

Published in final edited form as:

*Anticancer Res.* 2010 April ; 30(4): 1263–1272.

## Unique SNP in CD44 Intron 1 and its Role in Breast Cancer Development

Juhua Zhou<sup>1</sup>, Prakash S. Nagarkatti<sup>1</sup>, Yin Zhong<sup>1</sup>, Kim Creek<sup>1</sup>, Jiajia Zhang<sup>2</sup>, and Mitzi Nagarkatti<sup>1</sup>

<sup>1</sup>Department of Pathology, Microbiology and Immunology, School of Medicine, University of South Carolina, Columbia, SC 29209, U.S.A.

<sup>2</sup>Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208, U.S.A.

### Abstract

In the current study, we investigated if CD44 polymorphisms are associated with increased susceptibility to breast cancer. Direct nucleotide sequencing analysis identified a novel and unique single nucleotide polymorphism (SNP, designated as CD44 Ex2+14 A>G) in the CD44 intron 1 region in 84% of breast cancer patients, which was significantly higher than that seen in normal donors. Moreover, the breast cancer patients with homozygous unique SNP in CD44 intron 1 had breast cancer at earlier ages, larger tumor burden, more regional lymph node metastases at the time of diagnosis, and higher cancer recurrence rate. There was a strong association between the unique SNP in CD44 intron 1 and CD44 expression on peripheral blood mononuclear cells. Our results suggest that CD44 polymorphism is associated with breast cancer development, and CD44 polymorphism analysis may be effectively used in the risk assessment, prediction, prevention, diagnosis and genetic epidemiological analysis of breast cancer.

### Keywords

Gene polymorphism; breast cancer; CD44; patient age; tumor burden

---

Breast cancer is a leading cause of cancer-related deaths in women in the United States and worldwide. One of the most important risk factors for breast cancer is family history of the disease, indicating that genetic factors may be important determinants of breast cancer risk. A number of breast cancer susceptibility genes have been identified. The two most important breast cancer susceptibility genes, *BRCA1* and *BRCA2*, accounted for approximately 15% of breast cancer cases due to inherited mutations or alterations (1, 2). However, *BRCA1* and *BRCA2* mutations are present in far less than 1% of the general population (3). In addition to *BRCA1* and *BRCA2*, five other genes, *TP53*, *PTEN*, *LKB1*, *ATM* and *CHEK2*, are also well established as breast cancer susceptibility genes (4). The proportion of breast cancers in the general population explained by these germline gene

mutations is very small. It is estimated that all the currently known breast cancer susceptibility genes account for less than 25% of the familial aggregation of breast cancer (5). The evidence to date strongly suggests that majority of the familial clustering of breast cancer is unexplained and other breast cancer susceptibility genes still remain to be identified.

Inherited genetic variations, or gene polymorphisms, are viewed as major factors in breast cancer prevalence. However, the data thus far have been inconclusive. After reviewing a substantial number of studies that have investigated candidate genes for low-penetrance breast cancer susceptibility alleles, Dunning *et al.* (1999) concluded that there is no clear evidence that any of these polymorphisms are strongly associated with breast cancer risk (6). Recent findings on the associations between polymorphisms and breast cancer susceptibility have also been controversial. For instance, a previous study suggested that *BRCA1* carriers with longer lengths of the CAG repeat in the androgen receptor gene are at increased risk of breast cancer (3). However, a subsequent study failed to find any significant association of the androgen receptor repeat polymorphisms with risk of breast cancer in either *BRCA1/2* carriers or in non-carriers (7). Several reports documented that progesterone receptor gene polymorphism is associated with an increased breast cancer risk (8, 9). In contrast, other studies failed to show any significant associations between breast cancer risk and progesterone receptor polymorphisms (10–12). It is more likely, however, that the lack of correlation between polymorphisms and susceptibility to breast cancer may be due to genes that have yet to be identified or tested (6). Lately, analysis of 9 selected single nucleotide polymorphisms (SNPs) in a large sample size (>11,000 breast cancer cases and >14,000 controls) indicated that *CASP8* D302H was associated with decreased risk of breast cancer (13). However, analysis of a large number of SNPs (710 SNPs in 120 candidate genes) suggested that no single gene polymorphism was associated with breast cancer risk and the combined effects of several single nucleotide polymorphisms were associated with breast cancer risk (14). A recent genome-wide association study (assayed 266,722 SNPs) demonstrated that five novel SNPs, rs2981582 (*FGFR2*), rs3803662 (*TNRC9/LOC643714*), rs889312 (*MAP3K1*), rs13281615 (8q) and rs3817198 (*LSP1*) were associated with breast cancer risk, and assuming that the five most significant loci interact multiplicatively on disease risk, these loci explained an estimated 3.6% of the excess familial risk of breast cancer (15). Nevertheless, it is unlikely that these SNPs will be appropriate for predictive genetic testing, either alone or in combination with each other at this stage (15).

CD44 is a cell surface transmembrane glycoprotein and expressed in a variety of cells and tissues derived from hemopoietic, epithelial, endothelial, and mesodermal origins (16). Studies from our laboratory and elsewhere have shown that CD44 is involved in a number of physiological processes including lymphocyte migration and extravasation, lymph node homing, and lymphocyte activation and apoptosis (16–21). CD44 has a well documented role in tumor metastasis (22, 23). Recent experimental and clinical evidence shows that CD44 and its interaction with hyaluronan regulate breast cancer cell proliferation, migration, and invasion, as well as tumor-associated angiogenesis and these correlated with patient survival (24). Thus, we hypothesized that *CD44* polymorphisms may play an important role in breast cancer development.

A polymorphism in human *CD44* was identified by restriction digestion (25). To date, approximately 400 single nucleotide polymorphisms have been identified and are available in the online SNP database. However, *CD44* polymorphisms have not been extensively studied. It has been reported that a naturally occurring polymorphism of *CD44* coding sequence abolishes a hyaluronan-binding consensus sequence without preventing hyaluronan binding (26). *CD44* SNPs were selected for studying whether these polymorphisms were the risk factors for non-syndromic oral clefts (27). The role of *CD44* polymorphisms in breast cancer development is still unknown. Our investigation in this paper demonstrates that *CD44* polymorphism is associated with breast cancer and may be effectively used in the risk assessment, prediction, prevention, diagnosis and genetic epidemiological analysis of breast cancer.

## Patients and Methods

### Patient samples

All breast cancer patients in this study signed an Institutional Review Board-approved consent form. The blood samples, breast cancer specimens and adjacent normal breast tissues deposited in the Tissue Bank at South Carolina Cancer Center (Columbia, SC, USA) from 260 breast cancer patients during 2001 and 2008 were used in this research. Among breast cancer patients, 118 were Caucasian American (CA) women, 115 African-American (AA) women and 27 unknown ethnic women. The patients had breast cancers from stage I to stage IIIC. The patients were aged between 23 and 90, and had an average age of 55 years at the time of diagnosis with breast cancer.

