**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 $\mu$m 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 DH5$\alpha$, 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 $\beta$-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 $3.3 \times 10^{-8}$ 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 $3.9 \times 10^{-8}$ 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 = $D_{H\text{-}C}/(6*10^6*2)$ (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.

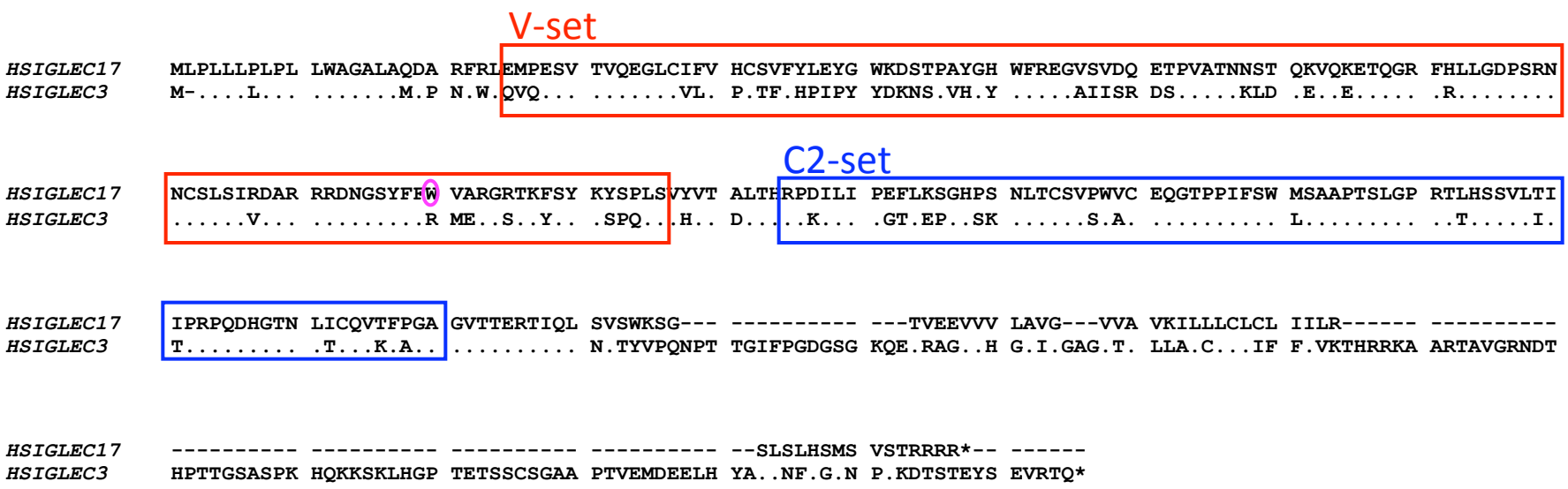
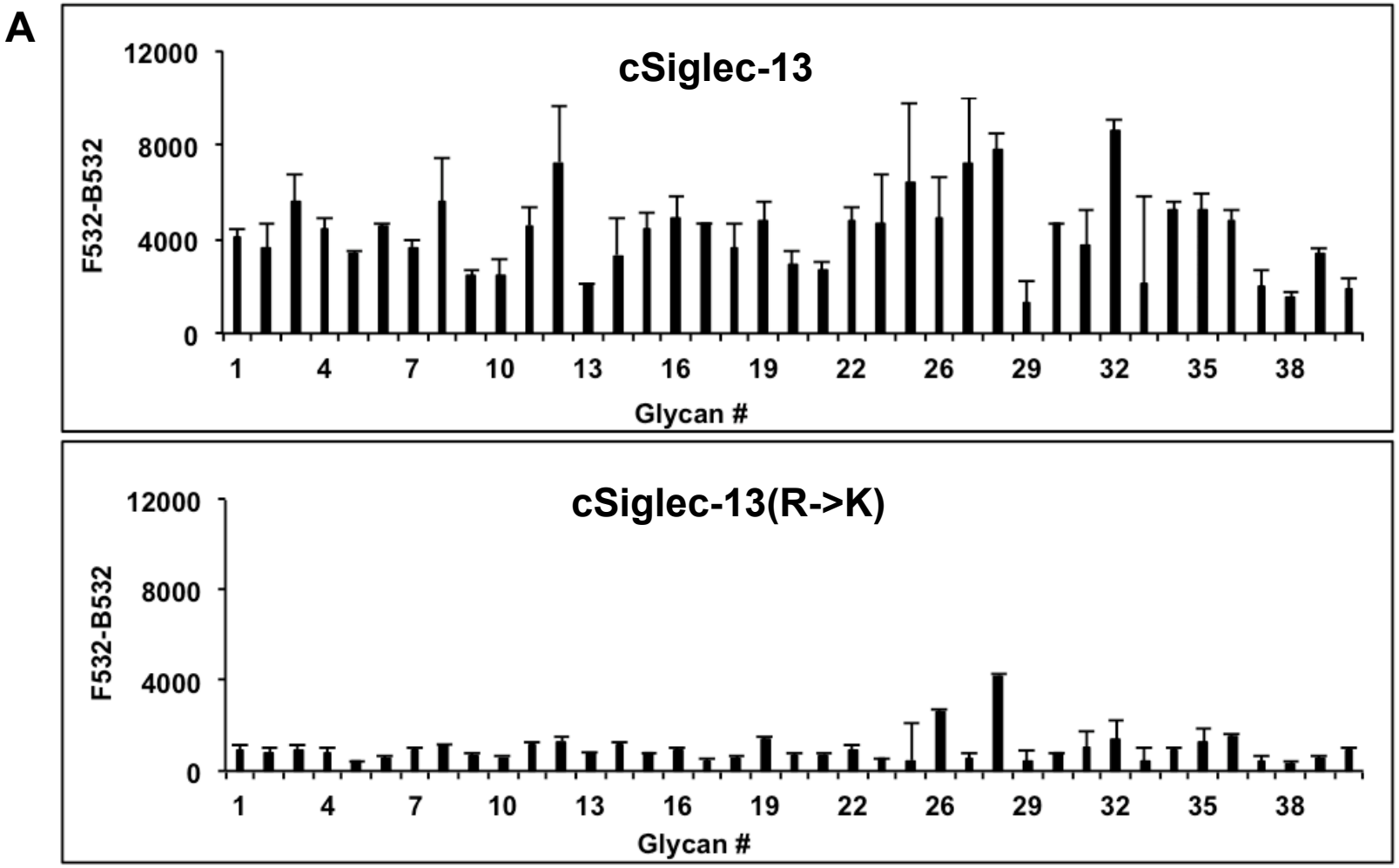


**Figure S3. Binding analysis of Siglec-Fc to sialoglycan-microarray slides.** Various sialoglycan-pairs (glycans numbers are as detailed in Padler-Karavani et al., 2011) with terminal