Published in final edited form as:

Breast Cancer Res Treat. 2008 July ; 110(2): 367–376. doi:10.1007/s10549-007-9720-7.

Genetic polymorphisms in uridine diphospho-glucuronosyltransferase 1A1 and breast cancer risk in Africans

Dezheng Huo,
Department of Health Studies, University of Chicago, Chicago, IL, USA

Hee-Jin Kim,
Department of Laboratory Medicine & Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Clement A. Adebamowo,
Department of Surgery, University of Ibadan and University College Hospital, Ibadan, Nigeria

Temidayo O. Ogundiran,
Department of Surgery, University of Ibadan and University College Hospital, Ibadan, Nigeria

Effiong E. Akang,
Department of Pathology, University of Ibadan and University College Hospital, Ibadan, Nigeria

Oladapo Campbell,
Department of Radiotherapy, University of Ibadan and University College Hospital, Ibadan, Nigeria

Adeniyi Adenipekun,
Department of Radiotherapy, University of Ibadan and University College Hospital, Ibadan, Nigeria

Qun Niu,
Department of Medicine, Section of Hematology/Oncology, University of Chicago, 5841 S. Maryland Ave, MC 2115, Chicago, IL 60637, USA

Lise Sveen,
Department of Medicine, Section of Hematology/Oncology, University of Chicago, 5841 S. Maryland Ave, MC 2115, Chicago, IL 60637, USA

James D. Fackenthal,
Department of Medicine, Section of Hematology/Oncology, University of Chicago, 5841 S. Maryland Ave, MC 2115, Chicago, IL 60637, USA

Donna Lee Fackenthal,
Department of Human Genetics, University of Chicago, Chicago, IL, USA

Soma Das,
Department of Human Genetics, University of Chicago, Chicago, IL, USA

Correspondence to: Olufunmilayo I. Olopade, folopade@medicine.bsd.uchicago.edu.

Nancy Cox,
Department of Medicine, Section of Genetic Medicine, University of Chicago, Chicago, IL 60637, USA

Anna Di Rienzo, and
Department of Human Genetics, University of Chicago, Chicago, IL, USA

Olufunmilayo I. Olopade
Department of Medicine, Section of Hematology/Oncology, University of Chicago, 5841 S. Maryland Ave, MC 2115, Chicago, IL 60637, USA; Department of Human Genetics, University of Chicago, Chicago, IL, USA

Olufunmilayo I. Olopade: folopade@medicine.bsd.uchicago.edu

Abstract

The UDP-glucuronosylatransferase 1A1 (*UGT1A1*) gene is involved in the metabolism of estrogen and detoxification of potential carcinogens. The number of TA repeats in the promoter region of *UGT1A1* has been linked to breast cancer risk, but results varied by race. We performed a comprehensive assessment of genetic polymorphisms in the *UGT1A1* gene, and examined these polymorphisms and TA repeats in relation to breast cancer risk in a case-control study in Nigeria. 512 breast cancer cases and 226 community controls were genotyped for *UGT1A1*. Compared with high-activity TA repeat genotypes, the odds ratios (OR) for low-activity and moderate-activity genotypes were 0.47 (95% confidence interval CI, 0.26–0.83) and 0.64 (95% CI, 0.39–1.06), respectively, in premenopausal women ($P = 0.009$ for trend), but no association was observed in postmenopausal women ($P = 0.24$). The effect of TA repeats was also differentiated by age: the OR was 0.39 (95% CI 0.21–0.71) for low-activity genotypes and 0.58 (95% CI 0.33–1.00) for moderate-activity genotypes in women <45 years old ($P = 0.002$ for trend), but no association was observed in women ≥45 years old ($P = 0.15$). Haplotype analysis showed that *UGT1A1* haplotypes were highly diverse with blocked structures. We found a specific haplotype in block 2 that was significantly associated with a 2.1-fold elevated risk (95% CI 1.05–4.39; $P = 0.04$). In contrast with previous studies, we found low-activity TA repeat alleles were protective against breast cancer among premenopausal indigenous Africans, suggesting that the role of *UGT1A1* in breast cancer development may vary by population, presumably due to different environmental and genetic modifier effects.

Keywords

African; Breast cancer; Case-control study; Genetic susceptibility; UGT1A1

Introduction

Uridine diphospho-glucuronosylatransferases (UGTs) catalyze the glucuronidation of a large number of substrates to form water-soluble and, in most cases, inactive metabolites that can be more easily excreted [1]. UGT1A1, a member of the UGT superfamily, is involved in the metabolism of estrogens and the detoxification of potential carcinogens [2, 3], and is expressed in mammary tissue [4]. It is also the primary isoenzyme responsible for glucuronidation of bilirubin, an antioxidant in blood. Therefore UGT1A1 has been

considered a good candidate gene in breast carcinogenesis. The number of TA repeats in the TATA box of the promoter region of *UGT1A1* gene has been shown to be inversely associated with its transcriptional activity in vitro, with 5 and 6 repeats (*36 and *1, respectively) for high activity, and 7 and 8 repeats (*28 and *37, respectively) for low activity. Homozygosity for *UGT1A1**28 has been associated with Gilbert's syndrome, characterized by a mild, chronic unconjugated hyperbilirubinemia in the absence of liver disease [5].

Several case-control studies have examined TA repeat polymorphisms and breast cancer risk, but the findings have varied by race. Guillemette et al. found that *UGT1A1**28 and *UGT1A1**37 alleles were associated with a 1.8-fold increased risk in premenopausal African Americans [4], but they did not find similar relationship in Caucasian women [6]. A study conducted in Russia showed that *UGT1A1**28 allele-containing genotypes were associated with breast cancer risk [7]. Adegoke et al. observed that the *UGT1A1**28 allele was associated with an increased risk of breast cancer among Chinese women younger than 40 years old, but not among women 40 years old or over [8]. Another study conducted in a Chinese population showed that the proportion having the *UGT1A1**28 allele was similar between cases and controls [9].