### Normal donors

All normal female donors signed an institutional review board-approved consent form and recruited in South Carolina Women's Care Study (Columbia, SC, USA) where they donated blood samples for our research. Among normal donors, 96 were Caucasian American women and 96 African-American women. In addition, genomic DNA samples from 40 healthy Caucasian American female normal donors provided by BioChain Institute, Inc. (Hayward, CA, USA) were also used as normal controls.

### Genomic DNA isolation

The genomic DNA was isolated from peripheral blood samples derived from female normal donors and breast cancer patients using the AGENCOURT® GENFIND™ v2 Blood & Serum Genomic DNA Isolation kit (Agencourt, Beverly, MA, USA) according to the manufacture's manual. The DNA concentration of genomic DNA samples was determined by spectrophotometer and the quality of genomic DNA samples was analyzed by agarose gel electrophoresis.

### Analysis of *CD44* polymorphisms

Direct DNA sequencing method was used to determine the polymorphisms in the *CD44* exon2 region including the *CD44* exon2 sequence and the boundary sequences between exon2 and upstream intron, and between exon2 and downstream intron. Briefly, 5 ng of genomic DNA from each sample were added into 50 µl of PCR reaction mixture. M13R-

tagged forward primer (5'-CCGGCCTTATTTGACTTTTTTAAGGAGTCTG-3') and M13F-tagged reverse primer (5'-CTCCAGTTGTCATACAGGTTGCA GATTGAC-3') were used to amplify the *CD44* exon2 region from genomic DNA samples by PCR using high fidelity DNA polymerase (Invitrogen, Carlsbad, CA, USA). The PCR products were sent out to MCLAB (South San Francisco, CA, USA) for cleanup and direct sequencing using primers M13R and M13F. At the same time, the PCR products were cloned into pCR2.1 TOPO vector (Invitrogen), transformed into One Short<sup>®</sup> TOP10F' cells (Invitrogen), and sent out to MCLAB for colony sequencing. After cloning and transformation, plasmid DNA samples were also prepared from certain number of samples (including 60 breast cancer patient samples and 40 normal donors) using QIAprep<sup>®</sup> Spin Miniprep kit (Qiagen, Valencia, CA, USA), and sent out to SeqWright (Houston, TX, USA) for sequencing confirmation. Sequencing analysis using BLAST searching for SNPs was carried out to determine *CD44* polymorphisms in the *CD44* exon2 region.

### FACS analysis of CD44 expression on peripheral blood mononuclear cells

Peripheral blood mononuclear cells (PBMC) were isolated from peripheral blood samples of normal donors using BioWhittaker Lymphocyte Separation Medium (Cambrex, Walkersville, MD, USA). PBMC were then stained with FITC-conjugated CD44 monoclonal antibody (BD, Franklin Lakes, NJ, USA) and CD44 expression was analyzed by Cytomics FC500 flow cytometer (Beckman Coulter, Fullerton, CA, USA). At the same time, CD44 polymorphic changes in the PBMC samples were also determined using direct DNA sequencing method. The association between *CD44* polymorphisms and CD44 expression on PBMC was determined by statistical analysis.

### Real-time PCR analysis of CD44 expression

PBMC were isolated from peripheral blood samples of normal donors using BioWhittaker Lymphocyte Separation Medium (Cambrex). Total RNA was prepared from PBMC samples using RNeasy Mini kit (Qiagen). 100 ng RNA in 20 µl reaction mixture each sample were used in cDNA synthesis using ThermoScript<sup>™</sup> RT-PCR system (Invitrogen). 1 µl cDNA in 20 µl SYBR<sup>®</sup> GREEN PCR mixture (Applied Biosystems, Foster City, CA, USA) each sample was used in real-time PCR analysis of CD44 expression by StepOne<sup>™</sup> Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) using forward primer (5'-CTGCAAGGCTTTCAATAGCACCTT-3') and reverse primer (5'-AATCAC CACGTGCCCTTCTATGAA-3') according to the manufacture's instruction. Relative level of CD44 expression was normalized by the number of CD44 molecules per nanogram of total RNA using human CD44 gene construct as a standard.

### Statistical analysis

Odds Ratio Generator Version 1.0.0 (Devilley G.J., 2005, Centre for Neuropsychology, Swinburne University, Australia) was used to calculate odds ratio (OR) and 95% confidence interval (CI) to analyze the statistical differences of *CD44* polymorphisms between two groups. The odds ratio ranges from 0 to positive infinity, with 1.0 indicating that the condition or event under study is equally likely in both groups. An odds ratio greater than 1 indicates that the condition or event is more likely in the second group. Two-sample

proportion test was used to confirm the statistical differences. Proportion trend test was used to correct the statistics for multiple hypothesis testing. Fisher's exact test and Wilcoxon-Mann-Whitney test was used to determine statistical difference between two groups, and the difference was considered significant at  $p < 0.05$ .

Hardy-Weinberg equilibrium (HWE) test of SNP was performed using Michael H. Court's (2005–2008) online calculator (<http://www.tufts.edu/~mcourt01/Documents/Court%20lab%20%20HW%20calculator.xls>). Tests in breast cancer patients ( $p=0.4962$ ) and female normal donors ( $p=0.7222$ ) did not show any significant deviation from HWE for any of the SNPs.

## Results

### Unique SNP in CD44 intron 1

The human *CD44* gene is located at the short arm of chromosome 11, containing 20 exons spanning about 50 kilobases of DNA. The gene is composed of two groups of exons. The exons 1–5 and 16–20, are expressed together as the standard form, designated CD44s. The 10 variable exons (exons 6–15) can be alternatively spliced and included within the standard exons at an insertion site between exons 5 and 16. *CD44* exon2 is critical for CD44 binding to its ligand, hyaluronan (28). Thus, we were interested in understanding the polymorphisms in the *CD44* exon 2 region and their role in breast cancer development. The direct nucleotide sequencing analysis indicated that a unique polymorphism in the boundary region between *CD44* exon 2 and its upstream intron (intron 1) was present in breast cancer patients, which was located in the upstream intron of *CD44* exon 2. The polymorphic change was A→G, and its position was at +14 from *CD44* exon2 (Figure 1). We designated this unique *CD44* polymorphism as *CD44* Ex2+14 A>G polymorphism. Both heterozygous SNP (Figure 1B) and homozygous SNP (Figure 1C) were identified in breast cancer patients. When compared to female healthy normal donors, female breast cancer patients had a significantly increased frequency of the unique SNP in *CD44* intron 1 (Figure 2A) (odds ratio=2.43; 95% confidence interval=1.58–3.74;  $p < 0.0001$ ). Caucasian American female breast cancer patients had a significantly higher percentage of the unique SNP than did Caucasian American female normal donors (Figure 2B) (odds ratio=2.23; 95% confidence interval=1.28–3.88;  $p < 0.005$ ). African-American female breast cancer patients also had a significantly higher percentage of the unique SNP than did African-American female normal donors (Figure 2C) (odds ratio=2.67; 95% confidence interval=1.15–6.19;  $p < 0.05$ ). In addition, the patients having the unique SNP were diagnosed with breast cancer at relative younger ages (Figure 2D). The results suggested that the unique SNP in *CD44* intron 1 may be associated with the risk for breast cancer.

