

DMD #14183

UDP-Glucuronosyltransferase 1A1 Gene Polymorphisms and Total Bilirubin Levels in an Ethnically Diverse Cohort of Women

Andrew L. Hong¹, Dezheng Huo¹, Hee-Jin Kim¹, Qun Niu, Donna L. Frankenthal, Shelly A. Cummings, Esther M. John, Dee W. West, Alice S. Whittemore, Soma Das, Olufunmilayo I. Olopade

Department of Medicine, Section of Hematology/Oncology, University of Chicago, Chicago, IL.
(ALH, HJK, QN, SAC, OIO)

Department of Health Studies, University of Chicago, Chicago, IL. (DH)

Department of Human Genetics, University of Chicago, Chicago, IL (DLF, SD, OIO)

Northern California Cancer Center, Fremont, CA (EMJ, DWW)

Stanford University School of Medicine, Stanford, CA (DWW, ASW)

DMD #14183

Running title: *UGT1A1* SNPs and Total Bilirubin in Women

Correspondence: Olufunmilayo I. Olopade, MD, Department of Medicine, Section of Hematology/Oncology, University of Chicago, 5841 S. Maryland Ave, MC 2115, Chicago, IL 60637. (email: folopade@medicine.bsd.uchicago.edu)

Text pages: 25

Tables: 5

Figures: 3

References: 39

Abstract: 225 words

Introduction: 716 words

Discussion: 1500 words

Nonstandard abbreviations list:

CAAT-enhancer binding protein (C/EBPalpha)

Hardy-Weinberg equilibrium (HWE)

Linkage disequilibrium (LD)

DMD #14183

Minor Allele Frequency (MAF)

One-way analysis of variance (ANOVA)

Phenobarbital-responsive enhancer module (PBREM)

Single Base Extension (SBE)

Single nucleotide polymorphisms (SNPs)

Tagging SNPs (tSNPs)

UDP-glucuronosyl transferase (UGT)

DMD #14183

Abstract

Objectives: To investigate variations in *UGT1A1* polymorphisms and haplotypes among African American and Caucasian women and to assess whether variants other than *UGT1A1**28 are associated with total serum bilirubin levels. **Methods:** The (TA)_n repeats and 14 single nucleotide polymorphisms (SNPs) in the *UGT1A1* gene were genotyped in 335 African Americans and 181 Caucasians. Total serum bilirubin levels were available in a subset of 125 women. **Results:** Allele frequencies of all SNPs and (TA)_n repeats were significantly different between African Americans and Caucasians. In Caucasians, three common haplotypes accounted for 71.8% of chromosomes while five common haplotypes accounted for only 46.6% of chromosomes in African Americans. Mean total serum bilirubin levels were significantly lower ($p=0.005$) in African Americans (0.36 mg/dl) than in Caucasians (0.44 mg/dl). The (TA)_n repeats explained a significant amount of variation in total bilirubin levels ($R^2=0.27$, $p<0.0001$) while other SNPs were less correlative. **Conclusions:** Significant variations in *UGT1A1* haplotype structure exist between African Americans and Caucasians in this relatively large cohort of women. *UGT1A1*'s correlation with total bilirubin levels was mainly due to (TA)_n repeats in Caucasians but a clear correlation was not observed in African Americans due to the high diversity of haplotypes and a small sample size. These data have implications in design of epidemiologic studies of cancer susceptibility and pharmacogenetic studies for adverse drug reactions in populations of African ancestry.

DMD #14183

Introduction

The *UDP-glucuronosyl transferase (UGT) 1A1* gene has been implicated in a number of processes, including conjugation of bilirubin from hemoglobin and hemoprotein turnover, detoxification of potential carcinogens, phase II drug metabolism, and estradiol metabolism (Senafi et al., 1994; Bosma et al., 1995; Malfatti et al., 2005). The nine *UGT1A* isoforms are expressed to varying degrees in the liver (Strassburg et al., 1997) and mammary tissue (Senafi et al., 1994; Chouinard et al., 2006). The number of TA repeats in the TATA promoter region of *UGT1A1* has been shown to be inversely associated with the transcriptional activity of *UGT1A1*, with 5 and 6 repeats (allele *36 and *1, respectively) associated with high *UGT1A1* activity, and 7 and 8 repeats (alleles *28 and *37, respectively) associated with low *UGT1A1* activity (Beutler et al., 1998). The genotype, (TA)₇/(TA)₇, has been associated with Gilbert's syndrome, which presents with mild hyperbilirubinemia (Bosma et al., 1995).

The unifying premise of *integrative epidemiology* suggested by Spitz et al. is that the same genes that are implicated in cancer risk may also be involved in a person's propensity to carcinogenic exposure and/or to modulation of therapeutic outcome (Spitz et al., 2005). Therefore, constructing genetic profiles that could be used to individualize therapy may also increase our understanding of cancer risk genes and may be applied to cancer development and prediction of outcome. To this end, differences in TA repeats of *UGT1A1* have been shown to be responsible for the toxic effects of irinotecan, an anti-cancer drug (Iyer et al., 2002), and have also been linked with cancer susceptibility (Guillemette et al., 2000; Adegoke et al., 2004). Of interest are studies suggesting an association of *UGT1A1* TA repeat polymorphisms with breast cancer among African Americans but not Caucasians (Guillemette et al., 2000; Guillemette et al., 2001).

DMD #14183

UGT1A1 is the main isoform that glucuronidates bilirubin (Tukey and Strassburg, 2000) and variants in the TATA promoter and phenobarbital-responsive enhancer module (PBREM) region have been studied extensively, but variation in serum bilirubin levels is not explained solely by TA repeats. Few studies have explored the polymorphisms throughout *UGT1A1* gene and assigned functions to these polymorphisms (Sai et al., 2004; Kaniwa et al., 2005). In addition, few studies have examined variants of *UGT1A1* related to total serum bilirubin within the normal range (<1.0 mg/dL) (Bosma et al., 1995; Sai et al., 2004; Lin et al., 2006). Nevertheless, it is important to evaluate these to understand if variants causing hyperbilirubinemia are similar to or different from those involving variation of total bilirubin levels within the normal range. Identifying common alleles of *UGT1A1* related to bilirubin has important implications in both cancer treatment and prevention. A recent clinical trial has shown that total bilirubin before treatment was a positive predictor of grade 4 neutropenia in cancer patients receiving irinotecan (Innocenti et al., 2004). Additionally, recent works have shown that the risk of cardiovascular disease is lower in those with higher bilirubin levels (Lin et al., 2006). It also has been hypothesized that bilirubin itself may be a potent antioxidant and its slight increase within the normal range may be beneficial in reducing risk of cancer (Grant and Bell, 2000).

Furthermore, population-based studies have shown that serum bilirubin levels vary with gender, race and smoking status and are 10-20% lower in African American than in Caucasians (Zucker et al., 2004). While there are well established differences in *UGT1A1* TA repeat allele frequencies between Caucasians and African Americans, with (TA)₅ and (TA)₈ being absent in Caucasians (Beutler et al., 1998; Guillemette et al., 2000; Guillemette et al., 2001), it remains unclear whether there are additional differences in other *UGT1A1* single nucleotide

DMD #14183

polymorphisms (SNPs) between the two racial groups, which may explain some of the observed racial difference in serum bilirubin levels.

