Human-specific derived alleles of CD33 and other genes protect against postreproductive cognitive decline

Flavio Schwarz1,b,c,d,1, Stevan A. Springera,b,d,1, Tasha K. Altheidea,b,c,d, Nissi M. Varkia,b,c,e, Pascal Gagneuxa,b,e,2, and Ajit Varkia,b,c,d,2

*Center for Academic Research and Training in Anthropogeny, University of California, San Diego, La Jolla, CA 92093; Glycobiology Research and Training Center, University of California, San Diego, La Jolla, CA 92093; Department of Medicine, University of California, San Diego, La Jolla, CA 92093; Department of Cellular & Molecular Medicine, University of California, San Diego, La Jolla, CA 92093; and *Department of Pathology, University of California, San Diego, La Jolla, CA 92093

Edited by Richard G. Klein, Stanford University, Stanford, CA, and approved October 27, 2015 (received for review September 11, 2015)

The individuals of most vertebrate species die when they can no longer reproduce. Humans are a rare exception, having evolved a prolonged postreproductive lifespan. Elders contribute to cooperative offspring care, assist in foraging, and communicate important ecological and cultural knowledge, increasing the survival of younger individuals. Age-related deterioration of cognitive capacity in humans compromises these benefits and also burdens the group with socially costly members. We investigated the contribution of the immunoregulatory receptor CD33 to a uniquely human postreproductive disease, Alzheimer’s dementia. Surprisingly, even though selection at advanced age is expected to be weak, a CD33 allele protective against Alzheimer’s disease is derived and unique to humans and favors a functional molecular state of CD33 resembling that of the chimpanzee. Thus, derived alleles may be compensatory and restore interactions altered as a consequence of human-specific brain evolution. We found several other examples of derived alleles at other human loci that protect against age-related cognitive deterioration arising from neurodegenerative disease or cerebrovascular insufficiency. Selection by inclusive fitness may be strong enough to favor alleles protecting specifically against cognitive decline in postreproductive humans. Such selection would operate by maximizing the contributions of postreproductive individuals to the fitness of younger kin.

CD33 | Siglec | postreproductive lifespan | Alzheimer’s disease | cognitive capacity

Natural selection operates on differential survival and reproductive success. Accordingly, most vertebrates lose fecundity as they age and die soon after their reproductive periods end. Humans, orcas, and pilot whales are the only vertebrate species known to have prolonged postreproductive lifespans (1, 2). In orcas, older members of the group influence the reproductive success and survival of subsequent generations, implying that activities of postreproductive individuals increase their inclusive fitness (3). Similarly, older humans communicate cultural and ecological information and often make influential decisions within groups and wider social networks. These contributions require maintenance of cognitive capacity (4–6). Dementia (defined as a decline in memory or other thinking skills severe enough to reduce a person’s ability to perform everyday activities) negates the informational value of postreproductive individuals, clouds critical decision-making by elders, sometimes results in disruptive behavior, and eventually diverts group resources toward the care of affected individuals. Behaviors and social structures that enhance the effectiveness of postreproductive individuals might therefore be expected to select specifically for the retention of cognitive capacity. In contrast, ancestral alleles that directly enhance survival and reproductive success during fertile years are favored, even if they limit retention of cognitive capacity in postreproductive age and predispose individuals to dementia. Extended postreproductive lifespans may thus evolve as a balance between these two opposing selective forces.

The brains of humans and other related primates differ not only in size and capability but also in their susceptibility to particular diseases. For instance, late-onset Alzheimer’s disease (LOAD) is considered to be unique to humans (7). In fact, whereas humans accumulate amyloid beta deposits and neurofibrillary tangles composed of hyperphosphorylated tau protein after age 40 y, postmortem brain samples from age-matched chimpanzees and other great apes do not show the complete pathology of LOAD (8–10). Human-unique neurodegenerative diseases could be byproducts of major differences in brain development that evolved along the human lineage. If so, derived protective alleles may be compensatory and restore functions that were altered as a consequence of human-specific brain evolution.

Genome-wide studies have associated the rs3865444 SNP in the promoter region of the CD33 gene with LOAD susceptibility (11–13). CD33 is the canonical member of the family of CD33-related SIGLEC genes, which have undergone multiple unique genetic and expression changes in the human lineage (14). CD33-related Siglecs are predominantly expressed on cells of the immune system and on the surface of endothelial cells, including those of blood vessels. These Siglecs provide a functional molecular state of CD33 resembling that of the chimpanzee. Thus, derived alleles may be compensatory and restore interactions altered as a consequence of human-specific brain evolution.