### Comparison of unique SNP in CD44 intron 1 between Caucasian Americans and African-Americans

It has been reported that African-Americans have higher mortality from breast cancer than Caucasian Americans although the incidence rate of breast cancer in African-Americans is lower than in Caucasian Americans (29). It is valuable to determine the difference in the unique SNP in *CD44* intron 1 between Caucasian Americans and African-Americans. The

direct nucleotide sequencing analysis revealed that African-American female normal donors had a significantly higher percentage of the unique SNP than Caucasian American female normal donors (Figure 3A) (odds ratio=2.65; 95% confidence interval=1.42–4.92;  $p<0.005$ ). African-American female breast cancer patients also had a significantly higher percentage of the unique SNP than Caucasian American female breast cancer patients (Figure 3B) (odds ratio=3.17; 95% confidence interval=1.43–7.01;  $p<0.01$ ). Proportion trend test confirmed that the frequency of the unique SNP in *CD44* intron 1 was highest in African-American breast cancer patients and the lowest in Caucasian American normal donors among four groups ( $p=0.0008$ ). In particular, African-American female normal donors had an approximately three-fold increased frequency of homozygous unique SNP as compared with Caucasian American normal donors (56% and 22%, respectively; Figure 3C). Similarly, African-Americans were diagnosed with breast cancer at significantly earlier ages than were Caucasian Americans (Figure 4D). These data indicated that the unique SNP in *CD44* intron 1 may correlate with breast cancer development.

### Characteristics of breast cancer patients having heterozygous and homozygous unique SNP in CD44 intron 1

A positive correlation of rs13181 (*ERCC2/XPD* Lys751Gln/A18911C polymorphism) heterozygous genotype with the risk of breast cancer was observed upon consideration of interactions between the mutant genotypes and anthropometric or lifestyle factors (30, 31). Further study on the association of the SNP rs13181 with predisposition to breast cancer showed a significant to highly significant positive association of greater than 2-fold for the rs13181 homozygous mutant (CC) (OR 4.412, 95% CI 2.413 to 8.068,  $p<0.0001$ ), heterozygous (AC) (OR 2.086, 95% CI 1.246 to 3.492,  $p=0.0056$ ) and combined mutant (AC+CC) (OR 2.672, 95% CI 1.647 to 4.334,  $p<0.0001$ ) genotypes (32). It is still possible that the effects of heterozygous genotype on cancer risk may be different from those of homozygous genotype. Our analysis showed that both heterozygous and homozygous unique *CD44* SNP carriers were diagnosed with breast cancer at younger ages than non-unique *CD44* SNP carriers (Figure 4A), however, the patients having the homozygous genotype of unique SNP in *CD44* intron 1 had a larger tumor burden (Figure 4B) and higher number of regional lymph nodes involved in cancer spread (Figure 4C) as compared with those having the heterozygous genotype of unique SNP and wild-type genotype when they were diagnosed with breast cancer. Moreover, the patients having the homozygous genotype of unique SNP in *CD44* intron 1 had difficulty in responding to treatment (Figure 4D) and a higher rate of cancer recurrence (Figure 4E). The results demonstrated that the homozygous genotype of unique SNP in *CD44* intron 1 correlates with more favorable breast cancer growth and development.

### Association of unique SNP in CD44 intron 1 and CD44 expression

When compared to female healthy normal donors, female breast cancer patients had a significantly increased frequency of the unique SNP in *CD44* intron 1 (Figure 2), suggesting that this unique *CD44* SNP may also play an important role in breast cancer development. The unique *CD44* SNP was located in the upstream intron of *CD44* exon2. The polymorphic change was A→G, and its position was at +14 from *CD44* exon 2 (Figure 1). *CD44* is a multi-functional adhesion molecule that can undergo alternative RNA splicing to generate



multiple isoforms bearing different cellular functions. RNA splicing is a complex process, which is catalyzed by the spliceosome, a large RNA-protein complex composed of five small nuclear ribonucleoproteins (snRNPs) (33). In the 5-splice site of pre-mRNA, the GU dinucleotide is followed by a less conserved consensus sequence, whereas the 3-splice site consists of a branch point most often containing an adenine residue, followed by the pyrimidine-rich region and the conserved AG dinucleotide (34). These consensus sequences are critical for successful RNA splicing. The unique SNP in *CD44* intron 1 may change the consensus sequence of the branch point at the 3' splicing site of *CD44* pre-mRNA and thereby affect *CD44* splicing, leading to decreased *CD44* expression. Thus, we explored the possibility that this unique SNP in the upstream intron region of *CD44* exon2 may affect *CD44* expression. Fluorescence-activated cell sorting (FACS) analysis showed that PBMC from female healthy normal donors having the homozygous genotype of the unique SNP in *CD44* intron 1 had significantly decreased *CD44*<sup>high</sup> cell populations than normal donors who did not have the unique SNP ( $22.72 \pm 1.80\%$  vs.  $41.30 \pm 2.40\%$ ; Figure 5A, 5C and 5D). The mean intensity of fluorescence for *CD44* expression was also less in individuals containing the unique SNP (Figure 5E). Real-time PCR assay also confirmed that PBMC from normal donors having the unique SNP in *CD44* intron 1 had a decreased *CD44* expression level (Figure 5F). PBMC from normal donors having the heterozygous genotype of the unique SNP in *CD44* intron 1 had a medium level of *CD44* expression (Figure 5B, 5D, 5E and 5F). These results suggested that there was an association between *CD44* polymorphism and decreased *CD44* expression. Considering our observation that the down-regulation of *CD44* expression on tumor-infiltrating lymphocytes by RNA interference technique decreased anti-tumor killing *in vitro* (unpublished observation), we propose that the unique *CD44* Ex2+14 A>G polymorphism affects normal *CD44* splicing, decreases *CD44* expression, inhibits anti-tumor immunity and facilitates breast cancer development.

## Discussion

Because *CD44* is involved in both immune surveillance and tumor metastasis (16, 22, 23, 35), it is not surprising that our results demonstrate that the unique single nucleotide polymorphism in *CD44* intron 1 is associated with breast cancer development. The association between *CD44* polymorphisms and breast cancer risk, however, has not been reported yet. For instance, analysis of a large number of SNPs (710 SNPs in 120 candidate genes) suggested that no single gene polymorphism was associated with breast cancer risk and the combined effects of several SNPs were associated with breast cancer risk (14). Genome-wide association study (assayed 266,722 SNPs) demonstrated that five novel SNPs, rs2981582 (*FGFR2*), rs3803662 (*TNRC9/LOC643714*), rs889312 (*MAP3KI*), rs13281615 (8q) and rs3817198 (*LSP1*) were associated with breast cancer risk (15). The difference between this investigation and the previous studies may be due to different methods used in analyzing gene polymorphisms. In the first report for the analysis of a large number of SNPs (14), an empirical SNP tagging approach by Taqman assay and oligonucleotide array methodology was used to capture common genetic variations in genes that are candidates for breast cancer based on their known function. In genome-wide association study (15), high-density oligonucleotide microarrays, Taqman assay and MALDI-TOF mass spectrometry were used to (query known SNPs to) identify novel breast cancer susceptibility loci. In

contrast, direct DNA sequencing method was used in this study to identify novel SNPs and perform fine *CD44* polymorphism analysis in the *CD44* exon2 region. Our investigations demonstrate for the first time that *CD44* polymorphism is associated with breast cancer.