Neu5Gc or Neu5Ac were spotted on Epoxy-coated slides, then developed using **(A)** chimpanzee Siglec-13 or its mutated 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

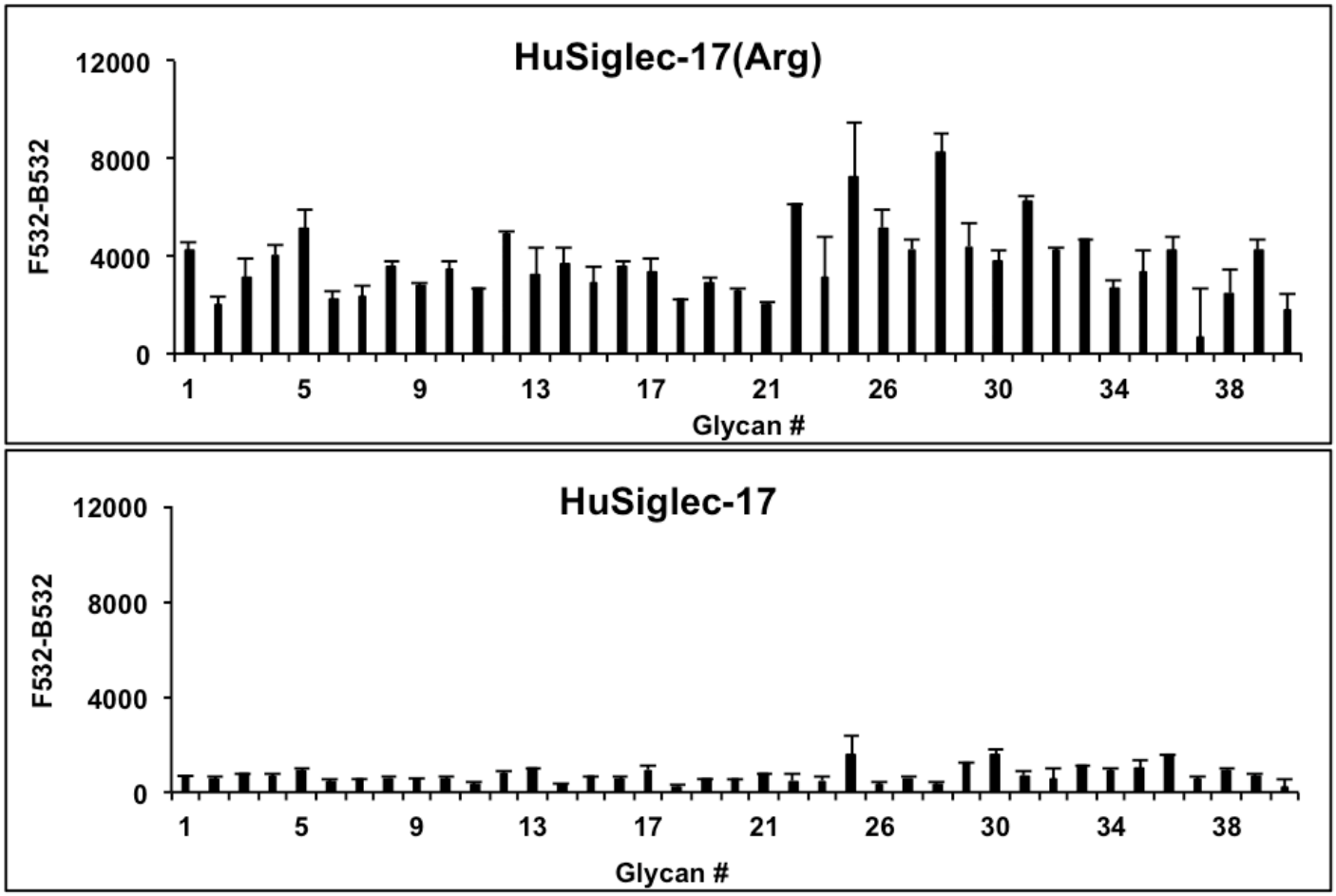

HuSiglec-17(Arg)
F532-B532
12000
8000
4000
0
1
5
9
13
17
21
26
30
34
38
Glycan #
HuSiglec-17
F532-B532
12000
8000
4000
0
1
5
9
13
17
21
26
30
34
38
Glycan #