We identified several genes with derived alleles that protect against age-related cognitive deterioration arising from neurodegenerative disease or cerebrovascular insufficiency. Selection by inclusive fitness may be strong enough to favor alleles protecting specifically against cognitive decline in postreproductive humans. Such selection would operate by maximizing the contributions of postreproductive individuals to the fitness of younger kin.

CD33 | Siglec | postreproductive lifespan | Alzheimer’s disease | cognitive capacity

Natural selection operates on differential survival and reproductive success. Accordingly, most vertebrates lose fecundity as they age and die soon after their reproductive periods end. Humans, orcas, and pilot whales are the only vertebrate species known to have prolonged postreproductive lifespans (1, 2). In orcas, older members of the group influence the reproductive success and survival of subsequent generations, implying that activities of postreproductive individuals increase their inclusive fitness (3). Similarly, older humans communicate cultural and ecological information and often make influential decisions within groups and wider social networks. These contributions require maintenance of cognitive capacity (4–6). Dementia (defined as a decline in memory or other thinking skills severe enough to reduce a person’s ability to perform everyday activities) negates the informational value of postreproductive individuals, clouds critical decision-making by elders, sometimes results in disruptive behavior, and eventually diverts group resources toward the care of affected individuals. Behaviors and social structures that enhance the effectiveness of postreproductive individuals might therefore be expected to select specifically for the retention of cognitive capacity. In contrast, ancestral alleles that directly enhance survival and reproductive success during fertile years are favored, even if they limit retention of cognitive capacity in postreproductive age and predispose individuals to dementia. Extended postreproductive lifespans may thus evolve as a balance between these two opposing selective forces.

The brains of humans and other related primates differ not only in size and capability but also in their susceptibility to particular diseases. For instance, late-onset Alzheimer’s disease (LOAD) is considered to be unique to humans (7). In fact, whereas humans accumulate amyloid beta deposits and neurofibrillary tangles composed of hyperphosphorylated tau protein after age 40 y, postmortem brain samples from age-matched chimpanzees and other great apes do not show the complete pathology of LOAD (8–10). Human-unique neurodegenerative diseases could be byproducts of major differences in brain development that evolved along the human lineage. If so, derived protective alleles may be compensatory and restore functions that were altered as a consequence of human-specific brain evolution.

Genome-wide studies have associated the rs3865444 SNP in the promoter region of the CD33 gene with LOAD susceptibility (11–13). CD33 is the canonical member of the family of CD33-related SIGLEC genes, which have undergone multiple unique genetic and expression changes in the human lineage (14). CD33-related Siglecs are predominantly expressed on cells of the immune system and on the surface of endothelial cells, including those of blood vessels. These Siglecs provide a functional molecular state of CD33 resembling that of the chimpanzee. Thus, derived alleles may be compensatory and restore interactions altered as a consequence of human-specific brain evolution.

We identified several genes with derived alleles that protect against age-related cognitive deterioration arising from neurodegenerative disease or cerebrovascular insufficiency. Selection by inclusive fitness may be strong enough to favor alleles protecting specifically against cognitive decline in postreproductive humans. Such selection would operate by maximizing the contributions of postreproductive individuals to the fitness of younger kin.

Significance

Most vertebrates die soon after they stop reproducing, but humans are an exception. Postreproductive humans care for offspring, assist in foraging, and communicate ecological and cultural knowledge, increasing the survival of younger individuals. Loss of cognitive capacity disrupts these benefits and burdens the group with the care of older members. We studied how the immunoregulatory receptor CD33 contributes to Alzheimer’s disease, a human-specific postreproductive condition. Surprisingly, a protective CD33 allele is derived and unique to humans, despite weak direct selection on older individuals. We identified several genes with derived alleles that protect against neurodegenerative disease and cerebrovascular insufficiency in old age. Selection by inclusive fitness may be strong enough to favor alleles that protect against cognitive decline in postreproductive humans.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

See Commentary on page 17.

1F.S. and S.A.S. contributed equally to this work.