So far, a number of breast cancer susceptibility genes have been identified. However, it is estimated that all the currently known breast cancer susceptibility genes, including *BRCA1* and *BRCA2* (1, 2), account for less than 25% of the familial aggregation of breast cancer (5). The data from the current study showed that the unique *CD44* polymorphism (*CD44* Ex2+14 A>G) in the *CD44* exon 2 upstream intron (intron 1) was identified in 84% of breast cancer patients and 68% of healthy normal donors (Figure 2A). Thus, *CD44* may represent one of the high penetrance genes for the genetic risk of breast cancer.

Recent statistics showed that although African-Americans have a lower incidence rate of breast cancer than whites, they have the highest mortality rate of any racial or ethnic group (29). For example, the incidence of breast cancer in South Carolina is lower in African-Americans compared with European American women by approximately 12% to 15%, but their mortality rate is twice as high as in European American women (36). Breast cancer in African-American women has been characterized by higher grade, later stage at diagnosis, and worse survival, even after controlling for stage at diagnosis. The Carolina Breast Cancer Study showed that the basal-like breast cancer subtype was more prevalent among premenopausal African-American women (39%) compared with postmenopausal African-American women (14%) and non-African-American women (16%) of any age ( $p<0.001$ ) and basal-like tumors had a high frequency of TP53 mutations (37). Our investigations revealed that African-Americans had a significantly increased frequency of the unique SNP in *CD44* intron 1, especially for the homozygous genotype of the unique SNP, than Caucasian Americans (Figure 3A, 3B and 3C). The high percentage of homozygous unique *CD44* SNP genotype in African Americans may represent one of the factors which is responsible for the high mortality rate of African-American breast cancer patients.

It has been reported that breast cancer patients more frequently carried the homozygous genotype AA of SNP rs1467465 of differentiation-associated human gene *ICB-1* (*C1orf38*) than did healthy women. Analysis of allele positivity revealed that AG or GG genotypes were significantly less frequent in breast cancer patients, suggesting that presence of G allele might have protective effects (38). A significantly higher risk of breast cancer was observed among homozygous carriers of *ESR1325* CC genotype (39) and heterozygous carriers of *HER-2/neu655* Ile/Val genotype (40). A positive correlation of rs13181 (*ERCC2/XPD* Lys751Gln/A18911C polymorphism) heterozygous genotype with the risk of breast cancer was observed upon consideration of interactions between the mutant genotypes and anthropometric or lifestyle factors (30, 31). Therefore, heterozygous and homozygous genotypes may have different impacts on breast cancer. The breast cancer patients having a homozygous unique *CD44* SNP genotype had a significantly larger tumor size (Figure 4B), higher number of regional lymph nodes involved in cancer spread (Figure 4C), lower curative rate (Figure 4D) and higher recurrence incidence (Figure 4E), suggesting that homozygous genotype of the unique SNP in *CD44* intron 1 facilitates breast cancer growth and metastasis.



Human *CD44* gene is composed of 20 exons, making it difficult to analyze *CD44* polymorphisms in all *CD44* exons. *CD44* function mainly depends on its binding to its ligands such as hyaluronan. *CD44* has two binding domains involved in the interaction between *CD44* and hyaluronan, one in exon 2 region and another in exon 5 region (26). Thus, at first, we analyzed *CD44* polymorphisms in exon 2 region. In the current study, the unique SNP in *CD44* intron 1 was identified in breast cancer patients. However, we cannot rule out that *CD44* polymorphisms in other exon regions may also play a critical role in breast cancer development. Additionally, analysis of *CD44* polymorphisms in a larger sample size is necessary to confirm the association between *CD44* polymorphisms and breast cancer risk. Nevertheless, this study provides the evidence for the first time that *CD44* may represent one of the high penetrance genes for the genetic risk of breast cancer and *CD44* polymorphisms may be effectively used in the risk assessment, prediction, prevention, diagnosis and genetic epidemiological analysis of breast cancer.

## Acknowledgments

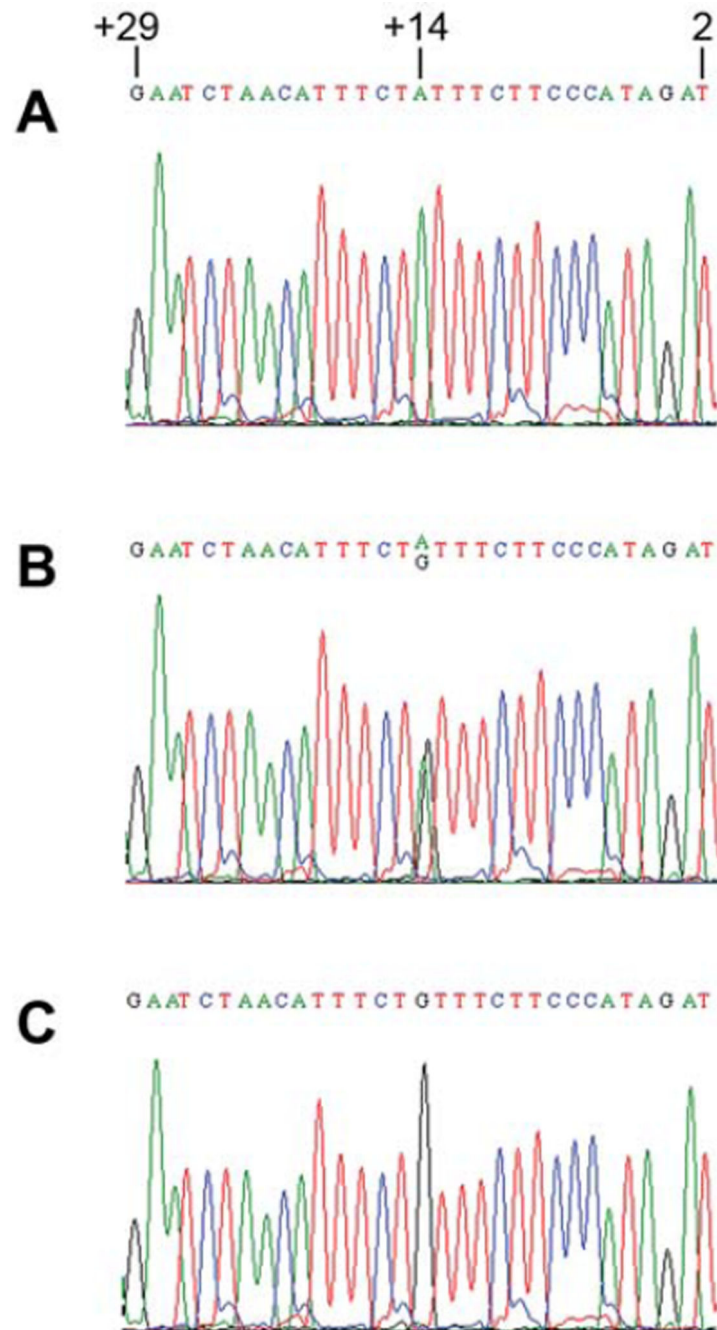
We thank Dr. Phillip J. Buckhaults and Ms. Ella S. Weinkle for assistance in obtaining genomic DNA samples, blood samples, breast cancer specimens and adjacent normal breast tissues from breast cancer patients in the Tissue Bank at South Carolina Cancer Center. This study was supported by the Innovative and Exploratory Grant Program (IEGP) of University of South Carolina School of Medicine (J. Zhou), NIH grants R01ES09098, R01DA016545 and P01AT003961 (P. Nagarkatti) and NIH grants R01AI053703, R01AI058300, and R01HL058641 (M. Nagarkatti).