The frequency of TA repeat polymorphisms varies among ethnic groups [4, 6, 8, 10–12], in particular between African and non-African populations [12, 13] and this variability is suggested to be maintained by natural selection [13]. Furthermore, breast cancer is a heterogeneous disease; for example, we found hormone receptor negative breast cancer was much more common in West Africans (Olopade et al. submitted) than in US populations of African Americans and non-African Americans [14]. The mechanisms through which the *UGT1A1* gene affects breast cancer risk may vary by breast cancer subtype, and the different distribution of subtypes across populations could explain the observed discrepancies regarding *UGT1A1* TA repeats and breast cancer risk. In this study, we performed a comprehensive assessment of the genetic variation in the *UGT1A1* gene and examined these polymorphisms, along with TA repeats, in relation to breast cancer risk in a case-control study. To our knowledge, this is the first study to examine variants other than TA repeats in the entire region of *UGT1A1* gene in relation to breast cancer risk.

Materials and methods

Study sample

Breast cancer cases were identified through the Surgical Oncology and Radiotherapy units of the University College Hospital (UCH), Ibadan, Nigeria. This hospital serves a population of 3 million people and is a referral center for other hospitals in the region. Based on referral patterns, the majority of breast cancer cases diagnosed in the region would probably be seen at the UCH. All consecutive female breast cancer cases aged 18 and above, with a clinical diagnosis of invasive breast cancers between March 1998 and December 2003, were eligible. After obtaining informed consent, trained nurse interviewers recruited patients into the study, administered a structured questionnaire, measured height and weight, and obtained blood samples for genotyping. Patients were recruited and interviewed on the same day they presented at the oncology clinic. If they did not have histological diagnosis before

presentation, biopsy was done and in the rare instances where histology report comes back as a diagnosis other than breast cancer, such person was removed from the study. All consecutive eligible cases were approached. The majority of them agreed to participate in the study, with a refusal rate of 4%.

During the period of case enrollment, a community adjoining the hospital was randomly selected by ballot from the list of all the communities in the hypothetical catchment area of the hospital. This community is thought to be stable, socio-economically diverse and represents the diversity of patients seen at the UCH. A community register was obtained and a meeting was arranged with community elders, during which the purpose of the study was explained to them. After obtaining assent of the elders, meetings were held at local gathering places to explain the purpose of the research to members of the community. Names were then randomly selected from the register and the individuals were invited to visit a clinic set up in the community for the study. Inclusion criteria for the controls were female, aged 18 years or older, absence of any type of cancer at the recruitment, and ability to give informed consent. After explaining the project to potential participants, a breast examination is done to rule out evidence of cancer by a trained nurse who also interviewed the participants; obtained anthropometric measurements and blood samples. Of persons invited, most came to the clinic for screening exams and only 2 individuals were found ineligible. In one case, the patient had suspected carcinoma of the head of pancreas and in the other a breast lump was found and carcinoma could not be ruled out.

The study was approved by the Institutional Review Boards of the University of Chicago and the University of Ibadan.

SNP identification and selection

Polymorphic variants occurring among Africans were identified by resequencing of the *UGT1A1* gene region in 24 Nigerian individuals of Yoruba origin [15]. A total of 66 polymorphisms were identified, including (TA)n repeat polymorphism (position 174990) and an indel polymorphism (A > AAAAGGGAGGGA) in intron 1 (position 179620). Based on these resequencing data, we made an optimal subset of map-based tagging SNPs (tSNPs) for this association study by using the algorithm implemented in LDSelect [16]. The criteria for the selection of tagging SNPs were minor allele frequency > 10% and a threshold of the r^2 value at 0.8. Briefly, the LDSelect program picks up bins composed of tagging SNPs and other SNPs in linkage to the tagging SNPs in each bin based on linkage disequilibrium (LD) between SNPs at a given r^2 value, in this case, 0.8. As a result, we obtained 20 bins with tagging SNP information. The average number of tagging SNPs was 1.4 per bin, and 15 bins had single tagging SNPs which is thought to reflect the short LD pattern in Africans. For technical reason, one SNP failed to be genotyped. In addition to this set of tagging SNPs, we also included two functional SNPs, (TA)n repeat and -3279 T/G polymorphism (position 171764).

Genotyping method

Genotyping was performed using the Beckman-Coulter SNPstream Genotyping System and procedures followed the specifications for the instrument (Fullerton, CA). Primer design for

PCR and Single Base Extension (SBE) was performed using Beckman-Coulter's Autoprimer software. PCR reactions were organized by SNP type. Reaction components included PCR primers at a final concentration of 50 nM, 0.2 U enzyme Hot Master Taq (Eppendorf, Hamburg, Germany), and 2–5 ng of genomic DNA per reaction. Amplifications were performed according to the manufacturer's conditions. All post-amplification steps were performed according to Beckman SNPstream specifications. Briefly, PCR clean-up was accomplished by treating reactions with 3 μl of Exo/SAP (USB, Cleveland, OH) with incubations of 30 min at 37°C followed by 10 min at 100°C. SBE reactions were performed using reagents and protocols specific to the Beckman SNPstream platform. SBE primers were present at a final concentration of 20 nM. Reactions were hybridized to Beckman Array plates and scanned by the SNPstream. Genotyping of SNPs was done by DNAPrint™ Genomics (Sarasota, FL).

