Investigating Association of Human-Specific Derived Alleles of *CD33* and Other Genes with Lifespan of Iranians

Mehdi Kargar (MSc)
Young Researchers and Elite Club, Shiraz Branch, Islamic Azad University, Shiraz, Iran

Hoshang Jamali (PhD)
Department of Microbiology, Jahrom Branch, Islamic Azad University, Jahrom, Iran

Mehdi Sadeghi (MSc)
Department of Biochemistry, Shiraz Branch, Islamic Azad University, Shiraz, Iran

Tel: +98-9173915663

Email: kargarmehdi53@yahoo.com

Corresponding author: Mehdi Kargar
Address: Young Researchers and Elite Club, Shiraz Branch, Islamic Azad University, Shiraz, Iran

ABSTRACT

Background and Objectives: Aging is a multi-agent phenomenon due to prolonged inflammation and stress. CD33 or Siglec3 is a membrane receptor that acts against aging by inhibiting inflammatory reactions. The aim of this study was to evaluate a possible relationship between *CD33* copy number and lifespan of an Iranian population.

Methods: The study included 50 individuals with cancer or Alzheimer's disease as the case group and 50 members of a family over 70 years old as the control group. Blood samples were collected and transferred to the laboratory. CD33 copy number was calculated using the QX100 Droplet Digital PCR system. A number of CD33 single-nucleotide polymorphisms including rs3865444, rs273634 and rs3852865 were genotyped using specific primers and the PCR method.

Results: The mean number of *CD33* copies among the case group (7.78) was significantly lower ($P < 0.05$) than control group (12.72). In the case group, the mean number of *CD33* copies was 7.83 among men and 7.73 among women. In the control group, the mean number of CD33 copies was 12.73 among men and 12.71 among women.

Conclusion: CD33rSiglecs counteract random molecular damage, which is the main driver of aging. Therefore, the *CD33rSiglec gene* number may be correlated with longevity. Our results indicate that there may be a link between reduced *CD33rSiglec* copy number and development of diseases.

Keywords: Gene Copy Number, Siglec-3, CD33 Antigens, Cancer.

INTRODUCTION

Aging is controlled partly by genetic factors, such as insulin/insulin-like growth factor 1, mammalian target of rapamycin, AMP-activated protein kinase and sirtuin signaling pathways (1). Understanding the details of this regulatory process may help develop measures for alleviating and controlling conditions associated with aging (2). Sialic acid-binding immunoglobulin-type lectins (Siglecs) have been reported to be involved in the early aging process (3). They are a superconducting family of type 1 lectins with salicilc acid linkage affinity and specific domains. Siglecs are categorized into two subgroups: CD33-related Siglecs and others including Siglec-1, Siglec-2 and Siglec-4 (4, 5). CD33 or Siglec-3 (also known as gp67, p67) is a membrane receptor found on myeloid-derived cells (6). There are two classes of Siglecs based on sequence homology and conservation. The first group consists of Sialoadhesin/Siglec-1, CD22/Siglec-2, MAG/Siglec-4 and Siglec-15 that share low sequence identity but are conserved across mammals. In contrast, the genes encoding CD33rSiglecs underwent extensive rearrangements, including duplication, conversion and pseudogenization. Therefore, they vary in number and sequence among mammals (2). For instance, mice and humans (the most studied organisms in this respect) express five and ten functional CD33rSiglecs, respectively (7). In humans, CD33rSiglecs are numbered (e.g., Siglecs-3, -5, -6, -7, -8, -9, -10, -11, -XII, -14 and -16), while murine CD33rSiglecs (other than Siglec-3) are identified by a distinct alphabetical nomenclature (8). Although there is no comprehensive information regarding Siglec expression patterns, it is known that many members are expressed in a cell type-specific manner. For instance, among the murine CD33rSiglecs, CD33 is expressed mainly in granulocytes, Siglec-E is expressed primarily in neutrophils, monocytes, microglia and dendritic cells, Siglec-F is mainly found in eosinophils and mast cells, Siglec-G is predominantly expressed in B cells and some dendritic cells, and Siglec-H is primarily expressed in plasma cytoid dendritic cells (9). Although it is not possible to identify clear CD33rSiglec orthologs between human and murine Siglecs, due to rapid Siglec evolution and deep divergence time between mice and humans, some Siglec receptors (for instance, Siglec-E and Siglec-9) are considered to be functional homologs (10). Notably, there is no evidence so far for a significant degree of functional redundancy among Siglecs. Despite the general low affinity of Siglecs towards sialylated structures, it appears that each Siglec has a unique sialoglycan specificity profile with regard to the type of sialic acid, its linkage and the composition of underlying glycan structure. Interestingly, CD33rSiglecs can transmit inhibitory signals into immune cells by phosphorylation of intracellular immunoreceptor tyrosine-based inhibition motif or immunoreceptor tyrosine-based inhibition motif-like domains, thus quenching pro-inflammatory cascades (11).