**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).

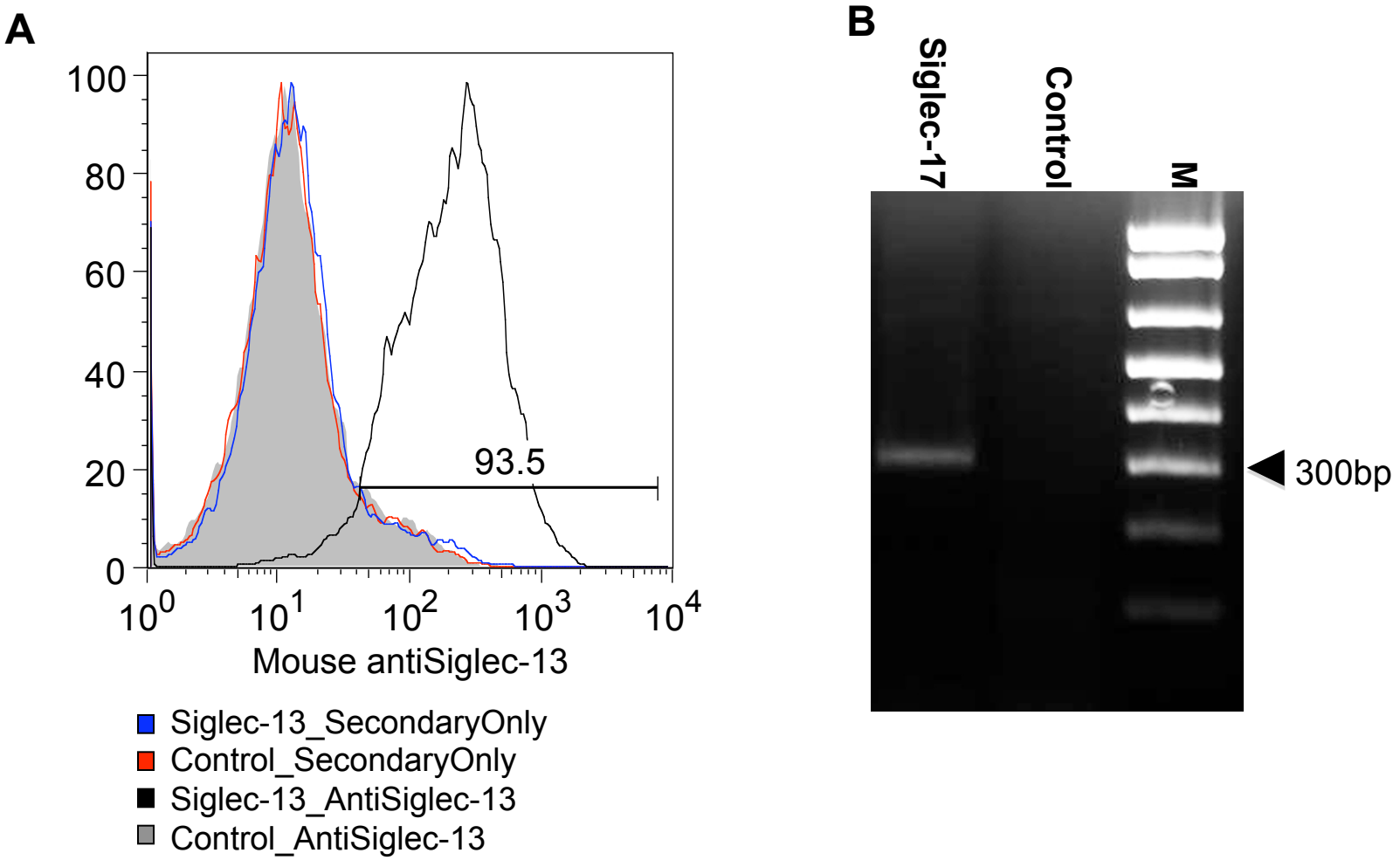
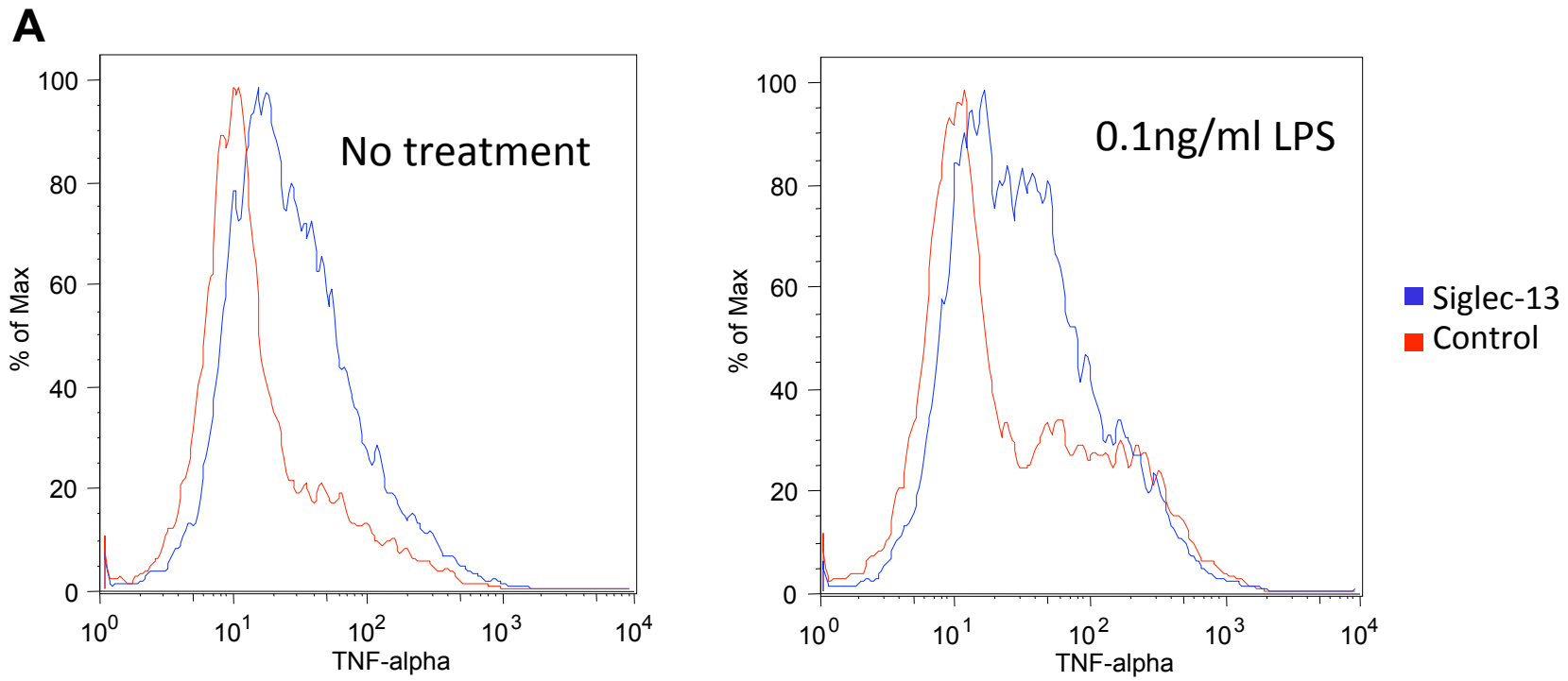
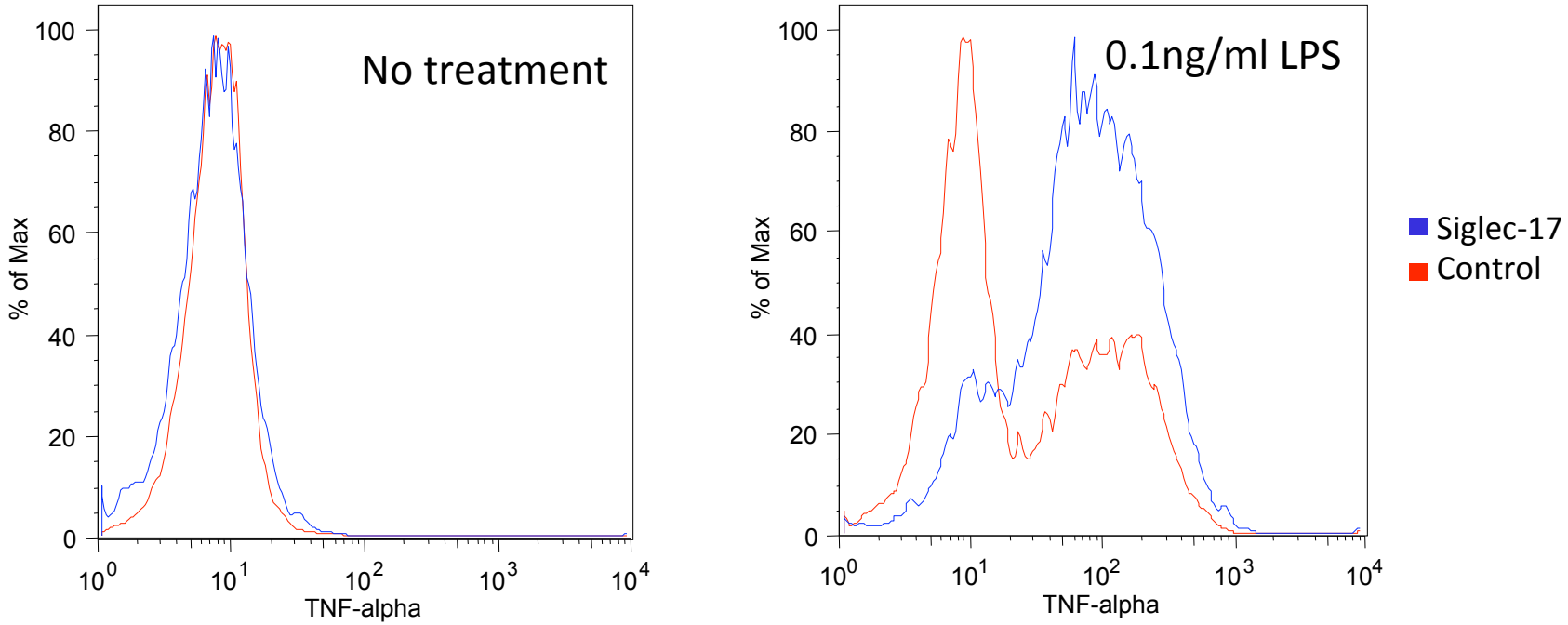


**Figure S5. Increased intracellular TNF level of Siglec transfected RAW264.7 cells in response to LPS