's theta were computed by DnaSP 5.10. Gaps in the sequence alignments were excluded from the analysis. Chromosomal interval is based on human hg18 assembly.

	Chromosomal Interval (Kb)	Total SNPs	Acquired sequence(bp)	Pi	Theta per site from S
SIGLEC9	56317872-56325571	35	5752	0.00116	0.00132
SIGLEC7	56336551-56348681	30	6579	0.00057	0.00099
SIGLEC17P	56359321-56368974	35	9046	0.00057	0.00084
SIGLEC10	56604984-56614221	52	9237	0.001	0.00127
SIGLEC13	56622379-56630898	9	5146	0.0004	0.00039
SIGELC8	56645884-56656926	32	7434	0.00064	0.00094

Table S3. PCR primers used for amplifying the regions surrounding the missing spot of human *SIGLEC13*.

Primers	5 Prime to 3 Prime
SIG13F1	ATGCCAAGTAAATGAAGCCAGA
SIG13R1	TTCTAGGTAAGGCTACATAAGC
SIG13F2	TGCTTATGTAGCCTTACCTAGA
SIG13R2	AGATTTACTCTAGAGTAGTCCC
SIG13F3	ACACCTGCTACCACAAGGCT
SIG13R3	CCTTCTGACCTCCAGAACTG
SIG13R4	CCAGGTTCAAGTGATTCTCC
SIG13F4	GCCCTAACCACCAATGTGATG
SIG13F5	TGGAGTGCAGTTGTGCGATG
SIG13R5	GTTTGTGGAGCTCTTGCCTC
SIG13F6	TCTAAGTCAGTCACAGCAGGACA
SIG13R6	TGACTCTAGGTAGTCAATCCTCC

Table S4. Polymorphism analysis for *SIGLEC17P* and *SIGLEC13* genomic regions. Sequences surrounding human *SIGLEC17P* (9Kb) and the 5Kb flanking region around the *SIGLEC13* deletion locus were acquired in 27 and 24 humans, respectively. Estimated recombination rates are 12 and 8 for *SIGLEC17P* and *SIGLEC13* regions (SI), respectively. Coalescense analysis shows significantly negative Fay and Wu's H values for both loci (* $p<0.05$). Tajima's D was not significant. DH and DHEW tests are significant for *SIGLEC17P*. See text for discussion.

Siglec	Sequence Length (Kb)	Number of SNPs	Tajima's D	Fay & Wu's H	P value for DH test	P value for DHEW test
SIGLEC17P	9	35	-1.0777	-7.3844*	0.05*	0.01*
SIGLEC13	5	9	0.04126	-2.69*	0.3	0.2

Supplement to Wang et al. PNAS submission

ANCIENT DNA ANALYSIS

Martina Lari [1], Ermanno Rizzi[2], Carlotta Balsamo[1], Giorgio Corti [2], Gianluca De Bellis[2], Laura Longo[3] and David Caramelli[1]

[1] Department of Evolutionary Biology, Laboratory of Anthropology, University of Florence, Via del Proconsolo 12, 50122 Florence, Italy.

[2] Institute for Biomedical Technologies (ITB), National Research Council (CNR), Via F.lli Cervi 93, 20090 Segrate, Milan, Italy

[3] Department of Enviromental Science, University of Siena, Via T. Pendola 62 - 53100 Siena, Italy.

MATERIALS & METHODS

1. Samples

We analyzed a new Neanderthal sample (MLS3) from Mezzena Rockshelter in the Lessini Mountains (Verona, Italia). The site is situated 250 m a.s.l. at the base of the Middle Eocene rock formation *Nummulites complanata*, located in the Avesa valley at the junction of two canyons, Vajo Gallina and Vajo Borago.The archaeological deposits are approximately 1.5 – 1.7 m deep and their stratigraphy is composed by three layers (i.e. III to I from the bottom to the top of the sequence) containing lithics, faunal and Neandertal human remains, attesting a lengthy or repeated use of the site. Several fragmented human remains were all found within sub layer Ib and include parts of the skull, of a jaw and post-cranial elements [1]. Starting from 0.5gr of a scapula fragment DNA was extracted at University of Florence in a clean-room facility exclusively dedicated to ancient DNA study and attending the most rigorous precautions to avoid contamination from modern human DNA as described in Lari et al. [2]. Immediately after extraction, DNA was amplified as described