## References

1. Peto J, Collins N, Barfoot R, et al. Prevalence of *BRCA1* and *BRCA2* gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst.* 1999; 91:943–949. [PubMed: 10359546]
2. Anglian Breast Cancer Study Group: Prevalence and penetrance of *BRCA1* and *BRCA2* mutations in a population-based series of breast cancer cases. *Br J Cancer.* 2000; 83:1301–1308. [PubMed: 11044354]
3. Ford D, Easton DF, Stratton M, et al. Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet.* 1998; 62:676–689. [PubMed: 9497246]
4. Antoniou AC, Easton DF. Models of genetic susceptibility to breast cancer. *Oncogene.* 2006; 25:5898–5905. [PubMed: 16998504]
5. Thompson D, Easton D. The genetic epidemiology of breast cancer genes. *J Mammary Gland Biol Neoplasia.* 2004; 9:221–236. [PubMed: 15557796]
6. Dunning AM, Healey CS, Pharoah PD, et al. A systematic review of genetic polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev.* 1999; 8:843–854. [PubMed: 10548311]
7. Kadouri L, Easton DF, Edwards S, et al. CAG and GGC repeat polymorphisms in the androgen receptor gene and breast cancer susceptibility in *BRCA1/2* carriers and non-carriers. *Br J Cancer.* 2001; 85:36–40. [PubMed: 11437399]
8. De Vivo I, Hankinson SE, Colditz GA, et al. A functional polymorphism in the progesterone receptor gene is associated with an increase in breast cancer risk. *Cancer Res.* 2003; 63:5236–5238. [PubMed: 14500352]
9. Huggins GS, Wong JY, Hankinson SE, et al. GATA5 activation of the progesterone receptor gene promoter in breast cancer cells is influenced by the +331G/A polymorphism. *Cancer Res.* 2006; 66:1384–1390. [PubMed: 16452193]
10. Feigelson HS, Rodriguez C, Jacobs EJ, et al. No association between the progesterone receptor gene +331G/A polymorphism and breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2004; 13:1084–1085. [PubMed: 15184270]