In this study, we aim to: 1) compare the variations in $(TA)_n$ and SNPs from the PBREM region to the 3'-conserved region of *UGT1A1* between African American and Caucasian women; 2) describe the linkage disequilibrium (LD) and construct haplotypes for these two populations; and 3) examine the relationship between *UGT1A1* polymorphisms and total bilirubin among individuals with total bilirubin levels within the normal range. To our knowledge, this is the first study to examine the entire region of *UGT1A1* gene in relation to total bilirubin level in Caucasian and African American women.

Materials & Methods

Study Sample

The study sample included individuals from two sources: Self-reported African American and Caucasian women who attended the Cancer Risk Clinic at the University of Chicago between 1992 and 2003, and self-reported African American women from the San Francisco Bay area who enrolled in the Northern California Family Registry for Breast Cancer between 1995 and 2000 (John et al., 2004). DNA samples were available for 227 individuals (46 African Americans and 181 Caucasians) aged 23-81 years who attended the Cancer Risk Clinic for a comprehensive cancer risk assessment; of these, 59% had previous diagnosis of breast cancer and 65% had a family history of breast cancer. DNA samples were also available for 244 newly diagnosed African American breast cancer patients aged 20-64 years identified through the Greater San Francisco Bay area cancer registry, and 45 African American women without breast cancer identified through random-digit dialing. Characteristics of the study population are shown

DMD #14183

in Table 1. All participants provided written consent and the study was approved by the Institutional Review Boards of the University of Chicago and the Northern California Cancer Center.

SNP Selection

Novel SNP identification was based on resequencing of the *UGT1A* gene cluster (Maitland et al., 2006). Briefly, optimal sets of mapped based tagging SNPs (tSNPs) were generated from LDSelect (Carlson et al., 2004) for African American and Caucasian populations (unpublished data). As expected, the number of tSNPs for Caucasians (n=10) was smaller than that for African Americans (n=18) since the number of genetic variations is generally smaller in the former. The criteria for selecting these tSNPs were setting the SNPs with a relative MAF >10% and a r^2 threshold of 0.8. An additional 6 SNPs were identified as well as (TA)_n. In total, there were 26 unique SNPs based on the re-sequencing of the *UGT1A1* genomic region. Of these, 14 SNPs in addition to (TA)_n were selected for the expanded genotyping in the current study while other SNPs were not pursued due to high repetitive sequences within primer sets.

PCR for (TA)_n

Variants were genotyped as described previously (Te et al., 2000). Briefly, approximately 5 ng of DNA was amplified by PCR in 10 μ l with primers flanking the polymorphic TA locus in the promoter region of the *UGT1A1* gene. The only protocol modifications to cycling were a final extension at 72°C for 5 min compared with 10 minutes as previously described. Control DNAs from individuals known to have a 6/6, 6/7, and 7/7 genotype were included in the PCR analysis. The amplified product yielded a 93, 95, 97, 99 bp fragment which corresponded to (TA)₅, (TA)₆, (TA)₇, and (TA)₈, respectively. Samples were

DMD #14183

diluted 1:4, subjected to ABI 377 3700 Automated DNA Sequencer for (TA)_n, and scored via Gene Mapper 3.7 (ABI, Foster City, CA).

SNP Genotyping

Genotyping was performed using the Beckman-Coulter SNPstream and procedures followed the specification 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. Reactions were composed of PCR primers at a final concentration of 50nM, 0.2 U enzyme Hot Master Taq (Eppendorf, Hamburg, Germany), and 2-5ng of genomic DNA per reaction. Amplification was performed according to manufacturer's conditions. All post amplification steps were performed according to Beckman SNPstream specifications. Briefly, PCR clean up was accomplished by treating reactions with 3ul of Exo/SAP (USB, Cleveland, OH) with incubation for 30 minutes at 37°C followed by 100°C for 10 minutes. SBE reactions were performed using reagents and protocols specific to the Beckman SNPstream platform. SBE primers were present at a final concentration of 20nM. Reactions were hybridized to Beckman Array plates and scanned by the SNPstream. Genotyping of SNPs was done by DNAPrint™ Genomics (Sarasota, FL).

Bilirubin Levels

To explore the relationship between UGT1A1 genotypes and total bilirubin levels, data on serum total bilirubin levels were obtained from the University of Chicago Hospital medical records. Total bilirubin values for patients diagnosed with cancer were collected prior to start of any treatments. If multiple values were available for a patient, the mean total bilirubin level was calculated. The standard deviation for these patients ranged from 0.05 mg/dl to 0.35 mg/dl, with

DMD #14183

a median of 0.11 mg/dl. For individuals who had known liver metastasis or significant hepatic dysfunction, their bilirubin data were excluded for the analysis.

Statistical Analysis

Hardy-Weinberg equilibrium (HWE) for all loci was examined among African Americans and Caucasians separately using a chi-square test. Differences in allele frequencies between African Americans and Caucasians were compared using a chi-square test or Fisher's exact test, as appropriate. Pair-wise LD was measured by Lewontin's D' (Hedrick, 1987) and the LD matrices were plotted using the GOLD software package (Abecasis and Cookson, 2000). Haplotypes were reconstructed using a Bayesian statistical method implemented in Phase 2.1 (Stephens et al., 2001; Stephens and Donnelly, 2003). To describe the haplotype diversity, we calculated the effective number of haplotypes separately for each racial group (Carlson et al., 2004). The effective number of haplotypes, analogous to the effective number of alleles (Hartl and Clark, 1997), was calculated as $N_e = 1/\sum p_i^2$, where p_i is the frequency of the i th haplotype. To find DNA regions in which there was no apparent historical recombination, we first searched for haplotype blocks using pair-wise D' implemented in HaploBlockFinder (Zhang and Jin, 2003). Then, the minimum blocks were expanded by adding SNPs at the ends using the following criteria. If the calculated D' values between haplotypes contained in the minimum blocks and the SNP at the end was greater than pre-specified value (e.g. 0.8), we concluded that the added marker belonged to the same block. We proposed this haplotype block finding method because multi-locus LD may not be adequately reflected by pair-wise LD. Within each block, haplotype-tagging SNP (or microsatellite marker) was identified manually with help from HaploBlockFinder.

DMD #14183

Since the distribution of total serum bilirubin was skewed to the right, a natural log transformation was done. For easier interpretation, geometric means and 95% confidence interval (CI) are presented. One-way analysis of variance (ANOVA) was used to test whether bilirubin level varied significantly by TA repeat or SNP genotypes, without assuming a genetic model. Coefficient of determination (R^2) obtained from the ANOVA was used to indicate the predictive value. Multiple linear regression models included TA repeats and individual SNPs to examine the independent effect of SNPs after adjusting for TA repeats. Correspondingly, a partial R^2 was calculated to indicate the additional predictive value beyond TA repeats. We also examined the independent effect of TA repeats, race, breast cancer status, and age on bilirubin levels using multiple linear regression models. Finally, we examined the relationship between the inferred haplotypes and bilirubin levels in the Caucasian sample using linear regression models. P values less than 0.05 are considered statistically significant.

Results

UGT1A1 Polymorphisms

The allele frequencies of *UGT1A1* SNPs and TA repeats are presented by self-reported race in Table 2. TA repeats and all SNPs followed HWE except for marker 11 in African Americans. The observed genotype frequency of marker 11 was 38 for genotype AA, 108 for AG, and 147 for GG, which differs from the expected frequency based on HWE (28.9 for AA, 126.2 for AG, and 137.9 for GG; $p=0.015$). The observed deficiency of heterozygosity of marker 11 may be due to chance alone because all markers flanking marker 11 were in HWE.