2To whom correspondence may be addressed. Email: avark@ucsd.edu or pgagneux@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1517951112/-/DCSupplemental.

74–79 PNAS January 5, 2016 vol. 113 no. 1 www.pnas.org/cgi/doi/10.1073/pnas.1517951112
innate immune system and modulate cellular reactivity upon recognition of cell surface sialic acids that serve as "self-associated molecular patterns" (15–17). CD33 encodes a type 1 transmembrane protein with two Ig-like extracellular domains, a single transmembrane span, and two intracellular inhibitory motifs (18). Engagement of sialic acid-containing ligands by the outermost V-set Ig-like domain of CD33 results in the phosphorylation of cytosolic domains, which leads to inhibition of proinflammatory cascades in monocytes and macrophages (19, 20). Human CD33 produces two alternative splice forms: a full-length CD33M and a minor form CD33m that lacks the exon 2 encoding the V-set domain (Fig. 1) (21, 22). Thus, CD33M and CD33m differ both in molecular weight and in their ability to bind sialic acids. A series of recent studies has provided a molecular link between the rs3865444 SNP and LOAD. First, CD33 was detected in brain microglia, phagocytic immune cells that respond to cellular damage in the central nervous system (23–25). Then, the rs3865444 SNP associated with LOAD was found to be in complete linkage disequilibrium with rs12459419, a SNP located at the third base pair of CD33 exon 2 (25, 26). Notably, the rs12459419 SNP was shown to influence the efficiency of exon 2 splicing. Humans homozygous for the LOAD-risk allele marked by rs3865444C have greater cell surface expression of CD33M compared with homozygotes with the LOAD-protective allele rs3865444A (25). Thus, both rs3865444 and rs12459419 polymorphisms predict the ratio of the two CD33 isoforms. Existing studies suggest that CD33M suppresses microglial uptake of amyloid beta peptide, which otherwise accumulates in the central nervous system and contributes to LOAD (27).

Here, we compare the regulation of CD33 in humans and chimpanzees and assess the splicing and CD33 expression of alleles that predispose and protect against LOAD. We ask whether the protective allele of CD33 is derived and human-unique and survey alleles of other human genes that are also reported to protect against cognitive decline.

**Results and Discussion**

**The Protective CD33 Allele Is Derived in Humans.** Because our closest evolutionary relatives do not seem to develop the complete pathology of LOAD, we decided to investigate the allelic state of CD33 in human and nonhuman primates (Fig. 1B). Examination of a recently sequenced dataset of 79 primate genomes (28) showed that chimpanzee, bonobo, and gorilla are fixed for rs3865444C, confirming that the protective allele is derived (in this dataset, orangutans lack information for the CD33 region). Analysis of the 1000 Genomes dataset shows the rs3865444A LOAD-protective allele is found at an overall frequency of 0.21 in the human population, ranging from 0.05 in Africans to 0.48 in Native Americans (Table 1). Analysis of the available genomic data from Neanderthal and Denisovan genomes showed only the ancestral rs3865444C allele (29, 30). The pattern for rs12459419 is identical: only modern humans have the protective allele, consistent with the observation that these two SNPs are in complete linkage disequilibrium. Thus, the LOAD-protective CD33 allele represents a derived state in humans, having evolved after our common ancestor with other hominids, and possibly after our common ancestor with Neanderthal and Denisovans 550,000–765,000 y ago (30).

**CD33M Is Expressed at Higher Levels in Humans Than in Chimpanzees.** We studied expression of CD33 on human and chimpanzee peripheral blood mononuclear cells, using antibodies specific for the two splice forms. Chimpanzee monocytes show lower cell surface expression of CD33M than human monocytes, independent of the allelic state of rs3865444A (Fig. 2A and Fig. S1). Immunohistochemistry analysis of brain microsections using antibodies that recognize both CD33 isoforms detected CD33 on human and chimpanzee microglia, but at different levels (Fig. 2B). Immunoblot data confirmed a fourfold increase in CD33M expression in human brain with rs3865444C/C compared with chimpanzee brain, whereas expression of CD33m was found to be similar in both species (Fig. 2B). Taken together, these data show that two CD33 isoforms are expressed both in human and chimpanzee cells. However, whereas CD33M expression is high in humans, CD33M is maintained at lower levels in the chimpanzee. Hence, it appears that at some point after the common ancestor with chimpanzees, the human lineage up-regulated CD33M expression as the result of an unknown selective pressure or as a byproduct of other adaptive changes. The increased CD33M expression in humans may have been deleterious in terms of LOAD risk, and an alteration in splicing efficiency to lower the amounts of the CD33M isoform followed.