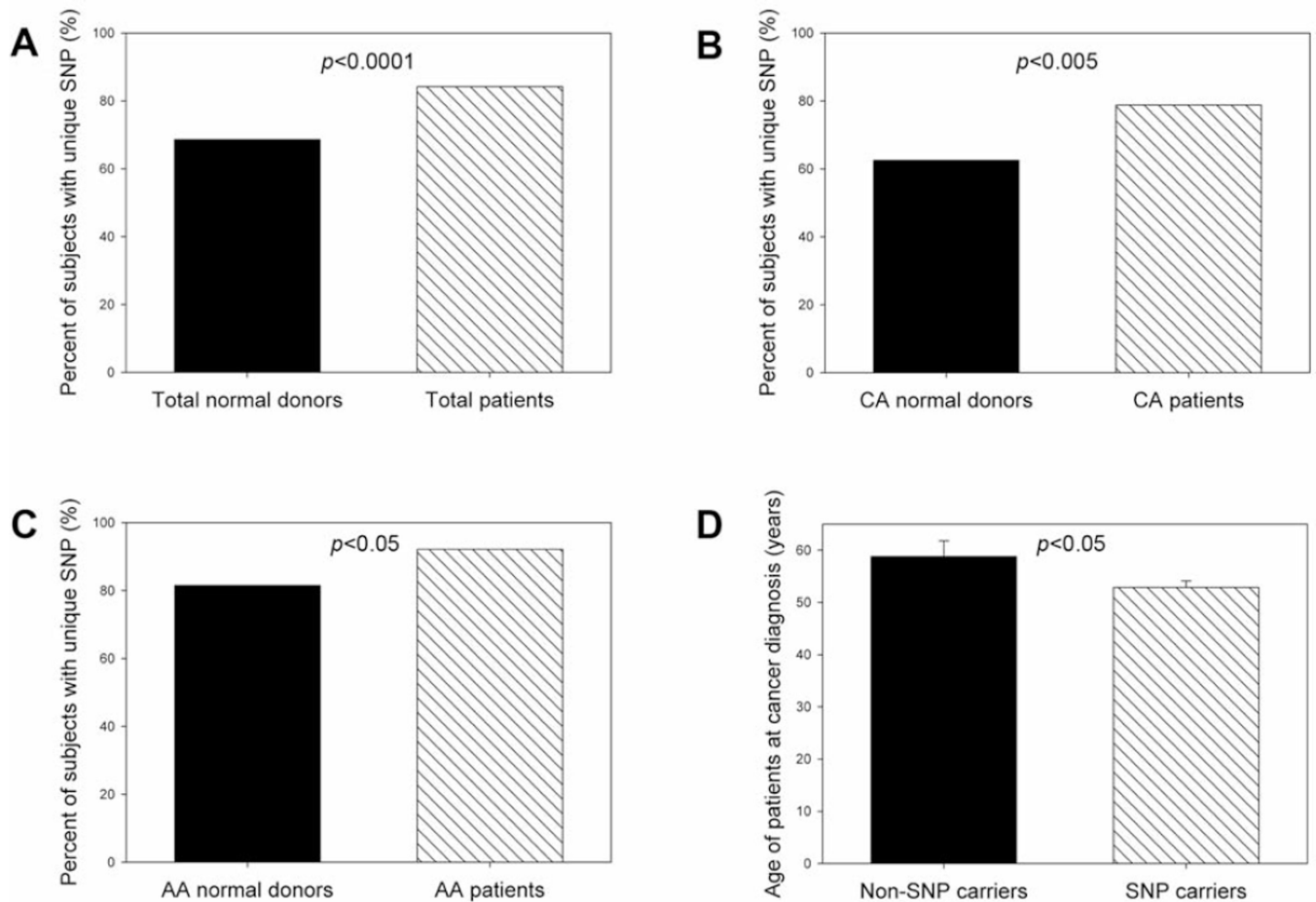
11. Fernandez LP, Milne RL, Barroso E, et al. Estrogen and progesterone receptor gene polymorphisms and sporadic breast cancer risk: a Spanish case-control study. *Int J Cancer*. 2006; 119:467–471. [PubMed: 16477637]
12. Johnatty SE, Spurdle AB, Beesley J, et al. Progesterone receptor polymorphisms and risk of breast cancer: results from two Australian breast cancer studies. *Breast Cancer Res Treat*. 2008; 109:91–99. [PubMed: 17592773]
13. Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in *CASP8* is associated with breast cancer risk. *Nat Genet*. 2007; 39:352–358. [PubMed: 17293864]
14. Pharoah PD, Tyrer J, Dunning AM, et al. Association between common variation in 120 candidate genes and breast cancer risk. *PLoS Genet*. 2007; 3:e42. [PubMed: 17367212]
15. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007; 447:1087–1093. [PubMed: 17529967]
16. Bourguignon LY, Zhu D, Zhu H. CD44 isoform-cytoskeleton interaction in oncogenic signaling and tumor progression. *Front Biosci*. 1998; 3:d637–D649. [PubMed: 9634539]
17. Seth A, Gote L, Nagarkatti M, et al. T-cell-receptor-independent activation of cytolytic activity of cytotoxic T lymphocytes mediated through CD44 and gp90MEL-14. *Proc Natl Acad Sci USA*. 1991; 88:7877–7881. [PubMed: 1881921]
18. Rafi A, Nagarkatti M, Nagarkatti PS. Hyaluronate-CD44 interactions can induce murine B-cell activation. *Blood*. 1997; 89:2901–2908. [PubMed: 9108410]
19. Chen D, McKallip RJ, Zeytun A, et al. CD44-deficient mice exhibit enhanced hepatitis after concanavalin A injection: evidence for involvement of CD44 in activation-induced cell death. *J Immunol*. 2001; 166:5889–5897. [PubMed: 11342603]
20. McKallip RJ, Fisher M, Do Y, et al. Targeted deletion of CD44v7 exon leads to decreased endothelial cell injury but not tumor cell killing mediated by interleukin-2-activated cytolytic lymphocytes. *J Biol Chem*. 2003; 278:43818–43830. [PubMed: 12904302]
21. McKallip RJ, Fisher M, Gunthert U, et al. Role of CD44 and its v7 isoform in staphylococcal enterotoxin B-induced toxic shock: CD44 deficiency on hepatic mononuclear cells leads to reduced activation-induced apoptosis that results in increased liver damage. *Infect Immun*. 2005; 73:50–61. [PubMed: 15618140]
22. Marhaba R, Zoller M. CD44 in cancer progression: adhesion, migration and growth regulation. *J Mol Histol*. 2004; 35:211–231. [PubMed: 15339042]
23. Hill A, McFarlane S, Johnston PG, et al. The emerging role of CD44 in regulating skeletal micrometastasis. *Cancer Lett*. 2006; 237:1–9. [PubMed: 15979783]
24. Gotte M, Yip GW. Heparanase, hyaluronan, and CD44 in cancers: a breast carcinoma perspective. *Cancer Res*. 2006; 66:10233–10237. [PubMed: 17079438]
25. Dadi HK, Greaves AM, Cox DW, et al. Identification of a polymorphism in human CD44. *Nucleic Acids Res*. 1991; 19:969. [PubMed: 2017391]
26. Telen MJ, Udani M, Washington MK, et al. A blood group-related polymorphism of CD44 abolishes a hyaluronan-binding consensus sequence without preventing hyaluronan binding. *J Biol Chem*. 1996; 271:7147–7153. [PubMed: 8636151]
27. Park JW, Cai J, McIntosh I, et al. High throughput SNP and expression analyses of candidate genes for non-syndromic oral clefts. *J Med Genet*. 2006; 43:598–608. [PubMed: 16415175]
28. Peach RJ, Hollenbaugh D, Stamenkovic I, et al. Identification of hyaluronic acid binding sites in the extracellular domain of CD44. *J Cell Biol*. 1993; 122:257–264. [PubMed: 8314845]
29. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2008. *CA Cancer J Clin*. 2008; 58:71–96. [PubMed: 18287387]
30. Bernard-Gallon D, Bosviel R, Delort L, et al. DNA repair gene *ERCC2* polymorphisms and associations with breast and ovarian cancer risk. *Mol Cancer*. 2008; 7:36. [PubMed: 18454848]
31. Terry MB, Gammon MD, Zhang FF, et al. Polymorphism in the DNA repair gene *XPB*, polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2004; 13:2053–2058. [PubMed: 15598760]
32. Mitra AK, Singh N, Garg VK, et al. Statistically significant association of the single nucleotide polymorphism (SNP) rs13181 (*ERCC2*) with predisposition to squamous cell carcinomas of the

- head and neck (SCCHN) and breast cancer in the north Indian population. *J Exp Clin Cancer Res.* 2009; 28:104. [PubMed: 19615095]
33. Nagai K, Muto Y, Pomeranz Krummel DA, et al. Structure and assembly of the spliceosomal snRNPs. Novartis Medal Lecture. *Biochem Soc Trans.* 2001; 29:15–26. [PubMed: 11356120]
34. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem.* 2003; 72:291–336. [PubMed: 12626338]
35. Rafi-Janajreh AQ, Nagarkatti PS, Nagarkatti M. Role of CD44 in CTL and NK cell activity. *Front Biosci.* 1998; 3:d665–d671. [PubMed: 9634546]
36. Smith ER, Adams SA, Das IP, et al. Breast cancer survival among economically disadvantaged women: the influences of delayed diagnosis and treatment on mortality. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:2882–2890. [PubMed: 18835941]
37. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA.* 2006; 295:2492–2502. [PubMed: 16757721]
38. Springwald A, Lattrich C, Seitz S, et al. Single nucleotide polymorphisms in human gene ICB-1 and breast cancer susceptibility. *Cancer Invest.* 2009; 27:669–672. [PubMed: 19308812]
39. Siddig A, Mohamed AO, Awad S, et al. Estrogen receptor alpha gene polymorphism and breast cancer. *Ann NY Acad Sci.* 2008; 1138:95–107. [PubMed: 18837889]
40. Siddig A, Mohamed AO, Kamal H, et al. *HER-2/neu* Ile655Val polymorphism and the risk of breast cancer. *Ann NY Acad Sci.* 2008; 1138:84–94. [PubMed: 18837888]



**Figure 1.** Identification of unique SNP in CD44 intron 1 in breast cancer patients. A chromatogram of wild-type sequences (A), and heterozygous (B) or homozygous SNP (C) sequences in CD44 intron 1 in the upstream of CD44 exon2 identified in breast cancer patients. Unique CD44 polymorphic change from A to G was seen at position +14 in the upstream of CD44 exon2 coding region. This unique CD44 single nucleotide polymorphism was designated as CD44 Ex2+14 A>G polymorphism. Number 2 represented the second nucleotide at the 5'-end of