in [2] in order to determine the mtDNA motif of a diagnostic fragments between nucletotide (nt) positions 16230 to 16262 using the primers pairs as described in [3]. Four different PCR reactions were performed. Two negative controls were introduced in every PCR reactions. Amplification products of the correct size were gel isolated and purified, that permits to check if the templates were originated from Neanderthal or modern humans. Four PCR products were cloned using the TOPO TA Cloning kit (Invitrogen), then several clones were cycle-sequenced by BigDye Terminator kit (Applied Biosystems) as described in ref. [3]; clones sequenced were determined in an Applied BioSystems 3100 DNA sequencer. With the same procedure but in a different day DNA was extracted from a rib of a Pan troglodytes specimen collected in Museum of Natural History of Florence, Anthropology and Ethnology Section, (catalog number: 5417). Control positive DNA from a human sample was obtained by oral swab and chelex extraction in a dedicated forensic lab in Florence. Extracted DNA was diluted 1:200 and subjected to the same PCR and sequencing conditions of the ancient and the museum samples as described below.

2. *Siglec gene amplifications*

2.1 Primers design.

Two different primers pairs were designed by means of Primer3 in order to detect the status of the Alu insertion in Siglec13 in the Neandertal sample. Due to the ancient DNA fragmentation, in fact, it is not possible to detect the presence of the entire insertion by PCR. For this reason we designed two different primer pairs; on the basis of Pan troglodytes sequence we designed a primer pair located between the end of the insertion and the subsequent sequence; another couple was designed on the basis of the human reference sequence that does not present the insertion (Fig 1). Primer3 default settings were used except that we searched for primers pairs amplifying short fragment (between 60 and 120 bp). With the same procedure a single primers pair was designed on the basis

of the human reference sequence in order to detect the insertion of the single G base in the Siglec P3 gene.

2.2 PCR conditions

5 µl of the extracted DNA (diluted 1:50) were used in PCR reactions with the following profile: 94°C for 10 min and 50 cycles of denaturation (94°C for 45 sec), annealing (57°C for 1 min), and extension step (72°C for 1 min); the 50µl reaction mix contained 2 units of AmpliTaq Gold polymerase, 1X reaction buffer (Applied Biosystems), 200 µM of each dNTP, 1.5 mM MgCl2, and 1 µM of each primer. Primer pairs sequences and amplicons lengths are listed in the table 1. For each primer pair three negative controls was introduced in each PCR. Amplification results were checked by agarose gel and the amplicons of expected size were excised and purified.

3. 454 sample preparation and pyrosequencing

Purified amplicons were not fragmented and were processed to obtain the single-stranded template DNA (sstDNA) library as in Roche GS FLX library preparation protocol. Quality and quantity of sstDNA were checked by Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA) and RiboGreen RNA Quantitation Kit (Invitrogen, Carlsbad, CA). SstDNA library was bound onto DNA capture bead and amplified by emulsion PCR (emPCR) as reported in Roche GS FLX protocol. Positive DNA beads were prepared as in Roche GS FLX protocol, counted (Multisizer 3 Coulter Counter; Beckman Coulter, Fullerton, CA, USA) and sequenced by FLX Genome Sequencer (FLX Roche/454 LifeSciences). In the obtained reads primers sequences were masked and the resulting portion was mapped on the corresponding reference sequence using the Amplicon Variant Analyzer application (AVA) by Roche, with default parameters. Finally, starting from the AVA multialignments, we generated the consensus sequences with a home-made Python script, which assigns for each position the most frequently base.