The allele frequencies of TA repeats and all SNPs genotyped were significantly different between Caucasian and African Americans (Table 2). Caucasians had only (TA)₆ and (TA)₇

DMD #14183

repeats, whereas African Americans had broader variability in the number of TA repeats, including (TA)₅ and (TA)₈. Marker 6 appeared to be a unique SNP in African Americans. The minor allele frequencies (MAF) of markers 13 and 15 in Caucasians and of marker 10 in African Americans were less than 5%.

Linkage Disequilibrium and Haplotype Structure of UGT1A1 Gene

Figure 1 shows pairwise LD measured by Lewontin's D' for the two racial groups. In African Americans, there were two LD clusters: markers 1 to 5 (corresponding to the range from promoter to intron 1) and markers 9 to 15 (3' UTR of exon 5 and conserved regions). A historical recombination hotspot appears to exist between the two clusters. A similar LD pattern was observed in Caucasians, except that there was a long-range LD between markers 3-5 and markers 11-15.

Haplotype analysis was performed separately for African Americans and Caucasians. The haplotype diversity was much higher in African Americans than in Caucasians, with 83 haplotypes inferred in African Americans versus 23 haplotypes inferred in Caucasians. The effective number of haplotypes (the number of equally frequent haplotypes that would be required to produce the same homozygosity as observed) was 17.7 and 5.4 in African Americans and Caucasians, respectively. In Caucasians, there were three common haplotypes (frequency > 5%), accounting for 71.8% of all chromosomes. In contrast, the 5 common haplotypes accounted for 46.6% of chromosomes in African Americans (Table 3). The top three haplotypes are the same between the two racial groups but with quite different frequencies.

Two haplotype blocks were observed for both Caucasians and African Americans, with one block consisting of markers 1-7, and the other block consisting of markers 9-15. Marker 8 did

DMD #14183

not fall into either block since it was in moderate LD with either block 1 or 2 ($D' \sim 0.6$).

Although the gene was divided into two haplotype blocks, there was weak but non-ignorable LD between the two blocks ($D' = 0.52$ in African Americans and $D' = 0.45$ in Caucasians). In Caucasians, the first block was defined by two major and one minor haplotypes, and only two markers (TA repeats and any one of markers 1, 3, 5 and 7) were required to tag them. The second block in Caucasians was defined by four common haplotypes, requiring three haplotype-tagging SNPs. In African Americans, five common haplotypes accounted for 85% of the first block and four markers were required to tag them. The second block in African Americans consisted of five common haplotypes and required four haplotype-tagging SNPs. In both blocks, haplotypes were more diversified in African Americans than in Caucasians.

Bilirubin Level, UGT1A1 Polymorphism, and Race

Total serum bilirubin levels were available for 125 patients (89 Caucasians and 36 African Americans). In the one-way ANOVA analysis, bilirubin levels were significantly associated with TA repeats (marker 2) and several other SNPs (markers 1, 4, 5, 9, 12, 13, and 14). Figure 2A shows the percentage of the variability in serum bilirubin levels explained by each marker (R^2). Specifically, the variability in bilirubin concentration explained by TA repeats was 27% ($p < 0.0001$). However, after adjusting for TA repeats, only marker 5 was marginally significant ($p = 0.05$). Figure 2B shows that 5% or less of variability in bilirubin level was explained by each SNP beyond TA repeats. The geometric means of bilirubin levels by TA repeat genotypes are presented in Table 4. Homozygous carriers of $(TA)_7$ had the highest serum bilirubin level, followed by heterozygous carriers of $(TA)_6/(TA)_7$. Homozygous carriers of $(TA)_6$ and heterozygous carriers of $(TA)_5/(TA)_6$ had low levels of bilirubin. There were only a few carriers of $(TA)_5/(TA)_7$ and $(TA)_6/(TA)_8$ and thus no inference was made for these categories. TA repeats

DMD #14183

was still a strong determinant of bilirubin levels (adjusted $p < 0.0001$) after adjusting for age, race, and breast cancer diagnosis. Interestingly, bilirubin levels were significantly lower (adjusted $p = 0.02$) in patients diagnosed with breast cancer (0.40 mg/dl) than in those without breast cancer (0.46 mg/dl).

Serum bilirubin levels were 22% higher ($p = 0.005$) in Caucasians (0.44 mg/dl) than in African Americans (0.36 mg/dl) (Table 4). This difference in bilirubin level was not explained by the distribution of TA repeats alone: after adjusting for TA repeats, Caucasians still had higher bilirubin level than African Americans ($p = 0.018$). Stratified by race, 29% of the variation in total bilirubin level was explained by TA repeats for Caucasians and 17% for African Americans (Figure 3).

Of the 89 Caucasian patients with bilirubin data, 3 patients were excluded from the analysis of haplotype and bilirubin since their haplotypes were constructed with a high level of uncertainty (probability $< 80\%$). Table 5 shows the diplotype configuration (haplotype pairs) and the relationship with total serum bilirubin levels in Caucasians. Bilirubin level varied significantly across the diplotypes in block 1 ($R^2 = 0.33$, $p < 0.0001$) and the diplotypes in block 2 ($R^2 = 0.22$, $p = 0.01$). However, this haplotype-phenotype correlation is mainly due to the effect of TA repeat polymorphism. After removing the effect of TA repeat polymorphism using linear regression models, neither block 1 (partial $R^2 = 0.06$, $p = 0.17$) nor block 2 haplotypes (partial $R^2 = 0.14$, $p = 0.14$) were significantly correlated with bilirubin. This is because $(TA)_n$ was one of the tagging markers for block 1 haplotypes and it was also moderately linked with haplotypes in block 2 (Table 5). In the 36 African Americans with bilirubin data, 20 distinct diplotypes were found in block 1 and 19 diplotypes were found in block 2. Therefore the study lacked statistical

DMD #14183

power to examine the relationship between *UGT1A1* haplotypes and bilirubin levels in African Americans.

Discussion

This is one of the first large scale studies to examine the distribution of *UGT1A1* variants other than TA repeats among an ethnically diverse cohort of women. Furthermore, we linked these variants with serum total bilirubin, which is an important *UGT1A1* phenotype.

The distribution of the number of TA repeats observed in this study is consistent with previous studies (Bosma et al., 1995; Beutler et al., 1998; Guillemette et al., 2001; Innocenti et al., 2002). (TA)₅ and (TA)₈ was each found in about 5% of African Americans but in none of the Caucasians. The allele frequency of (TA)₇ was slightly higher in African Americans (0.375) than in Caucasians (0.323). *UGT1A1* polymorphisms other than TATA box have not been well studied and several SNPs (M6, M8 and M10) identified in this study are novel. The allele frequencies and corresponding genotypes of these SNPs were quite different between African Americans and Caucasians. The MAFs of marker 9 (c.*211) observed in this study (0.276 for Caucasians and 0.382 for African Americans) were comparable to a previous study (labeled as 1813C>T), which reported a frequency of 0.220 for Caucasians and 0.350 for African Americans (Kaniwa et al., 2005). Furthermore, with the exception of SNP M5, the MAFs are comparable to that of the CEPH and Yoruba groups from the Environmental Genome and HapMap Projects (2003; 2006). For M5, the MAF for African Americans (0.431) is between the values from the EGP data set (Yoruba 0.292 and CEPH 0.455) which may be due to the larger sample size in this study or due to admixture of African Americans.