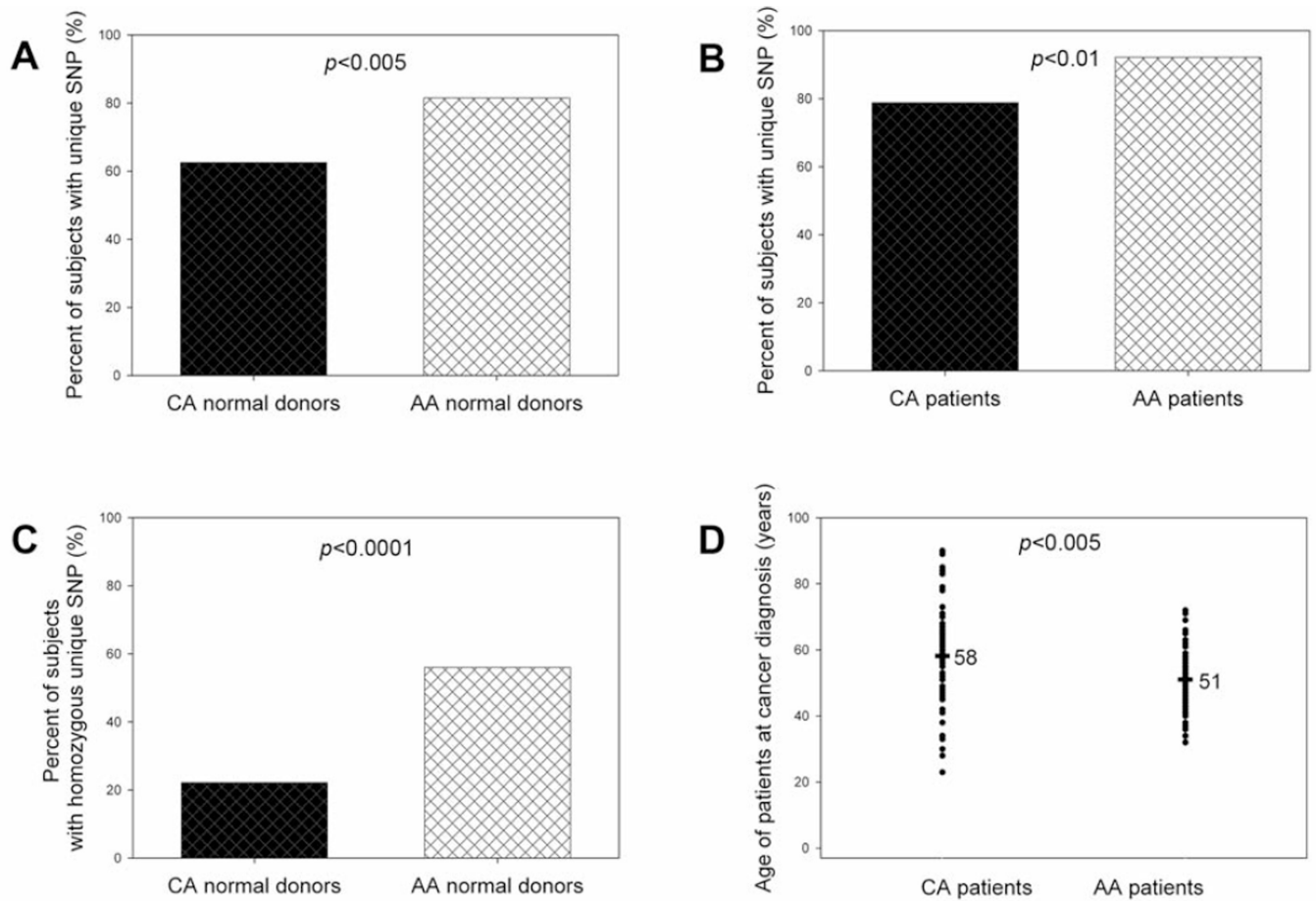
CD44 exon 2, whereas number +29 represented the nucleotide at position +29 in the upstream of CD44 exon2 (intron 1).



**Figure 2.**

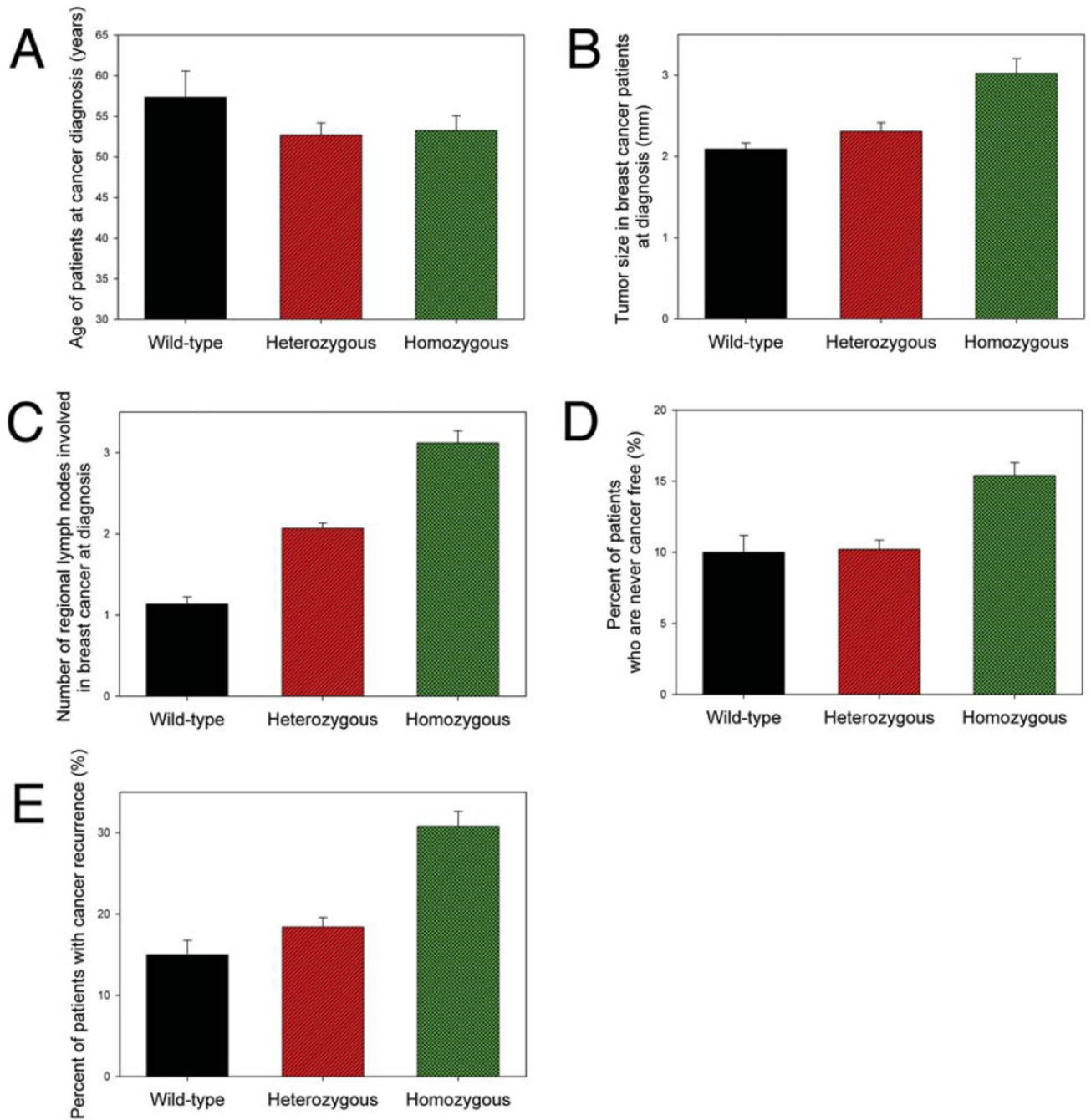
Frequency of unique SNP in CD44 intron 1 (CD44 Ex2+14 A>G) in female normal donors and breast cancer patients. Comparisons of percentages of individuals having the unique SNP in CD44 intron1 were made between: (A) total normal donors versus breast cancer patients, (B) Caucasian American (CA) normal donors versus CA breast cancer patients, (C) African-American (AA) normal donors and AA breast cancer patients, (D) non-unique SNP carriers versus unique SNP carriers when the patients were diagnosed with breast cancer. Fisher's exact test was used to determine the statistical difference of unique SNP frequencies between two groups. Wilcoxon-Mann-Whitney test was used to determine the statistical differences of patient ages between two groups.