treatment. (A)** Murine macrophage RAW264.7 cells were semi-stably transfected with cDNA encoding chimp Siglec-13 or a vector control and stimulated with LPS (0.1 ng/ml) for 2.5 hours at 37°C. Intracellular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

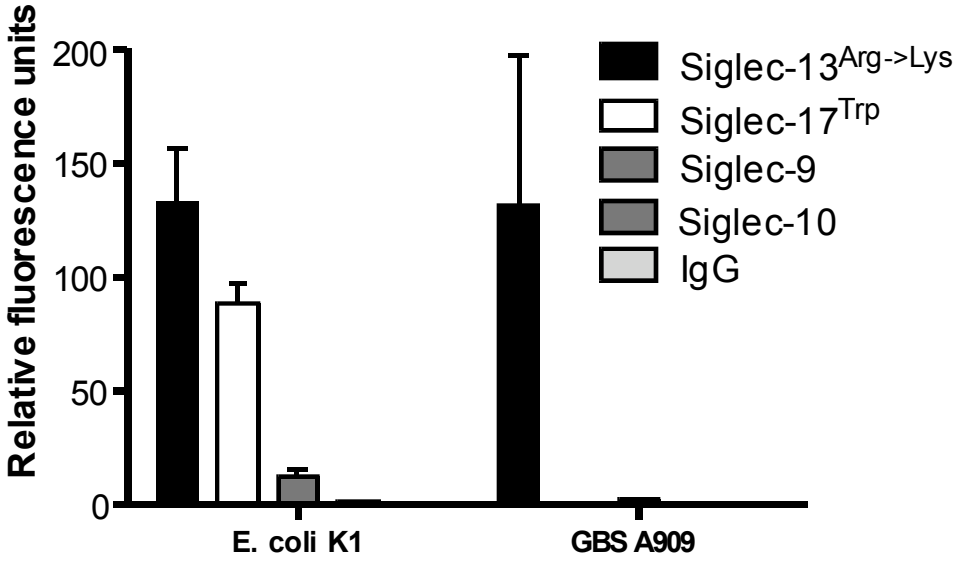