DMD #14183

We described the haplotype structure of *UGT1A1* gene from its PBREM region to 3'-conserved region. We found it was better to describe the haplotypes using two blocks for both Caucasians and African Americans with the advantage of low haplotype diversity within each block. For example, in African Americans, the effective number of haplotypes for blocks 1 and 2 was 4.6 and 5.2, respectively, less than that for the whole gene ($N_e=17.7$). Sai et al also reported that *UGT1A1* could be divided into two blocks among Japanese (Sai et al., 2004) and Kaniwa et al used the same two blocks to describe *UGT1A1* genes in African Americans, Caucasians and Japanese (Kaniwa et al., 2005). The boundaries of the two blocks proposed by Sai et al are similar to this study. In our study, the first block ranges from PBREM and promoter regions to intron 2, instead of to exon 1 as in the previous study (Sai et al., 2004). This discrepancy may be due to the arbitrary cutoff points of LD measures, as markers 6-7 in intron 2 were not in complete LD with markers 3-5 in intron 1. We believe it is better to include intron 1 in block 1 for African Americans and Caucasians as shown in Figure 1 and Table 3. In Caucasians, marker 1 (at -1352 upstream of transcription starting point) and TA repeats were sufficient to tag the common haplotypes in block 1: C7 0.323, A6 0.569, C6 0.107. Interestingly, these haplotype frequencies were similar to those tagged by markers -3279T>G and TA repeats for Caucasians in previous studies: G7 0.36-0.39, T6 0.45-0.53, G6 0.09-0 (Innocenti et al., 2002; Kaniwa et al., 2005), suggesting that these three haplotypes are the same although different markers were used. For African Americans, the haplotypes tagged by marker 1 and TA repeats in this study were also comparable to that tagged using markers -3279T>G and TA repeats in previous studies (Innocenti et al., 2002; Kaniwa et al., 2005). However, the haplotype diversity within block 1 was greater in African Americans than in Caucasians and thus more tagging markers are required (Table 3). Similarly, the haplotype diversity within block 2 was also greater in African

DMD #14183

Americans than in Caucasians. In this study, the haplotype block 2 extended from 3' UTR of exon 5 to the conserved region. To our knowledge, no previous study had examined this. These ethnic differences in SNP and haplotype distribution suggest that the presence and strength of the association between *UGT1A1* genotypes and phenotypes including drug toxicities and cancer susceptibility need confirmation in independent populations.

Consistent with a previous study conducted in Caucasians (Bosma et al., 1995), we showed that TA repeats were strongly correlated with total bilirubin within normal range, accounting for 29% of variations among Caucasians and 17% among African Americans. It is conceivable that the observed differences between the two groups is due to the diversity in the African Americans and the relatively small sample of African Americans patients with serum total bilirubin levels available for study. Nonetheless, our observation that serum total bilirubin levels were also lower among African American is consistent with previous population based studies (Zucker et al., 2004). TA repeat polymorphism was also shown to be responsible for variation of bilirubin in Asian individuals with a similar R^2 (Ki et al., 2003; Sai et al., 2004). In addition, an Asian-specific variant, 211G>A (*6, G71R) in exon 1 has been associated with variation in bilirubin (Huang et al., 2000; Ki et al., 2003; Sai et al., 2004). A study conducted in Japanese found that 1813C>T in 3'-untranslated region of exon 5 was weakly but significantly associated with bilirubin level after adjusting for TA repeats ($p=0.03$) (Sai et al., 2004). In this study, we showed that the same SNP, M9, was not significant after adjusting for TA repeats ($p=0.11$). Among Caucasian, we did not find other variant that was associated with bilirubin conditional on TA repeats, as shown in both the genotype analysis (Figure 2) and haplotype analysis (Table 5). Our sample size for African Americans was relatively small, thus we can not rule out whether there are other *UGT1A1* variants accounting for the variation in bilirubin in African Americans. This

DMD #14183

calls for further investigation of bilirubin levels in a larger study among African Americans as bilirubin can serve as an antioxidant and therefore may be a surrogate biomarker for risk.

A review of UGT literature suggests that bilirubin is selectively glucuronidated by *UGT1A1* (Tukey and Strassburg, 2000). A linkage analysis of pedigree data from the Framingham Study estimated the heritability of serum bilirubin to be $49\% \pm 6\%$ and the peak multipoint LOD score is located 1 cM away from the *UGT1A1* gene. The authors concluded that *UGT1A1* may be the major gene controlling serum bilirubin (Lin et al., 2003) but almost all subjects were Caucasians. Potentially, there are other factors such as other isoforms of UGT or other genes that moderate the variation in serum bilirubin because bilirubin levels also depend on the rate of bilirubin production, transportation and elimination. The presence of Gilbert alleles is not necessarily linked to hyperbilirubinemia (Kadacol et al., 2000). In addition, UGT genes in mice are regulated by factors such as CAAT-enhancer binding protein (C/EBPalpha) (Lee et al., 1997). Recent studies suggest these factors may influence human hepatic enzymes such as UGT2B7 (Ishii et al., 2000). Thus, these factors may play a role by enhancing or suppressing UGT activity and thus bilirubin in humans.

With respect to the racial difference in bilirubin levels, we observed the paradox as pointed out previously (Beutler et al., 1998): African Americans had lower bilirubin levels but had higher frequencies of (TA)₇ and (TA)₈ than Caucasians. In each stratum of TA repeats genotypes, (TA)₆/(TA)₆, (TA)₆/(TA)₇, and (TA)₇/(TA)₇, African Americans still had lower bilirubin levels than Caucasians. It has been documented that African Americans have lower hemoglobin levels than Caucasians (Perry et al., 1992) and the difference in hemoglobin existed after excluding subjects with iron-deficient erythropoiesis and α -thalassemia mutations (Beutler and West, 2005). Other genes such as *G6PD* and *OATP2* or environmental factors may account for this

DMD #14183

racial difference. For example, *OATP-C* has been correlated with conjugated bilirubin in a pilot study of 23 patients (Jeiri et al., 2004) This finding warrants further study because same total serum bilirubin in African American and Caucasian patients who are treated with glucuronidated drugs may indicate different toxicity profiles.

In conclusion, this study reconstructed the haplotypes of *UGT1A1* gene spanning from transcription-regulating region to 3'-conserved region with (TA)_n repeats and 14 SNPs and showed wider variation of *UGT1A1* haplotypes in African American women than in Caucasian women. The racial differences in haplotype distribution may be responsible for different toxicity profiles for African American and Caucasian patients treated with glucuronidated drugs and cancer susceptibility, suggesting that more than (TA)_n profile is needed in the African American population. Regarding functions of *UGT1A1* polymorphisms, it is clear that TA repeats is an important factor for total bilirubin variation in Caucasian population and there may exist other variants related to bilirubin within *UGT1A1* gene but these variants are less likely to be clinically significant. In African Americans, the role of TA repeats remains unclear but further studies with larger sample sizes are necessary to elucidate the function of other variants and haplotypes. Future studies are planned to examine the *UGT1A1* gene with a larger sample of African Americans as well as a sample of Nigerians.

DMD #14183

Acknowledgements

We would like to acknowledge Yu-Sheng Chen, Jim Frankenthal, and members of the Olopade lab for their advice and support.