CD33 is present on myeloid cells as well as on the surface of some lymphatic cells. The receptor binds to sialic acid and therefore can be considered a member of the Siglec family of lectins. The extracellular part of this receptor consists of two immunoglobulin domains (one LgV domain, and another LgC2 domain), which places CD33 in the immunoglobulin classification. Intracellular part of CD33 contains major inhibitors of the tyrosine immune receptor, which are involved in inhibiting cellular activity (12, 13). In a study conducted on 12 mammalian species, organisms with more *CD33* copy numbers showed a longer lifespan. The function of these genes is to reduce inflammation. In addition, the risk of death and early aging was found to be higher in mice with fewer *CD33* copy numbers (14). These data indicate that CD33rSiglecs regulate inflammatory damage and that the expansion of their number in the genome has coevolved with the extension of lifespan in mammals (2, 12-14). Schwarz et al. believe that the CD33rSIGLECs are co-evolved in mammals to help control the levels of reactive oxygen species during inflammation, thereby reducing cell damage and extending the lifespan of animals. Given that the number of working *CD33rSIGLEC* genes may vary among individuals (2), we investigated the possible association between *CD33rSIGLEC* gene numbers and life expectancy in people aged 70 years and older.

MATERIALS AND METHODS

This research was conducted at the Young Researchers Club of Islamic Azad University, Shiraz Branch, Iran. Fifty patients (24 men, 36 women) with cancer or Alzheimer's disease who were referred to Imam Reza Clinic and Namazi Hospital were selected as the case group. Fifty people (24 men, 36 women) aged over 70 years who were living in the Fars Province, were selected as the control group. Demographic data including age, gender, number of children, medical history and blood type were recorded using a questionnaire. Two ml blood samples were collected in EDTA tubes (FL Medical, Italy) and stored at -20 °C until analysis. The study protocol was approved by the ethics committee of the university (code: 95/2/19762), and written informed consent was taken from all subjects. DNA was extracted using the CBSA Blood Mini kit (CBSA, Iran), according to the manufacturer's guidelines. The quality and quantity of DNA were assessed by gel electrophoresis and spectrophotometry, respectively. The *CD33* gene is located on chromosome 19: 51729554-51729676 (RefSeq: NC_000019.9 NT_011109.16). Proprietary digital droplet PCR (ddPCR) CNV kit (BioRad. USA) and specific primers were used to determine *CD33* copy number. Each kit included a 20X master mix containing 18 μmol of primer and 5 μmol of fluorescent probe. The extracted DNA samples were diluted at a ratio of 1:100 to 1:50 and were activated to produce droplets to accurately count *CD33* copy numbers using the ddPCR method (15). Cycling conditions are presented in table 1. Primer sequences used for genotyping of *CD33* rs3865444, rs273634 and rs3852865 SNPs are presented in table 2. PCR conditions were as follows: initial denaturation at 95 °C for 5 minutes, followed by 35 cycles of denaturation at 94 °C for 90 seconds, annealing at 58 °C (for rs3865444), 55 °C (for rs273634) and 56 °C (for rs3852865) for 45 seconds, extension at 72 °C for 60 seconds, and final extension at 72 °C for 5 minutes. The PCR reaction mixture (25 μl) consisted 2 μl DNA sample (50 ng/ml), 1.5 μl MgCl$_2$ (50 mM), 2.5 μm dNTP (10 mM), primers (0.4 mM) and 2.5 units Taq DNA polymerase (Color Taq PCR Master Mix (2X, CBSA, Iran) (10). PCR products were subjected to electrophoresis on 1.5% agarose gel stained with 0.5 μl ethidium bromide (SBSA, Iran) and the results were visualized under ultraviolet light.