RESULTS

1. mtDNA amplification on Neanderthal MLS 3 sample

All the four different amplifications performed between nucleotide 16230 to 16262 gave positive result. Amplicons presented a classical mtDNA Neanderthal motif (16234 T, 16244 A, 16256 A, 16258 G) with the diagnostic transversion 16256 C/A (see Table 2) as observed in other Neanderthal Individuals [4-14].

2. Siglec genes amplifications

Each sample, as well as each extraction blanc controls, was amplified a single time with each primer pair. Amplification results are listed in the table 3. No band was seen in any PCR negative controls checked. The six positive amplicons were sequenced as previous described and the reads obtained were mapped on the relative reference sequences. Sequencing summary results are reported in the table 4. Final consensus sequences for each of the sequenced amplicon are listed below:

Neanderthal MLS Siglec P3:

cccatctgaccctcatgtctccacagggccctcgctcaggatgcaagattccggctggagat

Pan troglodytes Siglec 13_1:

ggttgcagtgagccaagatcgtgccactgcactccagcctggcgagagattgagactccgtctcaaaagaaaaaaaaaaaactaattgtttg caagaatatggacaagtgggagc (*)

Pan troglodytes Siglec P3:

cccatctgaccctcatgtctccacaggggccctcgctcaggatgcaagattccggctggagat

Homo sapiens Siglec 13_2:

gcagtgagccgagatggtgccactgcactccagcctggcgagagattgagactccgtctcaaaagaaaaaaaaatactaattgtttgcaaga

atatggacaagtgggagc (*)

Homo sapiens Siglec P3:

cccatctgaccctcatgtctccacagggccctcgctcaggatgcaagattccggctggagat

(*) Consensus sequence was manually corrected for gaps in the omopolymeric stretch reported in italics because it is know that 454 sequencing technology is not able to resolve this type of features. Detailed sequencing results for each amplicon (i.e. variants and coverage for each position) are reported table 5.

DISCUSSION

Contamination is a serious problem in ancient DNA studies, expecially when human-like samples such as Neanderthals are analyzed [15, 16]. Contamination often occurs through direct handling and washing, presumably because DNA permeates through dentinal tubules into the pulp cavity (in teeth) and the Haversian system (in bone) [17], although possibly not reaching the osteocytes [18, 19]. In that context, the excavation in itself and the handling of the samples just afterwards appear crucially at risk [17, 20]. This new Monti Lessini Neanderthal sample (MLS3) consisted of small clavicula fragment that has been not analyzed by classical morphometrical analysis and therefore were not handled. Usually, the analysis on nuclear DNA from ancient remains presents limitations due to a number of biochemical processes occurring after death. They include the small quantity of DNA available and the fragmentation of strands, caused by hydrolysis of the original template molecules. However, the MLS specimens recovered at Mezzena Rockshelter, present excellent state of preservation. This has been proved by several analysis reported in [2, 3]. To determine the degree of modern human contamination in the aliquot of MLS 3 DNA extract, we amplified cloning and sequencing the diagnostic fragments of the mtDNA HVR-I between positions 16,230nt to 16,262nt. We performed multiple PCRs, and we sequenced several clones from each PCR reaction (for a total