**CD33 and APOE Underwent Comparable Evolutionary Changes.** We noted that the evolutionary history of CD33 is similar to that of APOE, in that both genes have derived alleles that protect from a novel liability that is uniquely human (31–33). The APOE gene, which encodes for the plasma protein APOE, is polymorphic in humans. The three isoforms (e2, e3, and e4) have a distinct affinity for lipoprotein particles (34). Carriers of the minor e4 allele have higher levels of total cholesterol and accumulate atherosclerotic plaques in their arteries, leading to increased risks for cardiovascular disease, stroke, vascular dementia, and LOAD (31). The three alleles differ at two key nonsynonymous sites, resulting in amino acid differences at positions 112 and 158. Although the APOE of nonhuman mammals (including great apes) and the human e4 allele share the same residues at 112 and 158, studies of transgenic mice indicate that a threonine at position 61 in the APOE of nonhuman primates causes binding preferences that are functionally similar to the derived human e3...
Table 1. Examples of genes affecting cognitive functions in postreproductive age exhibiting disease-protective alleles uniquely in humans

<table>
<thead>
<tr>
<th>Gene</th>
<th>Associated disease</th>
<th>SNP</th>
<th>Derived allele</th>
<th>ALL</th>
<th>AFR</th>
<th>AMR</th>
<th>ASN</th>
<th>EUR</th>
<th>SAS</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD33</td>
<td>LOAD</td>
<td>rs3865444</td>
<td>A</td>
<td>0.21</td>
<td>0.05</td>
<td>0.48</td>
<td>0.19</td>
<td>0.31</td>
<td>0.16</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs12459419</td>
<td>T</td>
<td>0.21</td>
<td>0.05</td>
<td>0.48</td>
<td>0.19</td>
<td>0.31</td>
<td>0.16</td>
<td>25, 26</td>
</tr>
<tr>
<td>APOE</td>
<td>LOAD, cardiovascular disease</td>
<td>rs7412</td>
<td>T</td>
<td>0.08</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
<td>0.06</td>
<td>0.04</td>
<td>31</td>
</tr>
<tr>
<td>AGT</td>
<td>Sodium retention, sodium-sensitive</td>
<td>rs429358</td>
<td>T</td>
<td>0.85</td>
<td>0.73</td>
<td>0.90</td>
<td>0.91</td>
<td>0.84</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>hypertension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCG2</td>
<td>Hypertension</td>
<td>rs1017448</td>
<td>G</td>
<td>0.88</td>
<td>0.63</td>
<td>0.95</td>
<td>0.96</td>
<td>0.98</td>
<td>0.97</td>
<td>46</td>
</tr>
<tr>
<td>CAPN10</td>
<td>Type II diabetes</td>
<td>rs2975760</td>
<td>T</td>
<td>0.88</td>
<td>0.97</td>
<td>0.87</td>
<td>0.91</td>
<td>0.84</td>
<td>0.79</td>
<td>41, 76</td>
</tr>
<tr>
<td>TCF7L2</td>
<td>Type II diabetes</td>
<td>rs7903146</td>
<td>C</td>
<td>0.77</td>
<td>0.74</td>
<td>0.77</td>
<td>0.98</td>
<td>0.68</td>
<td>0.70</td>
<td>47, 77</td>
</tr>
<tr>
<td>EBF1†</td>
<td>Cardiovascular disease</td>
<td>rs2149954</td>
<td>C</td>
<td>0.66</td>
<td>0.61</td>
<td>0.73</td>
<td>0.77</td>
<td>0.65</td>
<td>0.59</td>
<td>78</td>
</tr>
<tr>
<td>COX-2</td>
<td>Myocardial infarction and ischemic</td>
<td>rs20417</td>
<td>C</td>
<td>0.80</td>
<td>0.65</td>
<td>0.79</td>
<td>0.96</td>
<td>0.85</td>
<td>0.81</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>stroke</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP3A5</td>
<td>Salt retention and hypertension</td>
<td>rs776746</td>
<td>C</td>
<td>0.62</td>
<td>0.18</td>
<td>0.80</td>
<td>0.71</td>
<td>0.94</td>
<td>0.67</td>
<td>80</td>
</tr>
<tr>
<td>PPARG</td>
<td>Type II diabetes</td>
<td>rs1801282</td>
<td>G</td>
<td>0.07</td>
<td>0.01</td>
<td>0.12</td>
<td>0.03</td>
<td>0.12</td>
<td>0.12</td>
<td>39</td>
</tr>
<tr>
<td>PON1</td>
<td>Dementia</td>
<td>rs2618516</td>
<td>T</td>
<td>0.34</td>
<td>0.26</td>
<td>0.24</td>
<td>0.30</td>
<td>0.38</td>
<td>0.52</td>
<td>81</td>
</tr>
</tbody>
</table>