DMD #14183

References

- (2003) The International HapMap Project. *Nature* **426**:789-796.
- (2006) Environmental Genome Project, The National Institute of Environmental Health Sciences.
- Abecasis GR and Cookson WO (2000) GOLD--graphical overview of linkage disequilibrium. *Bioinformatics* **16**:182-183.
- Adegoke OJ, Shu XO, Gao YT, Cai Q, Breyer J, Smith J and Zheng W (2004) Genetic polymorphisms in uridine diphospho-glucuronosyltransferase 1A1 (UGT1A1) and risk of breast cancer. *Breast Cancer Res Treat* **85**:239-245.
- Beutler E, Gelbart T and Demina A (1998) Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? *Proc Natl Acad Sci U S A* **95**:8170-8174.
- Beutler E and West C (2005) Hematologic differences between African-Americans and whites: the roles of iron deficiency and alpha-thalassemia on hemoglobin levels and mean corpuscular volume. *Blood* **106**:740-745.
- Bosma PJ, Chowdhury JR, Bakker C, Gantla S, de Boer A, Oostra BA, Lindhout D, Tytgat GN, Jansen PL, Oude Elferink RP and et al. (1995) The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. *N Engl J Med* **333**:1171-1175.
- Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L and Nickerson DA (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* **74**:106-120.

DMD #14183

Chouinard S, Tessier M, Vernouillet G, Gauthier S, Labrie F, Barbier O and Belanger A (2006)

Inactivation of the pure antiestrogen fulvestrant and other synthetic estrogen molecules by UDP-glucuronosyltransferase 1A enzymes expressed in breast tissue. *Mol Pharmacol* **69**:908-920.

Grant DJ and Bell DA (2000) Bilirubin UDP-glucuronosyltransferase 1A1 gene polymorphisms: susceptibility to oxidative damage and cancer? *Mol Carcinog* **29**:198-204.

Guillemette C, De Vivo I, Hankinson SE, Haiman CA, Spiegelman D, Housman DE and Hunter DJ (2001) Association of genetic polymorphisms in UGT1A1 with breast cancer and plasma hormone levels. *Cancer Epidemiol Biomarkers Prev* **10**:711-714.

Guillemette C, Millikan RC, Newman B and Housman DE (2000) Genetic polymorphisms in uridine diphospho-glucuronosyltransferase 1A1 and association with breast cancer among African Americans. *Cancer Res* **60**:950-956.

Hartl DL and Clark AG (1997) *Principles of population genetics*. Sinauer Associates, Sunderland, MA.

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. *Genetics* **117**:331-341.

Huang CS, Luo GA, Huang ML, Yu SC and Yang SS (2000) Variations of the bilirubin uridine-diphosphoglucuronosyl transferase 1A1 gene in healthy Taiwanese. *Pharmacogenetics* **10**:539-544.

Ieiri I, Suzuki H, Kimura M, Takane H, Nishizato Y, Irie S, Urae A, Kawabata K, Higuchi S, Otsubo K and Sugiyama Y (2004) Influence of common variants in the pharmacokinetic genes (OATP-C, UGT1A1, and MRP2) on serum bilirubin levels in healthy subjects. *Hepatol Res* **30**:91-95.

DMD #14183

Innocenti F, Grimsley C, Das S, Ramirez J, Cheng C, Kuttub-Boulos H, Ratain MJ and Di

Rienzo A (2002) Haplotype structure of the UDP-glucuronosyltransferase 1A1 promoter in different ethnic groups. *Pharmacogenetics* **12**:725-733.

Innocenti F, Undevia SD, Iyer L, Chen PX, Das S, Kocherginsky M, Karrison T, Janisch L,

Ramirez J, Rudin CM, Vokes EE and Ratain MJ (2004) Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *J Clin Oncol* **22**:1382-1388.

Ishii Y, Hansen AJ and Mackenzie PI (2000) Octamer transcription factor-1 enhances hepatic

nuclear factor-1alpha-mediated activation of the human UDP glucuronosyltransferase 2B7 promoter. *Mol Pharmacol* **57**:940-947.

Iyer L, Das S, Janisch L, Wen M, Ramirez J, Karrison T, Fleming GF, Vokes EE, Schilsky RL

and Ratain MJ (2002) UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharmacogenomics J* **2**:43-47.

John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, Ziogas A, Andrulis IL,

Anton-Culver H, Boyd N, Buys SS, Daly MB, O'Malley FP, Santella RM, Southey MC, Venne VL, Venter DJ, West DW, Whittemore AS and Seminara D (2004) The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res* **6**:R375-389.

Kadakol A, Ghosh SS, Sappal BS, Sharma G, Chowdhury JR and Chowdhury NR (2000)

Genetic lesions of bilirubin uridine-diphosphoglucuronate glucuronosyltransferase (UGT1A1) causing Crigler-Najjar and Gilbert syndromes: correlation of genotype to phenotype. *Hum Mutat* **16**:297-306.

DMD #14183

- Kaniwa N, Kurose K, Jinno H, Tanaka-Kagawa T, Saito Y, Saeki M, Sawada J, Tohkin M and Hasegawa R (2005) Racial variability in haplotype frequencies of UGT1A1 and glucuronidation activity of a novel single nucleotide polymorphism 686C> T (P229L) found in an African-American. *Drug Metab Dispos* **33**:458-465.
- Ki CS, Lee KA, Lee SY, Kim HJ, Cho SS, Park JH, Cho S, Sohn KM and Kim JW (2003) Haplotype structure of the UDP-glucuronosyltransferase 1A1 (UGT1A1) gene and its relationship to serum total bilirubin concentration in a male Korean population. *Clin Chem* **49**:2078-2081.
- Lee YH, Sauer B, Johnson PF and Gonzalez FJ (1997) Disruption of the c/ebp alpha gene in adult mouse liver. *Mol Cell Biol* **17**:6014-6022.
- Lin JP, Cupples LA, Wilson PW, Heard-Costa N and O'Donnell CJ (2003) Evidence for a gene influencing serum bilirubin on chromosome 2q telomere: a genomewide scan in the Framingham study. *Am J Hum Genet* **72**:1029-1034.
- Lin JP, O'Donnell CJ, Schwaiger JP, Cupples LA, Lingenhel A, Hunt SC, Yang S and Kronenberg F (2006) Association between the UGT1A1*28 allele, bilirubin levels, and coronary heart disease in the Framingham Heart Study. *Circulation* **114**:1476-1481.
- Maitland ML, Grimsley C, Kuttub-Boulos H, Witonsky D, Kasza KE, Yang L, Roe BA and Di Rienzo A (2006) Comparative genomics analysis of human sequence variation in the UGT1A gene cluster. *Pharmacogenomics J* **6**:52-62.
- Malfatti MA, Ubick EA and Felton JS (2005) The impact of glucuronidation on the bioactivation and DNA adduction of the cooked-food carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine in vivo. *Carcinogenesis* **26**:2019-2028.