number of 46 amplicons see table 2). The nucleotides at these site are very unlikely to reflect contamination, because they were consistently observed in amplicons also showing mutations typical of Neanderthals and not of modern humans; moreover, these substitutions had been previously observed in 5 other Neanderthals: Feldhofer 1, Vindija 75, El Sidrón 441, Vindija 80 (33.16) and in the previous MLS individuals (MLS1)(see table 1). Low proportion of modern human mtDNA contaminants is a basic pre-requisite for nuclear loci investigation in Neanderthal or other hominid samples [21, 22, 23]. In this light the Neanderthal sample analysed here is a good candidate for nuclear DNA amplification due to the fact that we only retrieved Neanderthal-type mtDNA sequences. For this reason we are confident that the results obtained on the SiglecP3 locus are endogenous even if we cannot exclude the possibility of a sporadic contamination due to the fact that at this sequence is identical to the sequence of the present day people. It is known that searching for large insertion in paleontological samples is a difficult task due to ancient DNA fragmentation. Our amplification scheme was developed in order to amplify a short fragment at the end of a Alu insertion in siglec13 that is different between Pan troglodytes in Homo sapiens as confirmed by PCR results on control samples (see table 3 and 4). Unfortunately we were not able to obtain sequence results from this locus in our Neandertal sample. An explanation could be found in the extreme fragmentation of the sample as observed also in mtDNA loci; in fact we could easily retrieve the very short mtDNA and SiglecP3 amplicons (82 and 63bp) but we failed to obtain at the same condition the longest Siglec 13 amplicons (i. e. 110 bp fragment). Even if not informative for Siglec 13 characterization, this feature confirms that the Neanderthal sample here analyzed is not contaminated with modern human DNA and supports for the authenticity of the SiglecP3 sequence results.

References

1. Longo, L., Boaretto, E., Caramelli, D., Giunti, P., Lari, M., Milani, L., Mannino, M.A., Sala, B., Thun Hohenstein, U., Condemi, S. (2012). Did Neandertals and anatomically modern humans coexist in northern Italy during the late MIS 3? Quat. Int. 259: 102-112.

2. Lari M., Rizzi E., Milani L., Corti G., Balsamo C., Vai S., Catalano G., Pilli E., Longo L., Condemi S., Giunti P., Hänni C., De Bellis G., Orlando L., Barbujani G., Caramelli D. (2010). The Microcephalin Ancestral Allele in a Neanderthal Individual. Plos One, Volume 5, Issue 5, e10648.

3. Caramelli D., Lalueza-Fox C., Condemi S., Longo L., Milani L., Manfredini A., de Saint Pierre M., Adoni F., Martina Lari M., Giunti P., Ricci S., Casoli A., Calafell F., Mallegni F., Bertranpetit J., Stanyon R:, Bertorelle G., Barbujani G. (2006). A highly divergent mtDNA sequence in a Neandertal individual from Italy. Curr Biol., 16(16): R650-2.

4. Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M., Pääbo, S. (1997). Neandertal DNA sequences and the origin of modern humans. Cell 90: 19-30.

5. Schmitz, R.W., Serre, D., Bonani, G., Feine, S., Hillgruber, F., Krainitzki, H., Pääbo, S., Smith, F.H. (2002). The Neandertal type site revisited; interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. Proc. Natl. Acad. Sci. USA 99, 13342-13347.

6. Ovchinnikov, I.V., Götherström, A., Romanova, G.P., Kharitonov, V.M., Lidén, K., Goodwin, W. (2000). Molecular analysis of Neandertal DNA from the northern Caucasus. Nature 404, 490-493.

7. Krings, M., Capelli, C., Tschentscher, F., Geisert, H., Meyer, S., von Haeseler, A., Grossschmidt, K., Possnert, G., Paunovic, M., Pääbo, S. (2000). A view of Neandertal genetic diversity. Nat. Genet. 26, 144-146.

8. Serre, D., Langaney, A., Chech, M., Teschler-Nicola, M., Paunovic, M., Mennecier, P., Hofreiter, M., Possnert, G., and Pääbo, S. (2004). No evidence of Neandertal mtDNA contribution to early modern humans. PLoS Biol. 2, 313–317.

9. Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H., Hansen N.F., Durand E.Y., Malaspinas A.S., Jensen J.D., Marques-Bonet T., Alkan C., Prüfer K., Meyer M., Burbano H.A., Good J.M., Schultz R., Aximu-Petri A., Butthof A., Höber B., Höffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Z., Gusic I., Doronichev V.B., Golovanova L.V., Lalueza-Fox C., de la Rasilla M., Fortea J., Rosas A., Schmitz R.W., Johnson P.L., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D., Pääbo S. (2010). A draft sequence of the Neandertal genome. Science, 328(5979):710-22.