Data were analyzed with SPSS software (version 19). The Chi-square test or Fisher's exact test was used to assess the differences between the groups. P-values less than 0.05 were considered as statistically significant.

RESULTS

Mean age of the patients and the controls was 66.44 ± 8.96 years and 77.94 ± 5.75 years, respectively. The Mann-Whitney U test showed a significant difference in the mean *CD33* copy numbers between patients and controls (P=0.001). Mean number of *CD33* copies in the patients (7.78) was less than in the control group (12.72). There was no significant difference in the mean number of *CD33* copies between men and women in both groups (Figure 1).

Table 1- Cycling conditions of the proprietary digital droplet PCR for assessment of *CD33* copy number

Process	Temperature (°C)	Time	Number of cycles
Enzyme activation	95	10 min	1
Denaturation	94	30 s	40
Annealing	55	1 min	40
Elongation	55	1 min	40
Enzyme inactivation	98	10 min	1

Table 2- Sequence of the primers used for detection of *CD33* SNPs.

| Row	*CD33* SNPs	Primer sequence
1	*CD33* rs3865444(G allele) F inner	CAGCCTCACCTAGATCCGTG
2	*CD33* rs3865444(T allele) R inner	CTATATCCTGCTGGACTAAACACTCA
3	*CD33* rs3865444(Forward outer primer)	AGCAGGAGACCTCAGCCC
4	*CD33* rs3865444(Reverse outer primer)	GAGGCTAGAAAGGAAATTGGAACT
5	*CD33* rs273634(G allele) F inner	AAGTCATCCAGCACGATATTGTAG
6	*CD33* rs273634 (A allele) R inner	GAGTGTTGTGGACGTCTCAACT
7	*CD33* rs273634 (Forward outer primer)	AGAAAAGGGCACCAAATGAG
8	*CD33* rs273634 (Reverse outer primer)	CAATCTGGCCACACTTATATGC
9	*CD33* rs3582865(T allele) F inner	GGGCCATCTGAGACAGGT
10	*CD33* rs3582865 (C allele) R inner	ACTTTAGACATGGGGTGGG
11	*CD33* rs3582865 (Forward outer primer)	GTGGAGGGGAAGGACTACAG
12	*CD33* rs3582865 (Reverse outer primer)	CCCTACAGAAACCTATATCCTGC

Frequency distribution of CD33 SNPs is presented in figure 2. Results of the Chi-square test showed no significant difference in the rs3865444 genotype distribution between the patients and controls (P=0.484). The highest frequency for rs3865444 SNP was observed in the case of G/G genotype in both the patients (42%) and the control (40%) groups. In the case of rs3852865 SNP, the most frequent genotypes were T/T genotype (38%) in the case group and T/C genotype (44%) in the control group. In addition, the difference in genotype distribution was not statistically significant (P=0.220). In the case of rs273634 SNP, the most frequent genotypes were G/G genotype (40%) in the case group and G/A genotype (52%) in the control group. There was a statistically significant difference between the two groups in terms of genotype distribution (P=0.038).

Table 3- Alleles and size of gene products.