In order to genotype the TA repeat in the promoter region, 25 ng of DNA was subjected to PCR amplification and analyzed by Genescan software (Applied Biosystems). One set of primers [GS-F-FAM (forward): 5′-/6-FAM/AGTCACGTGACACAGTCAAAC-3′ and GS-R (reverse, pigtailed): 5′-GTTTCTTTTGCTCCTGCCAGAGGTT-3′] was used to generate 104–110-bp fragments depending on the number of alleles. Reactions were denatured initially at 95°C for 10 min then cycled at 94°C for 30 s, 60°C for 30 s and 72°C for 30 s for 35 cycles. PCR products were diluted 1:10 prior to being run on the ABI 3100 Analyzer (Applied Biosystems).

In order to genotype the indel polymorphism, 10 ng of DNA was subjected to PCR amplification and analyzed by Genescan software (Applied Biosystems). One set of primers [FAM 179620 UGT1A1 FOR(forward): 5′-/56-FAM/CTCAGGGTGTTCTTGCTAC-3′ and 179620 UG T1A1 REV (reverse): 5′-CACTCTCACCAGCCCCAG-3′] was used to generate 154–166-bp fragments depending on the number of alleles. Reactions were denatured initially at 95°C for 10 min then cycled at 94°C for 30 s, 58°C for 30 s and 72°C for 30 s for 35 cycles. PCR products were diluted 1:8 prior to being run on the ABI 3100 Analyzer (Applied Biosystems).

Thirty-two patients had duplicate samples for genotyping and the concordance rate in genotypes in these duplicates was 98%. Most inconsistencies occurred in specimens from one patient whose data were excluded. The average genotyping success rate was 96% and similar between cases and controls.

Statistical analysis

Demographic characteristics and selected risk factors for breast cancer were compared between cases and controls using t-tests for continuous variables and chi-square tests for categorical data. Hardy-Weinberg equilibriums (HWE) for all loci were examined among cases and controls separately using a chi-square test. To examine whether genetic susceptibility for cancer risk is associated with a single locus, we compared case-control differences in genotype frequencies using chi-square tests. Logistic regression models were used to calculate odds ratios (OR) and 95% confidence intervals (CI). Multiple logistic regression models were also fit to adjust for age and ethnicities because they were different significantly between cases and controls. An analysis confined to Yorubas was also

conducted. For TA repeats, we also tested for the trend by coding homozygous high activity genotypes as 0, heterozygous activity genotypes as 1, and homozygous low activity genotypes as 2, based on a previous study [12]. Stratified analysis was conducted by menopausal status and age group (<45 and ≥45 years old), and interactions between *UGT1A1* genotypes and menopausal status or age were tested in logistic regression models.

Haplotype analysis was conducted to capture other unmeasured variants which may affect susceptibility to cancer. Haplotypes were reconstructed using a Bayesian statistical method implemented in PHASE 2.1 [17]. To describe the haplotype diversity, we calculated the effective number of haplotypes, defined as $Ne = 1/\Sigma Pi^2$, where *Pi* is the relative frequency of the *i*th haplotype [16]. The *UGT1A1* gene region was partitioned into haplotype blocks—regions of high LD with low haplotype diversity—using algorithms of Patil et al. [18] and implemented in HaploBlockFinder [19]. Within each block, we compared haplotype frequencies between cases and controls, and calculated the odds ratio for carriers of each haplotype versus non-carriers using logistic regression models. To account for the uncertainty in haplotype assignment and imputation of missing genotypes, 100 replicate data sets were generated from the posterior distributions in PHASE package and analyzed using logistic regressions in SAS 9.1 (SAS Institute, Cary, NC). The log ORs were combined using Rubin's formula [20]. In this multiple imputation procedure, 100 was chosen as the number of imputations to keep the efficiency of parameter estimation greater than 99.8% for the fraction of missing information up to 0.15.

Results

There were 738 participants genotyped for UGT1A1: 512 breast cancer cases and 226 controls. On average, cases were 5 years older than controls (Table 1). The majority of study participants were Yoruba, and other ethnicities include Hausa and Ibo. Many cases were diagnosed at an advanced stage of cancer. Cases were significantly more likely to report a family history of breast cancer. There were more post-menopausal women in cases than in controls. Cases had larger waist-to-hip ratio than controls and were more likely to drink alcohol than controls. The education level and socioeconomic status were slightly higher in cases than in controls. Cases and controls were comparable in the location of residence, age at menarche, hormone contraceptive use, and body mass index.

Table 2 displays the description of the TA repeat, the deletion/insertion polymorphism, and 19 single nucleotide polymorphisms, including chromosome location, nomenclature of the Human Genome Variation Society, dbSNP ID for known SNPs, and variants. In controls, all markers followed HWE except for markers 13 and 18 ($P = 0.02$ and 0.03, respectively). In cases, only TA repeats and markers 8 and 17 departed from HWE, but the tests for HWE were no longer significant after stratification by ethnicity. These departures from HWE may be due to chance alone. The allele frequencies of TA repeats were similar between cases and controls. The A allele frequency of marker 2 was slightly higher in controls than in cases ($P = 0.03$). No significant differences were observed for the remaining 19 markers.