*Allele frequencies are from the 1000 Genomes database.
†302 kb downstream.
‡Most common nonfunctional.

The Origins of Some Derived Protective Alleles May Predate the Origin of Modern Humans. It has been proposed that derived protective alleles arose and underwent selection primarily after human populations became agrarian, with the changes in selection caused by a sedentary lifestyle and increased consumption of carbohydrate- and lipid-rich diets (48, 49). This scenario clearly holds for lactase persistence and may also be true of some of the derived alleles in Table 1. However, surveys of the literature, and our current analyses using several different population-level metrics (FST, Rbp, iHS, Tajima’s D, and Fay and Wu’s H; Dataset S1), indicate that these genes do not consistently show signs of recent selection. In the 1000 Genomes data, some do show evidence of population structure and of possible deviations from neutrality; others appear to be evolving neutrally. Moreover, all the protective derived alleles in Table 1 are present in African populations, on the continent where modern humans originated and where contemporary nonagricultural societies persist. If inclusive selection operated on these other derived protective alleles, it may have been most effective near the origin of behaviorally modern humans ~100,000 y ago, a depth of time at which clear genomic signatures of selection and selective sweeps are degraded beyond recognition with existing statistical methods (48–50).
positive effects early in life via differential survival or reproductive success can have deleterious (pleiotropic) effects later in life, when natural selection is weak or absent. Derived alleles that protect against cognitive decline during a prolonged postreproductive period could result from kin selection late in life through increased survival of younger kin. Such alleles would operate by maintaining the cognitive capacity of grandmothers, as well as of active elderly decision makers and transmitters of valuable cultural information. Both of these features are apparently unique to humans compared with other primates. Indeed, the onset of dementia in otherwise physically fit humans would have a detrimental effect on these functions not only because of the dysfunctional and sometimes disruptive behavior of affected elders but also because of the time and resources eventually needed to care for older individuals whose decision making is compromised.

Taken together, these data indicate that the existence of a postreproductive lifespan in humans may be supported by selection to maintain cognitive function in older individuals. It would be interesting to search for similar derived alleles in the two mammalian species of toothed whales also known to have prolonged postreproductive lifespans (3, 66). Interestingly, in at least one instance (Orcinus orca; Orca, or killer whales), it has been suggested that ecological knowledge and leadership may have led to the evolution of menopause (3). Postmenopausal female killer whales lead foraging groups, and their leadership is most significant for the survival of the group when food is scarce (67). Helping by older female orcas results in differential survival of their male offspring, indicating kinship dynamics and inclusive fitness effects.

The evolutionary window in which cognition-based selection could operate in humans likely opened with the development of linguistic capabilities and the capacity for cumulative culture and shared intentionality. This indicates a time depth of >100 kya, which limits the potential to detect selective sweeps with existing statistical methods. However, polymorphic protective alleles do exist, and there could be ongoing selection for alleles that protect against cognitive decline in contemporary human populations. Blue zones are populations in which extreme human longevity occurs in the absence of modern medicine (68, 69). It is interesting that in these populations, there tends to be no “reirement” age: older individuals continue to be active, functional members of society.