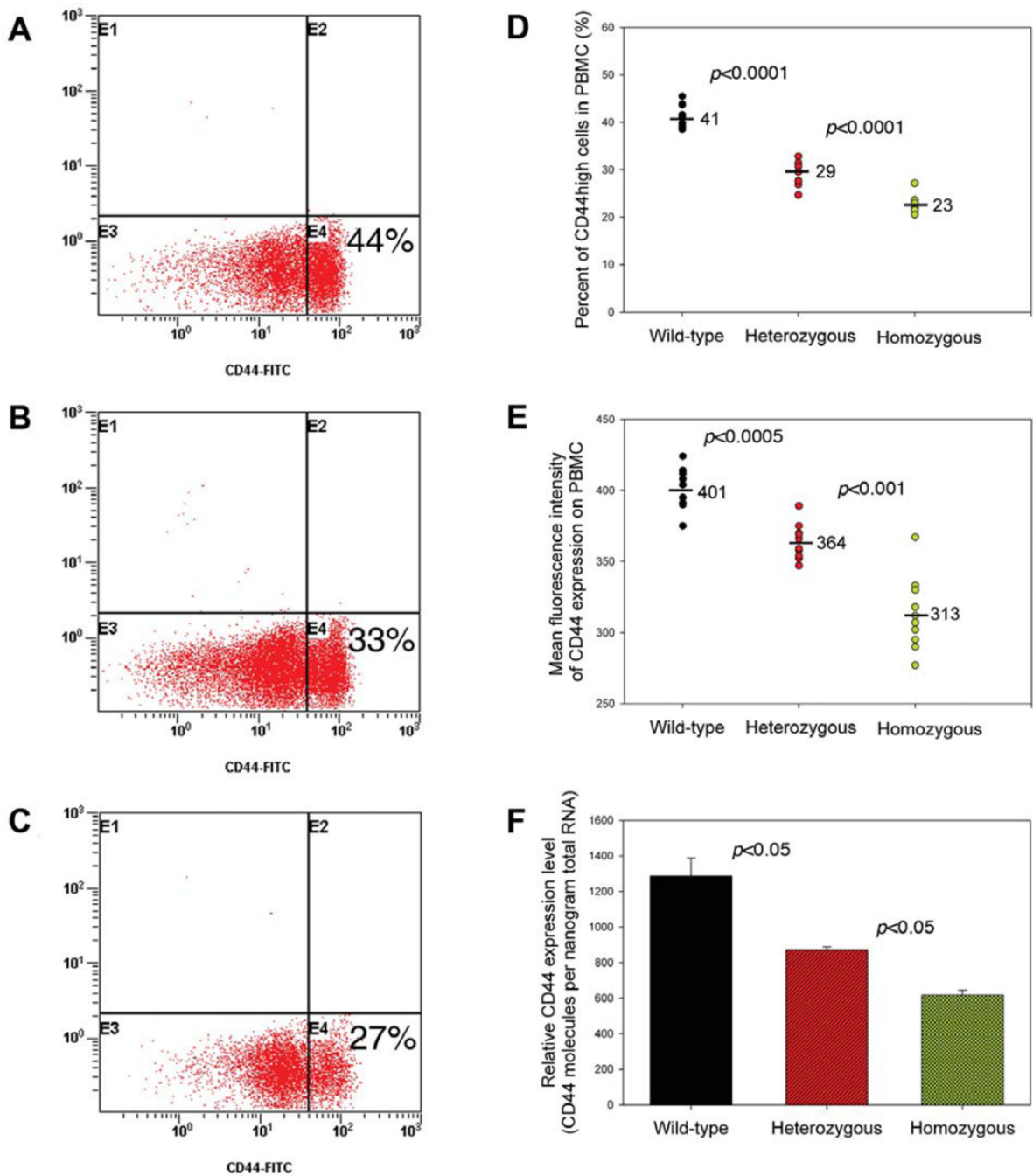


**Figure 3.**

Difference of frequencies of unique SNP in CD44 intron 1 (CD44 Ex2+14 A>G) between Caucasian American (CA) and African-American (AA) females. Comparisons of percentages of subjects having the unique SNP in CD44 intron1 were made between: (A) CA female normal donors and AA female normal donors, (B) CA female breast cancer patients and AA breast cancer patients, (C) CA female normal donors and AA female normal donors, (D) CA female breast cancer patients and AA breast cancer patients when the patients were diagnosed with breast cancer. Fisher's exact test was used to determine the statistical difference of unique SNP frequencies between two groups. Wilcoxon-Mann-Whitney test was used to determine the statistical differences of patient ages between two groups.



**Figure 4.** Difference of characteristics of breast cancer patients who did not have unique SNP in CD44 intron 1 (CD44 Ex2+14 A>G) and those that had heterozygous or homozygous unique SNP. The clinical characteristics included: (A) patient's age at cancer diagnosis, (B) size of tumor at diagnosis, (C) number of regional metastatic lymph nodes involved when the patients were diagnosed with breast cancer, (D) patients who were never cancer free, and (E) patients having cancer recurrence.



**Figure 5.** Association of unique SNP in CD44 intron 1 (CD44 Ex2+14 A>G) and CD44 expression on peripheral blood mononuclear cells (PBMC) from healthy normal donors. Representative histogram of CD44 expression on PBMC from normal donors who: (A) did not have the unique SNP, (B) had heterozygous unique SNP, and (C) had homozygous unique SNP. PBMC from female healthy normal donors were stained with FITC-conjugated anti-human CD44 antibody and analyzed by flow cytometer. The percentages of CD44<sup>high</sup> and CD44<sup>low</sup> PBMC populations are shown. (D) Comparison of percentages of CD44<sup>high</sup> PBMC

populations among normal donors who did not have the unique SNP, those having heterozygous unique SNP and those having homozygous unique SNP. (E) Comparison of mean fluorescence intensity of CD44 expression on PBMC among normal donors who did not have the unique SNP, and those having heterozygous or homozygous unique SNP. (F) Comparison of levels of CD44 expression on PBMC, as determined by real-time PCR assay, among normal donors who did not have the unique SNP, and those having heterozygous or homozygous unique SNP.