Table 3 presents the genotype frequency of the TA repeats, which was statistically significantly different between cases and controls (Fisher's exact test, $P = 0.049$). Cases

were more likely to have the 6/6 genotype, while controls were more likely to have the 5/6 and 6/7 genotypes. As a previous study showed that 5 and 6 repeats (allele *36 and *1, respectively) were associated with high *UGT1A1* transcriptional activity, and 7 and 8 repeats (alleles *28 and *37, respectively) represented low activity alleles [12], we grouped subjects into 3 categories: genotypes with low activity alleles only (7/7, 7/8, 8/8), genotypes with high activity alleles only (5/5, 5/6, 6/6), and genotypes with heterozygous activity alleles (5/7, 5/8, 6/7, 6/8) (see Table 3). Overall, there was no significant association between the TA repeat genotypes and risk of breast cancer ($P = 0.14$ for trend). However, we observed a statistically significantly differential effect of TA repeat genotype by menopausal status ($P = 0.01$ for interaction): in premenopausal women, low activity genotypes were associated with decreased risk of breast cancer ($P = 0.009$ for trend); while low activity genotypes was not associated with breast cancer risk in postmenopausal women ($P = 0.24$ for trend). After adjusting for age and ethnicity, the interaction remained statistically significant ($P = 0.03$ for interaction). The adjusted odds ratio for 7 or 8 allele-containing genotypes was 0.53 (95% CI 0.28–1.00) compared with 5 or 6 allele-containing genotypes in premenopausal women.

Similarly, we observed a significantly differential effect of the TA repeat genotype according to age group ($P = 0.002$ for interaction): low activity genotypes were associated with decreased risk of breast cancer in women <45 years old ($P = 0.002$ for trend) but not in women ≥45 years old ($P = 0.15$ for trend). After adjusting for age and ethnicity, this interaction was still significant ($P = 0.007$ for interaction). The adjusted odds ratio for 7 or 8 allele-containing genotypes was 0.47 (95% CI 0.24–0.94) compared with 5 or 6 allele-containing genotypes in women younger than 45. Analysis restricted to Yoruba only and the multivariate analysis adjusting for potential confounders listed in Table 1 gave similar results (data not shown).

For the whole *UGT1A1* gene region, haplotypes were highly diverse. There were 5 common haplotypes with a frequency greater than 5% in both cases and controls. But they only represented 36% and 37% of the chromosomes in cases and controls, respectively. As an index of haplotype diversity, the effective number of haplotype was 25.5 for cases and 26.4 for controls. To capture the haplotype diversity and efficiently compare haplotype frequencies between cases and controls, four haplotype blocks were identified (Fig. 1): block 1 (markers 1–5, 3.2 kb) spanned the upstream region of *UGT1A1*, also called the pheno-barbital-responsive enhancer module; block 2 (markers 6–13, 6.0 kb) ranged from intron 1 to intron 3; block 3 (markers 14–15, 0.4 kb) included two markers in introns 2 and 3; block 4 (markers 16–21, 6.9 kb) spanned the 3′ UTR of exon 5. These partitions were also consistent with recombination hotspots indicated by recombination probabilities calculated during haplotype reconstruction.

Table 4 shows the haplotype frequency by blocks in cases and controls. Percentages rather than absolute numbers (not integers) are presented because they are averaged from 100 imputations. No global tests for haplotype effects were significant for any block. However, carriers of the CTTC + CA haplotype in block 2 had a significantly higher risk of breast cancer than non-carriers of this haplotype (OR 2.14, 95% CI 1.05–4.39; $P = 0.04$). After adjusting for age and ethnicity, this association remained significant ($P = 0.04$).

Table 5 presents the frequency of haplotype pairs (diplotype) for CTTCC + CA. When stratified by menopausal status, the percentage of homozygosity or heterozygosity for CTTCC + CA was higher in cases than in controls among both premenopausal and postmenopausal women. The stratified analysis by age revealed that the association existed in both women ≥45 years old and women <45 years old. Although this haplotype was linked to the 6 repeat of TA (*UGT1A1**1), its association with breast cancer risk was still significant after adjusting for TA repeats because this haplotype existed in only 10.5% of all 6 allele-containing chromosomes (data not shown). On the other hand, after excluding subjects with the CTTTCC + CA haplotype, the interactions of TA repeats with menopausal status or age remained significant (data not shown).

Discussion

Previous studies have observed that the wild-type allele (*UGT1A1**1) was the most common allele among Caucasian, African American and Asian populations [4, 6, 8, 10–12]. We found that the variant allele *UGT1A1**28 was as common as *UGT1A1**1 allele, with about 40% frequency for each allele among indigenous Africans. *UGT1A1**36 and *UGT1A1**37 have been observed exclusively among African Americans [4, 12, 21], and we found that these two variant alleles were also present among indigenous Africans with slightly higher frequencies in the current study, suggesting that *UGT1A1**36 and *UGT1A1**37 can be used as markers for African ancestry.

All previous studies [4, 6, 8, 9] except one [7] showed that there was no overall difference between breast cancer cases and controls in the distribution of TA repeat genotypes. We found the distribution of TA repeat genotypes was significantly different between cases and controls, but the patterns were not systematic in the overall analysis. In contrast to the two previous studies conducted in African Americans [4] and in Chinese [8], which found that 7 or 8 allele-containing genotypes were associated with elevated risk of breast cancer in premenopausal women [4] or women younger than 40 years [8], we found that these low activity 7 or 8 allele-containing genotypes were associated with decreased risk of breast cancer among premenopausal women or those younger than 45 years old.