Conclusions and Perspectives

Survival beyond a fertile age is not a recent phenomenon in human groups, as hunter–gatherer populations exhibit aging structures in which about a third of the females are postreproductive (51). In contrast, chimpanzee females rarely live past fertile years, even when accorded full veterinary care in captivity (5, 52, 53). It has been suggested that the cooperative child-rearing in humans is crucially tied to the survival of grandmothers (5, 54) and also greatly contributed to the shortening of interbirth interval in humans (3–4 y compared with more than 4 y in chimpanzees) (55, 56). Unlike chimpanzees, young humans are unable to feed themselves until long after they reach weaning age and critically depend on the help and support of group members other than their mothers. Cooperative breeding played a key role in human cognitive evolution (57), and selection for grandmothers could have propelled the evolution of postmenopausal longevity, setting the social context for the subsequent evolution of many other features in our lineage (5, 54, 58–60). For humans, however, intergenerational information transfer is also an important factor for the survival of individuals, groups, and larger social networks characteristic of foraging societies (61). Such information transfer could enhance inclusive fitness via resource sharing and reduced mortality of younger individuals (62–64).

Antagonistic pleiotropy is considered a key mechanism in the evolution of senescence (65), as genes strongly selected for their

**Fig. 2.** Full-length CD33M is expressed at higher levels in humans compared with chimpanzees. (A) CD33M is expressed at higher levels on the cell surface of human peripheral blood monocytes. CD33M/*CD33* values reflect the ratio of WM53/HIM3.3 positive cells. The rs3865444 allelic state is indicated. Error bars reflect mean ± SEM. *P* values were calculated with an unpaired Student's *t* test. (B) CD33M is expressed at higher levels in human brain, whereas CD33m is expressed at comparable levels in human and chimpanzee. Human samples were from homozogous C/C subjects for the rs3865444 SNP. Tubulin is used for normalization of signals. Error bars reflect mean ± SEM; *n* = 3. *P* values were calculated with an unpaired Student's *t* test. Immunohistochemistry shows expression of CD33 both in chimpanzee and human brain. Mouse monoclonal (mAb) and rabbit polyclonal (rAb) CD33 antibodies detect both CD33 forms. CD68 is used as a marker for microglia.

**Materials and Methods**

CD33 rs3865444 Genotyping. Genomic DNA was extracted from blood and tissues with a DNeasy Blood & Tissue Kit (Qiagen). The allelic state of rs3865444 was defined by NcoI digestion of a 421-bp fragment amplified from genomic DNA, using primers 5′-GCGAACCCCATGTCTAAA-3′ and 5′-CCTACC-TCCCTCTGTGCC-5′. CD33 Expression on Blood Monocytes. Chimpanzee blood samples were collected as extra tubes only during routine noninvasive health screens of chimpanzee subjects at the Yerkes National Primate Center, Emory University, Atlanta (supported by NIH Base Grant ORIP/OD P51OD011132; routine collection covered under local institutional review board (IRB) approval by Emory University). All collections were made prior to the September 15, 2015 designation of captive chimpanzees as endangered species. Chimpanzee blood samples were shipped overnight on ice to the University of California, San Diego. Human blood was collected at about the same time into identical tubes from healthy volunteer donors (following informed consent, under the approval from the University of California, San Diego Human Subjects IRB), and stored overnight on ice, to ensure similar treatment conditions prior to analysis. All health and safety issues related to handling of human and nonhuman primate samples are covered by an institutional biosafety approval from the University of California, San Diego Environmental Health and Safety Committee. All individuals who handle the samples receive the required training regarding precautions for blood-borne pathogens. Peripheral blood mononuclear cells were isolated by centrifugation over Ficoll-Paque Premium (GE Healthcare). Contaminating red blood cells were lysed by incubation in

---

Schwarz et al.
ACK buffer (Gibco, Life Technologies). Cells were stained with an APC-conjugated anti-CD33 antibodies clone WM52 (BD Biosciences) and a FITC-conjugated anti-CD33 antibodies clone HIM3.4 (Biolegend), or isotype control antibodies. CD33 expression was measured with a BD FACSCalibur (BD Biosciences). Data were analyzed with FlowJo.

Immunohistochemistry. Frozen or paraffin-embedded samples of human brain were obtained from the National Cancer Institute-funded Co-operative Human Tissue Network. Frozen or paraffin embedded samples from chimpanzee brains were obtained from the Yerkes Primate Research Foundation.