**Figure S7**. Dn (Nonsynonymous substitution per nonsynonymous site) and Ds (Synonymous substitution per synonymous site) estimated from pairwise comparison of *SIGLEC3*, *SIGLEC5*, *SIGLEC6*, *SIGLEC8*, *SIGLEC10*, *SIGLEC12* and *SIGLEC11* sequences (solid squares) and resurrected SIGLEC17 sequence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

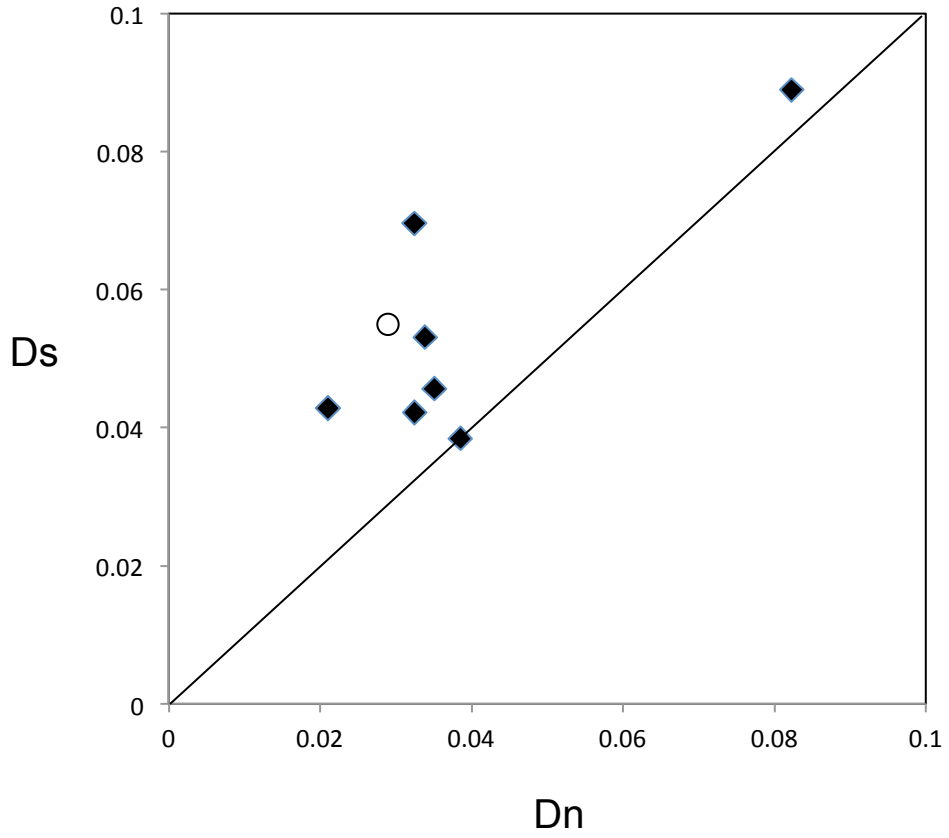


**Table S1. Ethnically and Geographically Diverse African populations checked for genotyping of *SIGLEC17P* and *SIGLEC13*.**

| 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.