There are several possible reasons for this discrepancy. First, other genes responsible for estrogen metabolism and environmental factors related to estrogen levels may vary across different racial populations. Although the plasma estradiol level was found to be elevated in carriers of the *UGT1A1**28 allele, the elevation was confined to homozygous carriers [22] or postmenopausal women with BMI > 27 kg/m^2 [6]. Second, UGT1A1 may also affect breast cancer development through the bilirubin pathway, as it is the primary isoenzyme responsible for bilirubin glucoronidation [1] and bilirubin is an antioxidant [23]. Researchers have hypothesized that low-activity UGT1A1 TA repeat alleles are associated with high bilirubin level, which may result in decreased cancer susceptibility due to increased antioxidant load [24]. A case-control study showed that bilirubin levels were lower in breast cancer cases than in healthy controls [25]. Therefore, bilirubin, as an antioxidant, may counteract the detrimental effect of elevated estradiol levels associated with low-activity *UGT1A1* alleles in premenopausal women. A study that examined the relationship between *UGT1A1* TA repeat polymorphisms and breast density, a strong predictor of breast cancer

risk, showed that homozygosity for the *28 allele was associated with lower breast density in premenopausal women and higher breast density in postmenopausal women [26]. This is in line with our study's findings. Third, the *UGT1A1* promoter variant may not be the true causal variant for breast cancer risk, but may be in linkage disequilibrium with the causal variant. Different haplotype structures in different populations could explain some of the inconsistencies across studies. Fourth, it is likely that the majority of our cases are hormone receptor negative as we found that West African breast cancer patients are more likely to be diagnosed with hormone receptor negative tumors (Olopade et al. submitted), suggesting that the estrogen metabolism pathway may be less important in this population.

To our knowledge, this is the first study that examined the *UGT1A1* gene polymorphisms other than TA repeats in relation to breast cancer risk. In the locus-by-locus analysis, none of the 19 SNPs and the insertion/deletion polymorphism were significantly different between cases and controls. However, in the haplotype analysis, we found that a specific haplotype in block 2 was significantly associated with a 2-fold elevated risk for breast cancer. Although the marker map was relatively dense with about 1 marker per 1 kb, the results of the haplotype analysis suggest that there is still an unmeasured variant that confers breast cancer risk.

Several limitations need to be considered when interpreting the results. First, controls were younger and more likely to be Yorubas than cases, but age- and ethnicity-adjusted analyses and the analyses restricted to Yorubas gave similar results, suggesting that there was no substantial population stratification. Second, there is no gold standard definition of a haplotype block, and several algorithms for haplotype block partitioning exist. Unfortunately, the block boundaries from different algorithms do not always match, and the haplotype diversity based method, which we used, tends to infer smaller block than other methods [27]. This ambiguity in haplotype block partitioning adds another barrier to the cross-population comparison. Finally, as in many genetic epidemiologic studies, the multiple comparison issue exists in this study. The analyses on TA repeats were confirmative in nature, and the tests for interaction and subgroup analyses were planned prospectively. However the analyses on other polymorphic markers were exploratory and thus the haplotype in block 2 was no longer statistically significant after the Bonferroni adjustment.

In summary, we observed that low-activity 7 or 8 alleles in the *UGT1A1* gene promoter were associated with decreased risk for breast cancer in indigenous African women who were premenopausal or under 45 years old, but not in women who were postmenopausal or 45 years and above. This finding suggests that the role of *UGT1A1* in breast cancer development may vary by population, presumably due to different environmental and genetic modifier effects. Both estrogen and antioxidant pathways should be explored. In addition, the mechanism by which the identified high-risk haplotype modulates cancer susceptibility warrants further investigation.

References

1. Tukey RH, Strassburg CP. Human UDP-glucuronosyltransferases: metabolism, expression, and disease. Annu Rev Pharmacol Toxicol. 2000; 40:581–616. [PubMed: 10836148]

2. Mackenzie PI, Walter Bock K, Burchell B, et al. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. Pharmacogenet Genomics. 2005; 15:677–685. [PubMed: 16141793]
3. Senafi SB, Clarke DJ, Burchell B. Investigation of the substrate specificity of a cloned expressed human bilirubin UDP-glucuronosyltransferase: UDP-sugar specificity and involvement in steroid and xenobiotic glucuronidation. Biochem J. 1994; 303(Pt 1):233–240. [PubMed: 7945246]
4. Guillemette C, Millikan RC, Newman B, Housman DE. Genetic polymorphisms in uridine diphospho-glucuronosyltransferase 1A1 and association with breast cancer among African Americans. Cancer Res. 2000; 60:950–956. [PubMed: 10706110]
5. Bosma PJ, Chowdhury JR, Bakker C, et al. The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. N Engl J Med. 1995; 333:1171–1175. [PubMed: 7565971]
6. Guillemette C, De Vivo I, Hankinson SE, et al. Association of genetic polymorphisms in UGT1A1 with breast cancer and plasma hormone levels. Cancer Epidemiol Biomarkers Prev. 2001; 10:711–714. [PubMed: 11401924]
7. Shatalova EG, Loginov VI, Braga EA, et al. Association of polymorphisms in SULT1A1 and UGT1A1 Genes with breast cancer risk and phenotypes in Russian women. Mol Biol (Mosk). 2006; 40:263–270. [PubMed: 16637266]
8. Adegoke OJ, Shu XO, Gao YT, et al. Genetic polymorphisms in uridine diphospho-glucuronosyltransferase 1A1 (UGT1A1) and risk of breast cancer. Breast Cancer Res Treat. 2004; 85:239–245. [PubMed: 15111762]
9. Cheng TC, Chen ST, Huang CS, et al. Breast cancer risk associated with genotype polymorphism of the catechol estrogen-metabolizing genes: a multigenic study on cancer susceptibility. Int J Cancer. 2005; 113:345–353. [PubMed: 15455371]
10. Balram C, Sabapathy K, Fei G, Khoo KS, Lee EJ. Genetic polymorphisms of UDP-glucuronosyltransferase in Asians: UGT1A1*28 is a common allele in Indians. Pharmacogenetics. 2002; 12:81–83. [PubMed: 11773869]
11. Innocenti F, Grimsley C, Das S, et al. Haplotype structure of the UDP-glucuronosyltransferase 1A1 promoter in different ethnic groups. Pharmacogenetics. 2002; 12:725–733. [PubMed: 12464801]
12. Beutler E, Gelbart T, Demina A. Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? Proc Natl Acad Sci USA. 1998; 95:8170–8174. [PubMed: 9653159]
13. Hall D, Ybazeta G, Destro-Bisol G, Petzl-Erler ML, Di Rienzo A. Variability at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates. Pharmacogenetics. 1999; 9:591–599. [PubMed: 10591539]
14. 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]
15. Maitland ML, Grimsley C, Kuttab-Boulos H, et al. Comparative genomics analysis of human sequence variation in the UGT1A gene cluster. Pharmacogenomics J. 2006; 6:52–62. [PubMed: 16314881]
16. Carlson CS, Eberle MA, Rieder MJ, et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet. 2004; 74:106–120. [PubMed: 14681826]
17. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet. 2003; 73:1162–1169. [PubMed: 14574645]
18. Patil N, Berno AJ, Hinds DA, et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science. 2001; 294:1719–1723. [PubMed: 11721056]
19. Zhang K, Jin L. HaploBlockFinder: haplotype block analyses. Bioinformatics. 2003; 19:1300–1301. [PubMed: 12835279]
20. Rubin, DB. Multiple imputation for nonresponse in surveys. Wiley; New York: 1987.