RefSeq	snp *CD33* rs3865444	Homozygote / heterozygote	Size
NC_000019.10:g.51224706C>A	*CD33* rs3865444 G allele	G/G	228 bp
NC_000019.9:g.51727962C>A	*CD33* rs3865444 T allele	T/T	342 bp
	CD33 rs3865444 G/T	G/T	524 bp
RefSeq	snp *CD33* rs273634	Homozygote / heterozygote	Size
NC__000019.10:g.51215581C>T	*CD33* rs273634 G allele	G/G	169 bp
NC_000019.9:g.51718837C>T	*CD33* rs273634 A allele	A/A	256 bp
	CD33 rs273634 G/A	G/A	379 bp
RefSeq	snp *CD33* rs3582865	Homozygote / heterozygote	Size
NC_000019.10:g.51210809G>A	*CD33* rs3582865 T allele	T/T	169 bp
NC_000019.9:g.51714065G>A	*CD33* rs3582865 C allele	C/C	256 bp
	CD33 rs3582865 T/C	T/C	379 bp

Figure 1- Frequency of *CD33* gene copies among patients with Alzheimer's disease and controls
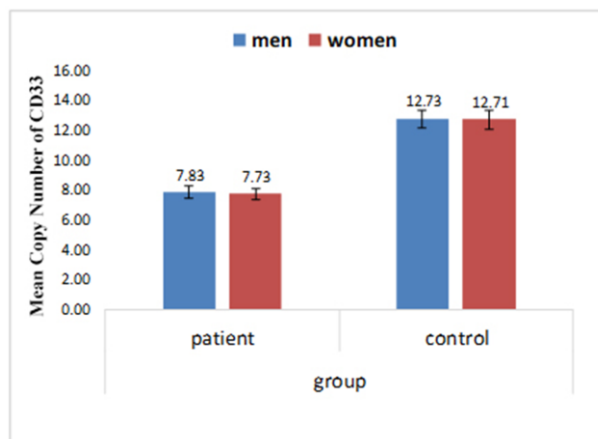
In the present study, *CD33* copy numbers did not differ significantly between men and women. In patients, the *CD33* copy number was highest in individuals with prostate cancer and lowest in those with kidney cancer. In the control group, the mean *CD33* copy number was highest in individuals with skin cancer and lowest those with gastric cancer (Figure 3).

Figure 2- Genotype distribution of *CD33* SNPs among the cases and controls.
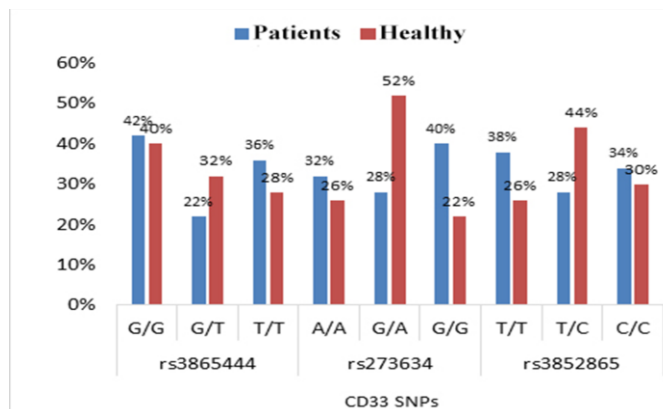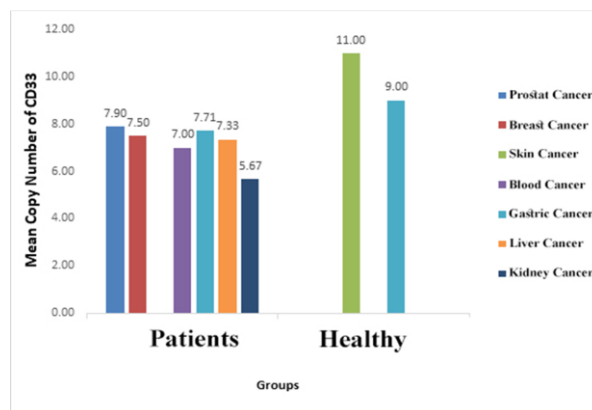
Figure 3- Mean number of *CD33* copies in patients with different types of cancers and control individuals