10. Beauval, C., Maureille, B., Lacrampe-Cuyaubere, F., Serre, D., Peressinotto, D., Bordes, J.G., Cochard, D., Couchoud, I., Dubrasquet, D., Laroulandie, V., et al. (2005). A late Neandertal femur from Les Rochers-de-Villeneuve, France. Proc. Natl. Acad. Sci. USA 102, 7085-7090.

11. Lalueza-Fox C., Sampietro M.L., Caramelli D., Puder Y., Lari M., Calafell F., Martínez-Maza C., Bastir M., Fortea J., de la Rasilla M., Bertranpetit J., Rosas. (2005). A: Neandertal evolutionary genetics: mitochondrial DNA data from the iberian peninsula. Mol. Biol. Evol. , 22:1077-1081.

12. Lalueza-Fox C., Krause J., Caramelli D., Catalano G., Milani L., Sampietro M.L., Calafell F., Martínez-Maza C., Bastir M., García-Tabernero A., de la Rasilla M., Fortea J., Pääbo S., Bertranpetit J., Rosas A. (2006). Mitochondrial DNA of an Iberian Neandertal suggests a population affinity with other European Neanderthals. Curr. Biol. , 16:R629-R630.

13. Burbano H.A., Hodges E., Green R.E., Briggs A.W., Krause J., Meyer M., Good J.M., Maricic T., Johnson P.L.F., Xuan Z., Rooks M., Bhattacharjee A., Brizuela L., Albert F.W., de la Rasilla M., Fortea J., Rosas A., Lachmann M., Hannon G.J., Pääbo S. (2010). Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. Vol. 328 no. 5979 pp. 723-725.

14. Krause, J., Orlando, L., Serre, D., Viola, B., Prüfer, K., Richards, M. P., Hublin, J., Hänni, C.,

Derevianko, A. P., Pääbo, S. (2007). Neanderthals in central Asia and Siberia. Nature 449, 902-904.

15. Wall J.D., Kim S.K. (2007). Inconsistencies in Neanderthal genomic DNA sequences. PLoS Genet 3: 1862-1866.

16. Coop G, Bullaughey K., Luca F., Przeworski M. (2008). The timing of selection at the human FOXP2 gene. Mol. Biol. Evol., 25: 1257–1259.

17. Gilbert M.T., Bandelt H.J., Hofreiter M., Barnes I. (2005). Assessing ancient DNA studies. Trends Ecol. Evol. 10:541-544.

18. Salamon M., Tuross N., Arensburg B., Weiner S. (2005). Relatively well preserved DNA is present in the crystal aggregates of fossil bones. Proc. Natl. Acad. Sci. U S A 102: 13783-13788.

19. Malmström H., Stora J., Dalen L., Holmlund G., Gotherstrom A. (2005). Extensive human DNA contamination in extracts from ancient dog bones and teeth. Mol. Biol. Evol. 22: 2040-2047.

20. Gilbert M.T., Willerslev E. (2006). Authenticity in ancient DNA studies. Med. Secoli. 18: 701-723.

21. Krause J., Lalueza-Fox C., Orlando L., Enard W., Green R.E., Burbano H.A., Hublin J.J., Hänni C., Fortea J., de la Rasilla M., Bertranpetit J., Rosas A., Pääbo S. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. Curr. Biol. 17: 1908-1912.

22. Lalueza-Fox C., Gigli E., de la Rasilla M., Fortea J., Rosas A., Bertranpetit J., Krause J. (2008). Genetic characterization of the ABO blood group in Neandertals. BMC Evol. Biol. 8: 342.

23. Green R.E., Krause J., Ptak S.E., Briggs A.W., Ronan M.T., Simons J.F., Du L., Egholm M., Rothberg J.M., Paunovic M., Pääbo S. (2006). Analysis of one million base pairs of Neanderthal DNA. Nature 444: 330-336.