21. Hong AL, Huo D, Kim HJ, et al. UDP-glucuronosyltransferase 1A1 gene polymorphisms and total bilirubin levels in an ethnically diverse cohort of women. Drug Metab Dispos. 2007; 35:1254–1261. [PubMed: 17478602]
22. Sparks R, Ulrich CM, Bigler J, et al. UDP-glucuronosyltransferase and sulfotransferase polymorphisms, sex hormone concentrations, and tumor receptor status in breast cancer patients. Breast Cancer Res. 2004; 6:R488–R498. [PubMed: 15318931]
23. Stocker R, Yamamoto Y, McDonagh AF, Glazer AN, Ames BN. Bilirubin is an antioxidant of possible physiological importance. Science. 1987; 235:1043–1046. [PubMed: 3029864]
24. Grant DJ, Bell DA. Bilirubin UDP-glucuronosyltransferase 1A1 gene polymorphisms: susceptibility to oxidative damage and cancer? Mol Carcinog. 2000; 29:198–204. [PubMed: 11170257]
25. Ching S, Ingram D, Hahnel R, Beilby J, Rossi E. Serum levels of micronutrients, antioxidants and total antioxidant status predict risk of breast cancer in a case control study. J Nutr. 2002; 132:303–306. [PubMed: 11823595]
26. Haiman CA, Hankinson SE, De Vivo I, et al. Polymorphisms in steroid hormone pathway genes and mammographic density. Breast Cancer Res Treat. 2003; 77:27–36. [PubMed: 12602902]
27. Indap AR, Marth GT, Struble CA, Tonellato P, Olivier M. Analysis of concordance of different haplotype block partitioning algorithms. BMC Bioinformatics. 2005; 6:303. [PubMed: 16356172]

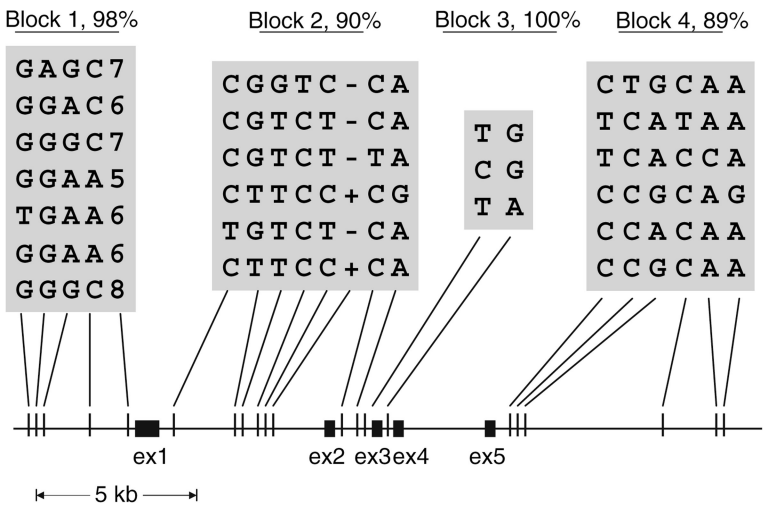

Fig. 1. Haplotype blocks in *UGT1A1* gene region. Most frequent haplotypes in each block and percentage of chromosome coverage are shown, with 21 markers mapped to *UGT1A1* gene region

Table 1

Selected characteristics of women genotyped for *UGT1A1* in Nigerian case control study of breast cancer

Characteristic	Cases (n = 512)	Controls (n = 226)	*P*-value
Age in years, mean ± SD	46.2 ±11.0	41.7 ± 14.1	<0.001
Ethnicity, n (%)			<0.001
Yoruba	388 (75.8)	214 (94.7)	
Others	124 (24.2)	12 (5.3)	
Education, n (%)			0.04
No formal	131 (25.6)	57 (25.2)	
Elementary	117 (22.9)	52 (23.0)	
Secondary	62 (12.1)	66 (29.2)	
Vocational	100 (19.5)	30 (13.3)	
College or above	102 (19.9)	21 (9.3)	
Location of residence, n (%)			0.75
Urban	420 (82.0)	188 (83.2)	
Rural	92 (18.0)	38 (16.8)	
Socioeconomic status, n (%)			0.02
Low	258 (50.4)	138 (61.1)	
Middle	245 (47.9)	87 (38.5)	
High	9 (1.8)	1 (0.4)	
Stage			
I	3 (0.7)		
II	77 (19.0)		
III	135 (33.3)		
IV	191 (47.0)		
Family history of breast cancer, n (%)	50 (9.8)	11 (4.9)	0.03
Menopausal status, n (%)			0.003
Pre-menopausal	288 (56.2)	155 (68.6)	
Post-menopausal, natural	213 (41.6)	70 (31.0)	
Post-menopausal, artificial	11 (2.1)	1 (0.4)	
Age at natural menopause, mean ± SD	47.7 ± 6.1	50.0 ± 6.1	0.01
Age at menarche, mean ± SD	15.3 ± 2.1	15.6 ± 2.1	0.07
Hormone contraceptives, n (%)	114 (22.3)	48 (21.2)	0.77
BMI in kg/m^2, mean ± SD	25.2 ± 5.4	24.7 ± 5.4	0.23
WHR, mean ± SD	0.833 ± 0.088	0.815 ± 0.078	0.009
Alcohol drink, n (%)	112 (21.9)	35 (15.5)	0.046