DMD #14183

- Perry GS, Byers T, Yip R and Margen S (1992) Iron nutrition does not account for the hemoglobin differences between blacks and whites. *J Nutr* **122**:1417-1424.
- Sai K, Saeki M, Saito Y, Ozawa S, Katori N, Jinno H, Hasegawa R, Kaniwa N, Sawada J, Komamura K, Ueno K, Kamakura S, Kitakaze M, Kitamura Y, Kamatani N, Minami H, Ohtsu A, Shirao K, Yoshida T and Saijo N (2004) UGT1A1 haplotypes associated with reduced glucuronidation and increased serum bilirubin in irinotecan-administered Japanese patients with cancer. *Clin Pharmacol Ther* **75**:501-515.
- Senafi SB, Clarke DJ and Burchell B (1994) 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* **303** (Pt 1):233-240.
- Spitz MR, Wu X and Mills G (2005) Integrative epidemiology: from risk assessment to outcome prediction. *J Clin Oncol* **23**:267-275.
- Stephens M and Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* **73**:1162-1169.
- Stephens M, Smith NJ and Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**:978-989.
- Strassburg CP, Oldhafer K, Manns MP and Tukey RH (1997) Differential expression of the UGT1A locus in human liver, biliary, and gastric tissue: identification of UGT1A7 and UGT1A10 transcripts in extrahepatic tissue. *Mol Pharmacol* **52**:212-220.
- Te HS, Schiano TD, Das S, Kuan SF, DasGupta K, Conjeevaram HS and Baker AL (2000) Donor liver uridine diphosphate (UDP)-glucuronosyltransferase-1A1 deficiency causing Gilbert's syndrome in liver transplant recipients. *Transplantation* **69**:1882-1886.

DMD #14183

Tukey RH and Strassburg CP (2000) Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu Rev Pharmacol Toxicol* **40**:581-616.

Zhang K and Jin L (2003) HaploBlockFinder: haplotype block analyses. *Bioinformatics* **19**:1300-1301.

Zucker SD, Horn PS and Sherman KE (2004) Serum bilirubin levels in the U.S. population: gender effect and inverse correlation with colorectal cancer. *Hepatology* **40**:827-835.

DMD #14183

Footnotes

a) Unnumbered footnote:

Financial Support: Supported by NCI grant CA-RO1 89085-01A, the Falk Medical Research Trust and the United States National Cancer Institute, National Institutes of Health, under Request for Application CA-95-003 as part of the Breast Cancer Family Registries, and through cooperative agreements with the Northern California Cancer Center. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of collaborating centers in the Breast CFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the Breast CFR.

b) Reprints sent to: Olufunmilayo I. Olopade, MD, Department of Medicine, Section of Hematology/Oncology, University of Chicago, 5841 S. Maryland Ave, MC 2115, Chicago, IL 60637. (email: folopade@medicine.bsd.uchicago.edu)

c) Numbered footnotes

¹These authors contributed equally to this work

DMD #14183

Figure Legend

Figure 1. Linkage disequilibrium (Lewontin D') of 15 consecutive markers in *UGT1A1* in Caucasians (A) and African Americans (B). These graphical displays were generated by GOLD and are a visual representation of the LD statistics. The scales on the right indicate the Lewontin D' values. A color scheme closer to red ($D' = 1.00$) represents strong association of the SNPs (linkage disequilibrium) whereas a color scheme closer to blue ($D' = 0.00$) represents areas of linkage equilibrium. The SNPs are arranged by the physical order and distance on *UGT1A1* along the x and y axes.

Figure 2. (A) R-square calculated from simple linear regressions of bilirubin on each *UGT1A1* gene polymorphic marker. (B) Partial R-square calculated from multiple linear regressions of bilirubin on each *UGT1A1* gene polymorphic marker adjusting for TA repeat polymorphism in all subjects.

Figure 3. Box plot of bilirubin levels by race and TA repeat polymorphisms of *UGT1A1*. Box represents lower quartile, median and upper quartile. Whiskers extend to at most 1.5 times the [interquartile range](#). Dots represent observations outside of whiskers.

DMD #14183

Tables

Table 1. Characteristics of Study Sample

| | African Americans | Caucasians | Total |
|---------------------------------|----------------------|-----------------|-----------------|
| No. of total | 335 | 181 | 516 |
| Enrollment site | | | |
| San Francisco Bay area | 289 | | 289 |
| Chicago | 46 | 181 | 227 |
| Age, mean \pm SD | 49.2 \pm 10.0 | 48.0 \pm 10.0 | 48.8 \pm 10.0 |
| No. with breast cancer | 282 | 91 | 373 |
| No. of related individuals | 6 | 20 | 26 |
| Serum total bilirubin available | 36 | 89 | 125 |

DMD #14183

Table 2. Polymorphisms in *UGT1A1* gene by race

| Marker | Location | Reference SNP ID | Position* | cDNA position† | Major > Minor Allele | Allele frequency‡ | |
|--------|---------------------|------------------|-----------|----------------|----------------------|---------------------|----------------------------|
| | | | | | | Caucasians (2n=362) | African Americans (2n=670) |
| M1 | PBREM region§ | rs3755319 | 173691 | c.-1352 | C>A | 0.570 | 0.333 |
| M2 | Promoter | | ~174990 | c.-53 | TA(5) | 0 | 0.063 |
| | | | | | TA(6) | 0.677 | 0.521 |
| | | | | | TA(7) | 0.323 | 0.375 |
| | | | | | TA(8) | 0 | 0.041 |
| M3 | Intron 1 | rs3771342 | 178771 | c.864+2865 | G>T | 0.107 | 0.218 |
| M4 | Intron 1 | rs4148325 | 179417 | c.865-2371 | C>T | 0.322 | 0.417 |
| M5 | Intron 1 | rs4148326 | 179570 | c.865-2218 | C>T | 0.575 | 0.431 |
| M6 | Intron 2 | | 181937 | c.996+18 | C>T | 0 | 0.100 |
| M7 | Intron 2 | rs1018124 | 182226 | c.996+307 | A>G | 0.081 | 0.129 |
| M8 | Intron 2 | | 182521 | c.997-82 | T>C | 0.106 | 0.352 |
| M9 | 3' UTR of Exon 5 | rs10929303 | 187524 | c.*211 | C>T | 0.276 | 0.382 |
| M10 | 3'-Conserved region | | 188164 | c.*851 | C>A | 0.128 | 0.035 |
| M11 | 3'-Conserved region | rs4148329 | 188170 | c.*857 | C>T | 0.500 | 0.314 |
| M12 | 3'-Conserved region | rs6717546 | 188227 | c.*914 | G>A | 0.430 | 0.508 |
| M13 | 3'-Conserved region | rs1500476 | 192593 | c.*5280 | C>T | 0.009 | 0.207 |
| M14 | 3'-Conserved region | rs6431631 | 194144 | c.*6831 | A>C | 0.289 | 0.179 |
| M15 | 3'-Conserved region | rs17863803 | 194435 | c.*7122 | A>G | 0.014 | 0.115 |

*Physical locations are based on GenBank reference sequence AF297093.1

†Physical locations based on GenBank reference sequence NM_000463.2 and ascribing to HGSV nomenclature, <http://www.hgvs.org/mutnomen/>

‡Minor allele frequencies of total sample are presented except for TA repeat.