DISCUSSION

Previous studies have reported that the number of CD33rSIGLEC genes may be correlated with maximum lifespan (2). Hence, in this study, we investigated the possible association between CD33rSIGLEC gene numbers and life expectancy in people aged 70 years and older. We identified G/G as the most common genotype for rs3865444 SNP in patients with Alzheimer's disease and control individuals. Two previous studies found that rs3865444c SNP is preserved in the genome of all primates and is not changing through generations. Moreover, rs3865444A was found to be late-onset Alzheimer's disease-protective with overall frequency of 0.21 among human population; ranging from 0.05 in the African population to 0.48 in the Native American population (15, 16). The same pattern was observed for rs12459419 allele. In other words, only alleles in modern humans are protected. It has been shown that these two SNPs are in linkage disequilibrium. In fact, *CD33* allele is protected in humans but has evolved compared to its ancestors. In a study by Schwarz et al., it was observed that the allelic sequence of Alzheimer's disease is completely preserved in primates but changed slightly in the human genome. Moreover, variation of phenotypic symptoms of Alzheimer's disease may be due to *CD33* allele modification during evolution and C>A allele replacement on chromosome 19 of late-onset Alzheimer's disease-protective rs3865444A SNP (2). Assessment of the sequences around rs3865444 and rs12459419 SNPs showed that both rs3865444C and rs12459419C SNPs are ancestral alleles. In 2011, Naj et al. reported that rs3865444 SNP has undergone a change in adenine in people with Alzheimer's disease.

The overall frequency of this altered allele in the human genome was 0.21%, 0.05% in black people, 0.48% in Native Americans, 0.19% in South Asians and 0.31% in Europeans (17).

The evolutionary theory of germ line and disposable soma theory predict that long-lived species assure their longevity through investments in more resilient somatic tissues (18). Moreover, genes involved in management of cellular stress and repair mechanisms contribute to lifespan. Indeed, it has been experimentally shown that lifespan of eight mammalian species correlates with the ability of their primary fibroblasts to cope with stress (19). Siglecs are capable of modulating cellular inflammatory responses and the number of genes encoding CD33rSiglecs varies widely between species (7).

In 2016, Walker et al. claimed that the allelic alteration in *CD33* rs3865444 SNP is a risk factor for Alzheimer's disease (3). They also indicated that the disease symptoms are associated with the type of allelic alteration. In 2013, Mortland et al. examined *CD33* SNPs in children with acute myeloid leukemia (AML) who were treated with gemtuzumab ozogamicin (20). They reported that changes in the *CD33* gene could significantly affect the clinical outcomes of the disease, including disease phenotype and side effects. According to this study, rs12459419 and rs35112940 SNPs were observed in children treated with gemtuzumab ozogamicin-AAMLO3p10 but were absent in AML patients who were undergoing chemotherapy. In addition, it was shown that individuals with rs12459419 TT genotype are at a higher risk of developing AML compared to those with rs12459419 CC or CT genotypes.

In the present study, *CD33* copy numbers did not differ significantly between men and women. In patients, the *CD33* copy number was highest in individuals with prostate cancer and lowest in those with kidney cancer. In the control group, the mean *CD33* copy number was highest in individuals with skin cancer and lowest those with gastric cancer. Raj et al. examined the relationship between Alzheimer's disease symptoms and *CD33* exon 2 copy numbers. They reported an association between *CD33* rs3865444C SNP and Alzheimer's disease. A higher exon 2 copy numbers were associated with *CD33* overexpression at cellular level. Whereas, a unique type of CD33 (monotype) exists at the level of ancestral cells in American, African and European populations which extends exon 2 copy number and increases rs3865444C allelic frequency, ultimately causing a significant increase in occurrence of phenotypic variation as well as symptoms of Alzheimer's disease (21).

CONCLUSION

CD33 SIGLEC could maximize lifespan due to a general expansion of genes encoding for cell surface receptors that interact with pathogens to initiate immune responses.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

All contributing authors declare no conflicts of interest.

REFERENCES

1. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. *The hallmarks of aging*. Cell. 2013; 153(6): 1194-217. doi: 10.1016/j.cell.2013.05.039.

2. Schwarz F, Fong JJ, Varki A. *Human-specific evolutionary changes in the biology of siglecs*. Adv Exp Med Biol. 2015; 842: 1-16. doi: 10.1007/978-3-319-11280-0_1.