Frozen brain samples were homogenized in lysis buffer [1% (vol/vol) Mayer’s HEPES, 0.1% (vol/vol) SDS, 5% (vol/vol) 2-mercaptoethanol, and 0.1% (vol/vol) proteinase K] overnight at 4 °C in a humid chamber. Sections were incubated with biotinylated mouse anti-mouse IgG antibodies, HRP-conjugated streptavidin, biotinyl tyramide enhancement, and HRP-conjugated streptavidin, followed by AEC substrate (Vector Laboratories), followed with Mayer’s nuclear counterstain, using established protocols. Deparaffinized tissue sections were treated with antigen retrieval in Decloaking Chamber (Biocare) and overlaid with mouse anti-CD33 monoclonal antibodies (clone PWS44; Leica Biosystems Inc.) overnight at 4 °C in a humid chamber. Sections were incubated with biotinylated anti-mouse IgG antibodies, HRP-conjugated streptavidin, biotinyl tyramide enhancement, and HRP-conjugated streptavidin, followed by AEC substrate. Nuclei were counterstained with Mayer’s hematoxylin.

Immunoblot. Frozen brain samples were homogenized in lysis buffer [1% (vol/vol) Nonidet P-40, 20 mM Tris at pH 8, 150 mM NaCl]. Lysates were spun twice at 20,000 × g. Protein concentration of the supernatant was measured using a BCA kit (Pierce). Proteins were separated by SDS/PAGE and transferred to nitrocellulose membranes. Membranes were incubated with rabbit anti-CD33 (St Cruz Biotechnology, Inc.) or mouse anti-tubulin antibodies (Sigma), as well as secondary antibodies (LI-COR). Signals were acquired with an Odyssey instrument (LI-COR) and analyzed by Image Studio software (LI-COR).

Statistical Analysis. Unpaired Student’s t test was used for comparisons involving two groups. Prism 6 Program (GraphPad) was used for most of the statistical analyses.

Genomic Data. Human genomes were accessed from the 1000 Genomes Project server (www.1000genomes.org), using tabix to extract regions in variant call format (VCF 4.1) (71). Neandertal and Denisovan VCFs were downloaded from their respective servers (Neandertal, cdna.eva.mpg.de/neandertal/taita/Alta/Neandertal/VCF); Denisovan, cdna.eva.mpg.de/denisova/VCF/hg19_1000g) (29, 30). Primate VCFs were downloaded from the Great Ape Genome Project server (biologiaevolutiva.org/greatape1) (28). Bed coordinates defining the genomic regions surrounding CD33 and other genes were retrieved using the University of California, Santa Cruz, genome browser (72). Primate VCFs are referenced to hg18, bed coordinates were lifted to hg19 using CrossMap (crossmap.sourceforge.net). New coordinates were checked against hg19 and corrected by hand when necessary. Variants in each region of each species were extracted with vcftools (73), and variants of each CD33 SNP were called directly from each file.

Polymorphism Tests. Bed coordinates of the loci in Table 1 were retrieved from build hg19, using the University of California, Santa Cruz, Genome Browser. A region containing each gene, plus 100 kb upstream and 100 kb downstream, was retrieved in VCF format from the 1000 Genomes Project database, using tabix. Patterns of polymorphism at each gene and its surrounding region were analyzed for each population and each region (as defined by the 1000 Genomes Project), using the selectionTools pipeline and custom Perl scripts (74). Several statistical tests were evaluated: frequency-based methods (Tajima’s D, Fay and Wu’s H), linkage disequilibrium-based measures (LD, r2, iHS); and population differentiation-based methods (Fs). Each is suited to detecting selection at different timescales. Frequency-based and population differentiation-based methods are better suited to detecting events farther in the past than linkage disequilibrium-based methods (75). Outputs from selectionTools were formatted with custom Perl scripts and visualized in R as heatmaps. Plots were examined for evidence of deviation from the null expectation. The values of each test statistic from the 200 kb surrounding the gene of interest were used to generate histograms that describe the empirical null expectation. Test statistic values from within each gene itself were compared with this expectation to look for statistical outliers whose positions lie within in the gene of interest.

ACKNOWLEDGMENTS. We thank Andrea Garcia-Binaghi for help with immunohistochemistry and the members of the A.V. and P.G. groups for suggestions. This work was supported by grants from the National Institutes of Health (PO1HL107150 to A.V. and R01GM095882 to P.G.) and by the Harold and Leila Mathers Foundation. This work was supported by grants from the National Institutes of Health (PO1HL107150 to A.V. and R01GM095882 to P.G.) and by the Harold and Leila Mathers Foundation.