Note: *SD* standard deviation, *BMI* body mass index, *WHR* waist-to-hip ratio

Table 2

Polymorphisms in *UGT1A1* gene and allele frequency in cases and controls

Marker	Location	dbSNP ID	Position[a]	cDNA position[b]	Variant	MAF of cases[c]	MAF of controls[c]
1	Upstream	rs4124874	171764	c.–3279	G > T	0.093	0.096
2	Upstream	rs10929302	171887	c.–3156	G > A	0.294	0.356
3	Upstream		172092	c.–2951	A > G	0.455	0.505
4	Upstream	rs3755319	173691	c.–1352	C > A	0.278	0.263
5	Promoter		174990	c.–53	*36, $(TA)_5$	0.107	0.104
					*1, $(TA)_6$	0.443	0.394
					*28, $(TA)_7$	0.406	0.459
					*37, $(TA)_8$	0.054	0.043
6	Intron 1		176248	c.864+342	C > T	0.104	0.106
7	Intron 1	rs3771342	178771	c.864+2865	G > T	0.169	0.135
8	Intron 1	rs4148324	178830	c.864+2924	T > G	0.446	0.482
9	Intron 1	rs4148325	179417	c.865–2371	C > T	0.445	0.502
10	Intron 1	rs4148326	179570	c.865–2218	C > T	0.379	0.375
11	Intron 1		179620	c.865–2168ins	Del > Ins	0.165	0.135
12	Intron 2		181937	c.996+18	C > T	0.159	0.142
13	Intron 2	rs1018124	182226	c.996+307	A > G	0.144	0.116
14	Intron 2		182521	c.997–82	T > C	0.415	0.399
15	Intron 3	rs10445705	182895	c.1304–79	G > A	0.090	0.085
16	Exon 5 UTR	rs10929303	187524	c.*211	C > T	0.462	0.491
17	Downstream	rs4148329	188170	c.*857	C > T	0.251	0.224
18	Downstream	rs6717546	188227	c.*914	A > G	0.440	0.432
19	Downstream	rs1500476	192593	c.*5280	C > T	0.285	0.274
20	Downstream	rs6431631	194144	c.*6831	A > C	0.161	0.173
21	Downstream	rs17863803	194435	c.*7122	A > G	0.129	0.111

[a] Physical positions are based on GenBank reference sequence AF297093.1

[b] HGVS nomenclature of sequence variations, http://www.hgvs.org/mutnomen/

[c] Minor allele frequencies are presented except for TA repeat

Note: *MAP* minor allele frequency, *UTR* untranslated region

Table 3
Genotype of *UGT1A1* TA repeats and risk of breast cancer

Genotype	Cases N (%)	Controls N (%)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)[a]
All subjects				
5/5	10 (2.0)	2 (0.9)		
5/6	38 (7.6)	26 (11.7)		
5/7	36 (7.2)	15 (6.8)		
5/8	13 (2.7)	1 (0.5)		
6/6	100 (19.9)	28 (12.6)		
6/7	177 (35.3)	88 (39.6)		
6/8	20 (4.0)	5 (2.3)		
7/7	90 (17.9)	45 (20.3)		
7/8	15 (3.0)	11 (5.0)		
8/8	3 (0.6)	1 (0.5)		
5/5, 5/6, 6/6	148 (29.5)	56 (25.2)	1.0 (reference)	1.0 (reference)
5/7, 5/8, 6/7, 6/8	246 (49.0)	109 (49.1)	0.85 (0.58–1.25)	0.88 (0.59–1.31)
7/7, 7/8, 8/8	108 (21.5)	57 (25.7)	0.72 (0.46–1.12)	0.75 (0.47–1.19)
P for trend test			0.14	0.22
Premenopausal women				
5/5, 5/6, 6/6	84 (29.8)	30 (19.7)	1.0 (reference)	1.0 (reference)
5/7, 5/8, 6/7, 6/8	139 (49.3)	77 (50.7)	0.64 (0.39–1.06)	0.68 (0.39–1.17)
7/7, 7/8, 8/8	59 (20.9)	45 (29.6)	0.47 (0.26–0.83)	0.53 (0.28–1.00)
P for trend test			0.009	0.05
Postmenopausal women[b]				
5/5, 5/6, 6/6	62 (29.7)	25 (36.2)	1.0 (reference)	1.0 (reference)
5/7, 5/8, 6/7, 6/8	100 (47.8)	32 (46.4)	1.26 (0.68–2.32)	1.34 (0.71–2.51)
7/7, 7/8, 8/8	47 (22.5)	12 (17.4)	1.58 (0.72–3.48)	1.40 (0.62–3.17)
P for trend test			0.24	0.36
Women < 45 years				
5/5, 5/6, 6/6	75 (31.9)	25 (19.1)	1.0 (reference)	1.0 (reference)
5/7, 5/8, 6/7, 6/8	111 (47.2)	64 (48.9)	0.58 (0.33–1.00)	0.67 (0.36–1.22)
7/7, 7/8, 8/8	49 (20.9)	42 (32.1)	0.39 (0.21–0.71)	0.47 (0.24–0.94)
P for trend test			0.002	0.032
Women ≥ 45 years				
5/5, 5/6, 6/6	73 (27.3)	31 (34.1)	1.0 (reference)	1.0 (reference)
5/7, 5/8, 6/7, 6/8	135 (50.6)	45 (49.5)	1.27 (0.74–2.18)	1.29 (0.75–2.24)
7/7, 7/8, 8/8	59 (22.1)	15 (16.5)	1.67 (0.82–3.38)	1.51 (0.73–3.10)
P for trend test			0.15	0.24