§Phenobarbital responsive module

Table 3. Common haplotypes and haplotype blocks within *UGT1A1* gene*

| Caucasians (2n=362) | | African Americans (2n=670) | |
|---------------------|---------|----------------------------|---------|
| M1-M15 | | M1-M15 | |
| A6GCTCATCCTGCAA | (28.2%) | A6GCTCATCCTGCAA | (11.2%) |
| A6GTCATTCCACCA | (24.0%) | A6GTCATTCCACCA | (8.2%) |
| C7GTCCATCCTGCAA | (19.6%) | C7GTCCATCCTGCAA | (13.1%) |
| | | C6GCTCACCCCGCAG | (7.8%) |
| | | C7GTCCATTCCATAA | (6.3%) |
| Total | (71.8%) | Total | (46.6%) |
| Ne† | 5.4 | Ne | 17.7 |

| Caucasians (2n=362) | | African Americans (2n=670) | | | | | |
|--|--|--|--|--|---|--------|---------|
| M1-M7 | M8 | M1-M7 | M8 | | | | |
| <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;">{</div> <div style="border: 1px solid black; padding: 2px;">A6GCTCA</div> (56.9%) <div style="border: 1px solid black; padding: 2px;">C7GTCCA</div> (31.8%) <div style="border: 1px solid black; padding: 2px;">C6TCCCG</div> (8.3%) </div> | <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;">{</div> <div style="margin-right: 10px;">T</div> (89.5%) <div style="margin-right: 10px;">C</div> (10.5%) </div> | <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;">{</div> <div style="border: 1px solid black; padding: 2px;">CCTGCAA</div> (50.6%) <div style="border: 1px solid black; padding: 2px;">TCCACCA</div> (27.3%) <div style="border: 1px solid black; padding: 2px;">CACACAA</div> (13.0%) <div style="border: 1px solid black; padding: 2px;">CCCGTA</div> (5.5%) </div> | <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;">{</div> <div style="border: 1px solid black; padding: 2px;">A6GCTCA</div> (24.5%) <div style="border: 1px solid black; padding: 2px;">C7GTCCA</div> (36.4%) <div style="border: 1px solid black; padding: 2px;">C6TCCCG</div> (8.5%) <div style="border: 1px solid black; padding: 2px;">C6GCTCA</div> (9.4%) <div style="border: 1px solid black; padding: 2px;">A5GCTTA</div> (6.3%) </div> | <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;">{</div> <div style="margin-right: 10px;">T</div> (64.9%) <div style="margin-right: 10px;">C</div> (35.1%) </div> | <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;">{</div> <div style="border: 1px solid black; padding: 2px;">CCTGCAA</div> (33.9%) <div style="border: 1px solid black; padding: 2px;">TCCACCA</div> (17.5%) <div style="border: 1px solid black; padding: 2px;">TCCATAA</div> (17.0%) <div style="border: 1px solid black; padding: 2px;">CCCGCAG</div> (11.2%) <div style="border: 1px solid black; padding: 2px;">CCACAA</div> (5.8%) </div> | | |
| Total | (97.0%) | (100%) | (96.4%) | Total | (85.1%) | (100%) | (85.4%) |
| Ne | 2.3 | | 2.9 | Ne | 4.6 | | 5.2 |

*Braces indicate haplotype blocks and boxes indicate haplotype-tagging single nucleotide polymorphism (SNP) or microsatellite marker.

†Ne, effective number of haplotype

DMD Fast Forward. Published on May 3, 2007 as DOI: 10.1124/dmd.106.014183
This article has not been certified and formatted. The final version may differ from this version.

DMD #14183

Table 4. Geometric means (95% confidence interval) of serum total bilirubin levels by TA repeat polymorphism and race

| | No. | Geometric mean (95% CI) | R ² | P value |
|-------------------|-----|-------------------------|----------------|---------|
| TA repeat | | | 0.271 | <0.0001 |
| 5 6 | 7 | 0.344 (0.278-0.427) | | |
| 5 7 | 1 | 0.260 (0.260-0.260) | | |
| 6 6 | 55 | 0.376 (0.351-0.403) | | |
| 6 7 | 46 | 0.427 (0.386-0.473) | | |
| 6 8 | 2 | 0.304 (0.131-0.707) | | |
| 7 7 | 13 | 0.672 (0.525-0.860) | | |
| Race | | | 0.064 | 0.005 |
| Caucasians | 89 | 0.439 (0.408-0.472) | | |
| African Americans | 36 | 0.361 (0.321-0.405) | | |

DMD #14183

Table 5. Geometric means (95% confidence interval) of serum total bilirubin levels by *UGT1A1* haplotypes in Caucasians

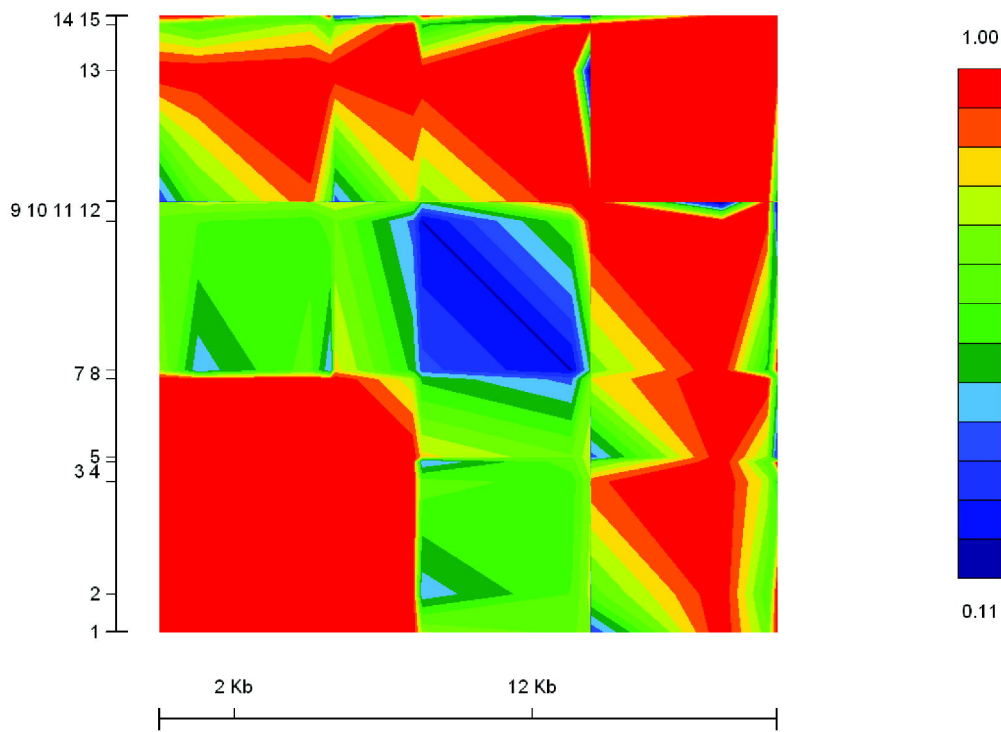
| Haplotype* | No. | Geometric mean (95% CI) | R ² (p-value) | Partial R ² (p-value)† |
|-----------------|-----|----------------------------|-----------------------------|--------------------------------------|
| Block 1 | | | 0.327 (<0.0001) | 0.060 (0.17) |
| A6GCTCA/A6GCTCA | 30 | 0.371 (0.337-0.409) | | |
| A6GCTCA/C7GTCCA | 29 | 0.445 (0.403-0.491) | | |
| A6GCTCA/C6.C.C. | 13 | 0.431 (0.371-0.501) | | |
| C7GTCCA/C7GTCCA | 10 | 0.697 (0.503-0.966) | | |
| C6TCCC./C6TCCCG | 2 | 0.382 (0.156-0.935) | | |
| C6TCCCG/C7GTCCA | 2 | 0.315 (0.256-0.386) | | |
| Block 2 | | | 0.220 (0.01) | 0.146 (0.14) |
| TCCACCA/CCTGCAA | 26 | 0.367 (0.332-0.405) | | |
| CCTGCAA/CCTGCAA | 22 | 0.494 (0.421-0.579) | | |
| CCTGCAA/C.CAC.A | 14 | 0.435 (0.366-0.518) | | |
| TCCACCA/TCCACCA | 8 | 0.408 (0.313-0.532) | | |
| CCTGCAA/CCCGCA. | 7 | 0.524 (0.357-0.769) | | |
| TCCA.../CACACAA | 5 | 0.470 (0.333-0.664) | | |
| TCCACCA/CCCGCAA | 2 | 0.358 (0.221-0.578) | | |
| CACACAA/CCCACAA | 1 | 0.310 | | |
| CCCGCAA/CCCGCAA | 1 | 0.920 | | |

*Markers 1 and 2 were used to tag haplotypes in block 1. Markers 9, 11, and 12 were used to tag haplotypes in block 2. Dots indicate that alleles can not be fully determined by haplotype-tagging markers.