3. Walker DG, Dalsing-Hernandez JE, Campbell NA, Lue LF. *Decreased expression of CD200 and CD200 receptor in Alzheimer's disease: a potential mechanism leading to chronic inflammation*. Exp Neurol. 2009; 215(1): 5-19. doi: 10.1016/j.expneurol.2008.09.003.

4. Vlad SC, Miller DR, Kowall NW, Felson DT. *Protective effects of NSAIDs on the development of Alzheimer disease*. Neurology. 2008; 70(19): 1672-7. doi: 10.1212/01.wnl.0000311269.57716.63.

5. Szekely CA, Zandi PP. *Non-steroidal anti-inflammatory drugs and Alzheimer's disease: the epidemiological evidence*. CNS Neurol Disord Drug Targets. 2010; 9(2): 132-9.

6. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. *The Global Distribution of Clinical Episodes of Plasmodium Falciparum Malaria*. Natu. 2005; 434(3): 214-217.

7. Angata T, Margulies EH, Green ED, Varki A. *Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms*. Proceedings of the National Academy of Sciences of USA. 2004; 101(5): 13251-13256.

8. Macauley MS, Crocker PR, Paulson JC. *Siglec-mediated regulation of immune cell function in disease*. Nat Rev Immunol. 2014; 14(10): 653-66. doi: 10.1038/nri3737.

9. Pillai S, Netravali IA, Cariappa A, Mattoo H. *Siglecs and immune regulation*. Annual Review of Immunology. 2012; 30(6): 357-392.

10. La¨ubli H, Pearce OM, Schwarz F, Siddiqui SS, Deng L, Stanczak MA, et al. *Engagement of myelomonocytic Siglecs by tumor-associated ligands modulates the innate immune response to cancer*. Proc Natl Acad Sci U S A. 2014; 111(39): 14211-6. doi: 10.1073/pnas.1409580111.

11. Crocker PR, Paulson JC, Varki A. *Siglecs and their roles in the immune system*. Nat Rev Immunol. 2007; 7(4): 255-66.

12. Zotova E, Bharambe V, Cheaveau M, Morgan W, Holmes C, Harris S, et al. *Inflammatory components in human Alzheimer's disease and after active amyloid-beta42 immunization*. Brain. 2013; 136(3): 2677-2696.

13. Yuan Q, Chu C, Jia J. *Association studies of 19 candidate SNPs with sporadic Alzheimer's disease in the North Chinese Han population*. Neurol Sci. 2012; 33(1): 1021-1028. doi: 10.1007/s10072-011-0881-0.

14. Winblad B, Graf A, Riviere ME, Andreasen N, Ryan JM. *Active immunotherapy options for Alzheimer's disease*. Alzheimers Res Ther. 2014; 6(1): 74-86.

15. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. *A high-coverage genome sequence from an archaic Denisovan individual*. Science. 2012; 338(6104): 222-226.

16. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. *The complete genome sequence of a Neanderthal from the Altai Mountains*. Nature. 2014; 505(7481): 43-49.

17. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, et al. *Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1are associated with late-onset Alzheimer's disease*. Nat Genet. 2011; 43(5): 436-441.

18. Kirkwood TB. *Comparative life spans of species: why do species have the life spans they do?* Am J Clin Nutr. 1992; 55(6 Suppl):1191S-1195S. doi: 10.1093/ajcn/55.6.1191Sa.

19. Kapahi P, Boulton ME, Kirkwood TB. *Positive correlation between mammalian life span and cellular resistance to stress*. Free Radic Biol Med. 1999; 26(5-6): 495-500.

20. Mortland L, Alonzo TA, Walter RB, Gerbing RB, Mitra AK, Pollard JA, et al. *Clinical significance of cd33 non-synonymous single nucleotide polymorphisms (snps) in pediatric patients with acute myeloid leukemia treated with gemtuzumab-ozogamicin-containing chemotherapy*. Clin Cancer Res. 2013; 19(6): 1620-1627.

21. Raj T, Ryan KJ, Replogle JM, Chibnik LB, Rosenkrantz L, Tang A, et al. *CD33: increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility*. Hum Mol Genet. 2014; 23(10): 2729-36. doi: 10.1093/hmg/ddt666.