[a] Adjusted for age and ethnicity in logistic regression models

[b] Excluded women with artificial menopause

Note: *OR* odds ratio, *CI* confidence interval

Table 4
Association between haplotypes in blocks 1–4 in *UGT1A1* gene and breast cancer risk

	Haplotype percentage		Unadjusted OR (95% CI)[a]	Adjusted OR (95% CI)[a,b]
	Cases (n = 512)	Controls (n = 226)		
Block 1, loci 1–5				
GAGC7	28.6	33.6	**0.72 (0.52–1.00)***	0.79 (0.57–1.11)
GGAC6	26.1	22.6	1.20 (0.87–1.65)	1.22 (0.87–1.72)
GGGC7	11.1	12.1	0.93 (0.63–1.38)	0.98 (0.65–1.47)
GGAA5	10.3	10.2	0.95 (0.63–2.19)	0.88 (0.58–1.34)
TGAA6	8.9	9.4	0.88 (0.58–1.33)	0.93 (0.61–1.44)
GGAA6	8.3	6.1	1.39 (0.86–2.26)	1.33 (0.80–2.20)
GGGC8	5.3	3.8	1.42 (0.78–2.56)	1.27 (0.68–2.38)
Block 2, loci 6–13				
CGGTC – CA	41.7	44.3	0.82 (0.59–1.15)	0.85 (0.60–1.20)
CGTCT – CA	13.2	13.3	0.96 (0.67–1.39)	1.01 (0.69–1.48)
CGTCT – TA	10.5	10.2	0.99 (0.66–1.47)	0.94 (0.62–1.43)
CTTCC + CG	10.0	9.6	1.10 (0.72–1.68)	1.09 (0.70–1.70)
TGTCT – CA	8.9	9.8	0.88 (0.58–1.34)	0.94 (0.60–1.45)
CTTCC + CA	**5.4**	**2.6**	**2.14 (1.05–4.39)***	**2.14 (1.02–4.46)***
CGTTC – CA	2.4	3.4	0.67 (0.33–1.34)	0.79 (0.39–1.61)
CGGCT – TA	2.3	2.3	1.03 (0.46–2.29)	0.91 (0.39–2.14)
Block 3, loci 14–15				
TG	49.0	52.2	0.77 (0.53–1.13)	0.82 (0.55–1.23)
CG	41.8	39.5	1.12 (0.80–1.57)	1.08 (0.76–1.54)
TA	9.2	8.3	1.14 (0.74–1.78)	1.31 (0.82–2.08)
Block 4, loci 16–21				
CTGCAA	24.9	22.3	1.09 (0.78–1.52)	1.14 (0.80–1.62)
TCATAA	23.9	23.1	1.08 (0.77–1.50)	1.07 (0.75–1.51)
TCACCA	16.1	18.1	0.88 (0.61–1.26)	0.90 (0.62–1.32)
CCGCAG	13.0	10.7	1.18 (0.79–1.75)	1.23 (0.82–1.87)
CCACAA	7.7	7.1	1.09 (0.68–1.75)	1.18 (0.72–1.93)
CCGCAA	3.9	6.6	0.63 (0.36–1.09)	0.62 (0.34–1.12)
TCACAA	4.0	5.0	0.77 (0.42–1.42)	0.73 (0.38–1.41)
CCATAA	3.4	3.4	0.93 (0.47–1.81)	0.93 (0.46–1.87)

[a] Reference groups for odds ratio are non-carriers of each haplotype

[b] Adjusted for age and ethnicity in logistic regression models

* $P < 0.05$

Note: *OR* odds ratio, *CI* confidence interval

Table 5
Association of diplotype in block 2 in *UGT1A1* gene and breast cancer risk

No. of haplotype CTTCC + CA	Cases (n = 512) %	Controls (n = 226) %	Unadjusted OR (95% CI)	Adjusted OR (95% CI)[a]
All subjects				
0	89.4	94.7	1.0 (reference)	1.0 (reference)
1 or 2	10.6	5.3	2.14 (1.05–4.39)	2.14 (1.02–4.46)
1	10.3	5.3		
2	0.3	0.0		
Premenopausal women				
0	88.3	94.8	1.0 (reference)	1.0 (reference)
1 or 2	11.7	5.2	2.44 (1.02–5.83)	2.11 (0.82–5.45)
Postmenopausal women[b]				
0	90.4	94.5	1.0 (reference)	1.0 (reference)
1 or 2	9.6	5.5	1.85 (0.53–6.45)	1.99 (0.56–7.04)
Women < 45 years				
0	88.0	95.2	1.0 (reference)	1.0 (reference)
1 or 2	12.0	4.8	2.72 (1.05–7.04)	2.24 (0.77–6.55)
Women ≥ 45 years				
0	90.7	94.1	1.0 (reference)	1.0 (reference)
1 or 2	9.3	5.9	1.66 (0.58–4.81)	1.83 (0.63–5.32)

[a] Adjusted for age and ethnicity in logistic regression models

[b] Excluded women with artificial menopause

Note: *OR* odds ratio, *CI* confidence interval