†Partial R² and corresponding p values were calculated in linear regressions after controlling for TA repeats.

Figure 1

A



B

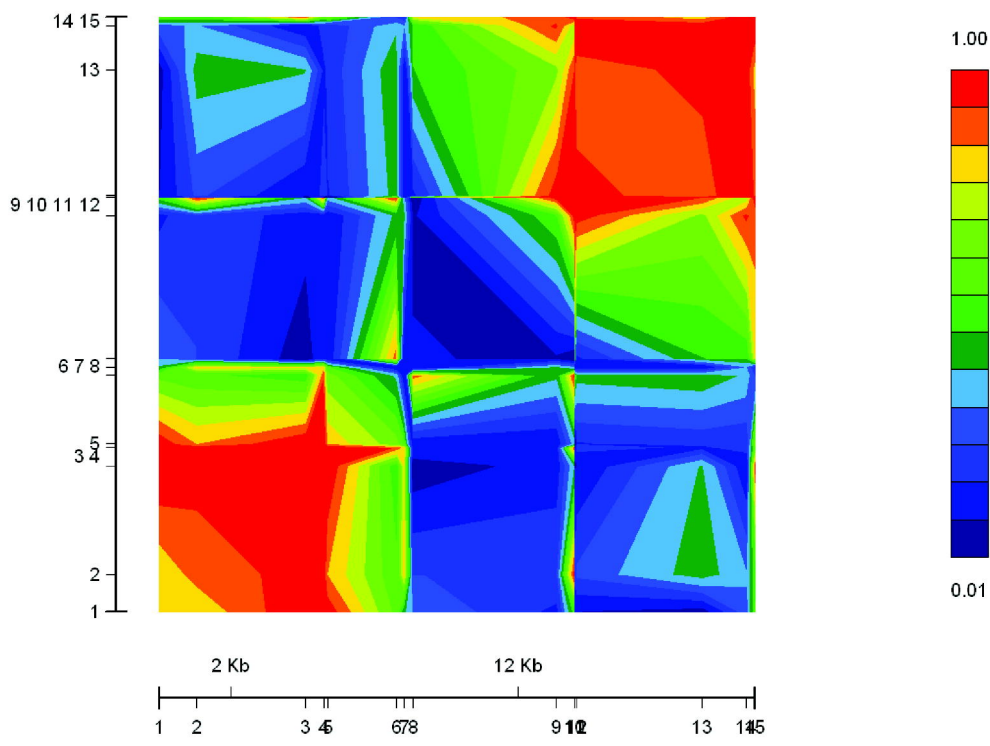
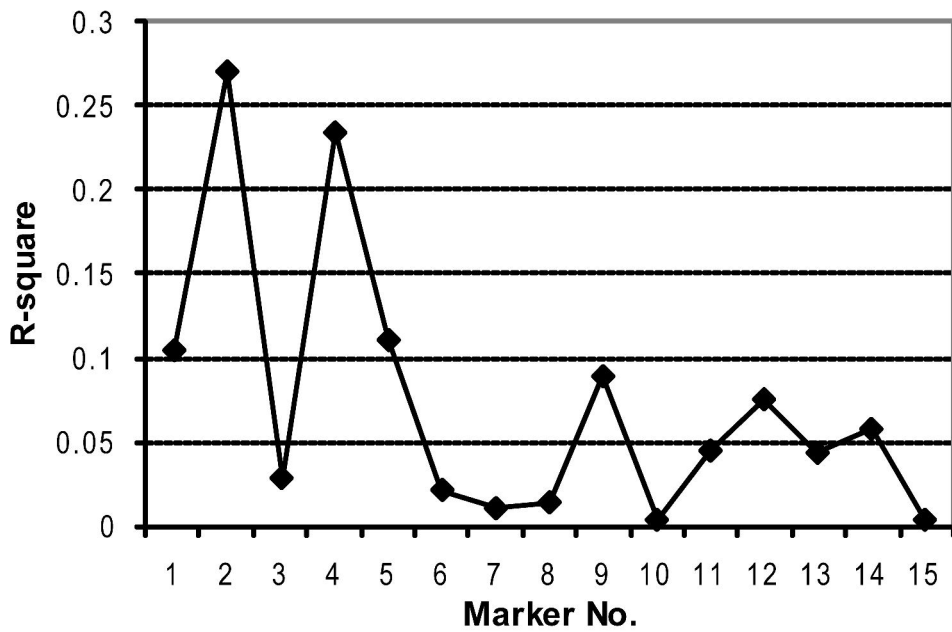


Figure 2

A.



B.

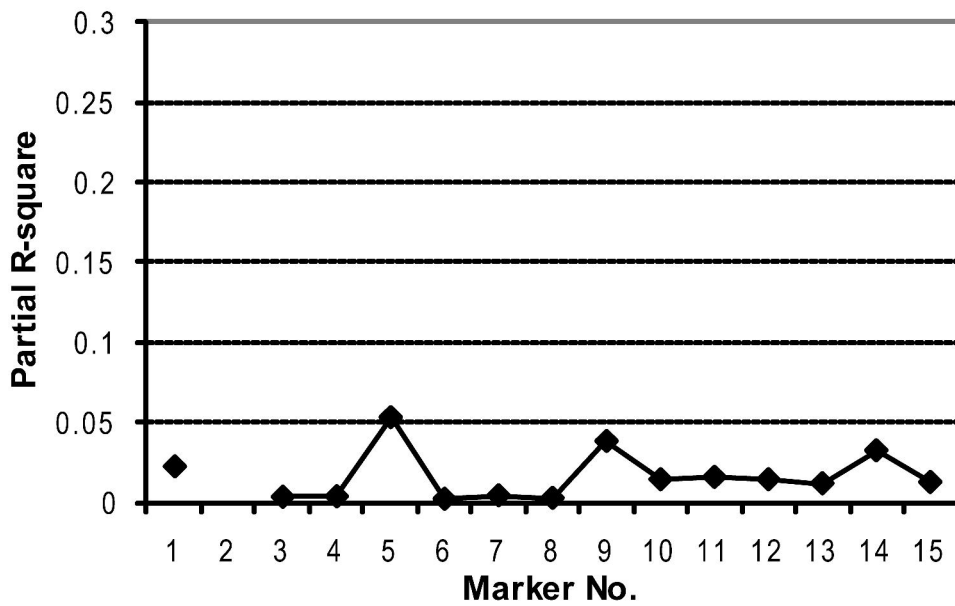


Figure 3

