UGT1A1 is a major locus influencing bilirubin levels in African Americans

Guanjie Chen1, Edward Ramos1, Adebowale Adeyemo1, Daniel Shriner1, Jie Zhou1, Ayo P Doumatey1, Hanxia Huang1, Michael R Erdos2, Norman P Gerry3, Alan Herbert4, Amy R Bentley1, Huichun Xu1, Bashira A Charles1, Michael F Christman3 and Charles N Rotimi1

Total serum bilirubin is associated with several clinical outcomes, including cardiovascular disease, diabetes and drug metabolism. We conducted a genome-wide association study in 619 healthy unrelated African Americans in an attempt to replicate reported findings in Europeans and Asians and to identify novel loci influencing total serum bilirubin levels. We analyzed a dense panel of over two million genotyped and imputed SNPs in additive genetic models adjusting for age, sex, and the first two significant principal components from the sample covariance matrix of genotypes. Thirty-nine SNPs spanning a 78 kb region within the UGT1A1 displayed P-values < 5 x 10^{-8}. The lowest P-value was 1.7 x 10^{-22} for SNP rs887829. None of SNPs in the UGT1A1 remained statistically significant in conditional association analyses that adjusted for rs887829. In addition, SNP rs10929302 located in phenobarbital response enhancer module was significantly associated with bilirubin and displayed a P-value of 1.37 x 10^{-11}; this enhancer module is believed to have a critical role in phenobarbital treatment of hyperbilirubinemia. Interestingly, the lead SNP, rs887829, is in strong linkage disequilibrium (LD) (r^2 < 0.74) with rs10929302. Taking advantage of the lower LD and shorter haplotypes in African-ancestry populations, we identified rs887829 as a more refined proxy for the causative variant influencing bilirubin levels. Also, we replicated the reported association between variants in SEMA3C and bilirubin levels. In summary, UGT1A1 is a major locus influencing bilirubin levels and the results of this study promise to contribute to understanding of the etiology and treatment of hyperbilirubinemia in African-ancestry populations.

European Journal of Human Genetics (2012) 20, 463–468; doi:10.1038/ejhg.2011.206; published online 16 November 2011

Keywords: GWAS; replications; bilirubin; African Americans

INTRODUCTION

Bilirubin is an end product of heme metabolism. Heme, a component of hemoglobin in red blood cells, is broken down by heme oxygenase resulting in bilirubin.1,2 Glucuronidation by UDP-glycosyltransferase in hepatocytes3 is an important step in facilitating the excretion of bilirubin into bile.4 A very high level of bilirubin, especially in children, is neurotoxic. In contrast and because of its potent antioxidant effect, moderately elevated bilirubin has been shown to be protective against several adult oxidative-stress-mediated diseases including diabetes mellitus, diabetic nephropathy, cancer and cardiovascular disease.5–9 Moderately higher levels of bilirubin within the normal range have been associated with reduced risk of respiratory disease and all cause mortality.10 Additionally, high serum bilirubin is apparently beneficial for survival in the elderly population.11

Previously, linkage studies identified a serum bilirubin QTL on chromosome 2q (containing the UDP glycosyltransferase 1 family, UGT1A) in European-ancestry populations, a finding that was replicated in American Indians.12–14 Recently, multiple genome-wide association studies (GWAS) identified variants in UGT1A to be associated with serum bilirubin levels in European- and East-Asian-ancestry populations.15–17 The aim of the present study is to investigate if these findings extend to African-ancestry populations, and to identify novel loci by conducting a GWAS and replication study of serum bilirubin in apparently healthy unrelated African Americans. Multiple studies have reported that African-ancestry populations have lower bilirubin levels compared with other ancestral groups.18,19 Understanding the genetic and non-genetic basis of circulating levels of bilirubin is particularly important in populations of African ancestry given the reported higher prevalence of low bilirubin related health conditions (eg, diabetes, the heart and kidney diseases) in African Americans.

MATERIALS AND METHODS

Ethics statement

Ethical approval for the Howard University Family Study (HUFS) was obtained from the Howard University IRB and written informed consent was obtained from each participant.

Study sample

The HUFS is a population-based family study of African Americans in the Washington, D.C. metropolitan area.20 The major objective of the HUFS was to enroll and examine a randomly ascertained sample of African American families (ie, pedigree data collected), along with a set of unrelated individuals (no pedigree data collected), from the Washington, D.C. metropolitan area for the study of the genetic and environmental bases of common complex diseases including hypertension, obesity, diabetes and associated phenotypes. In order
to maximize the utility of this sample for the study of multiple common traits, families were not ascertained based on any phenotypes. To estimate heritability of serum bilirubin, we used 1314 individuals within 328 pedigrees from HUFS. Heritability was estimated using SOLAR.21 For the GWAS, we included the subset of apparently healthy individuals (defined as no type 2 diabetes, no hypertension and no other major morbidty) with bilirubin measurements and Affymetrix 6.0 (Santa Clara, CA, USA) SNP genotype data. The 619 individuals that met these criteria comprised 330 subjects enrolled as individuals and 289 unrelated subjects from families (1 randomly selected from related persons in each pedigree).

**Serum total bilirubin measurement**

Total bilirubin was measured on a Cobas Integra 400 Plus (Roche Diagnostics, Indianapolis, IN, USA) using the Diazo method. Briefly, this method is a colorimetric method based on the reaction between bilirubin and diazotized sulfanilic acid solution to produce a complex known as azobilirubin whose maximum absorbance is pH dependant; to preserve the pH of the reaction an oxalic acid/ sulfanilic acid buffer is used. The color intensity of the reaction is proportional to the concentration of total bilirubin present in the sample and is determined by the increase in absorbance at 552 nm. The minimum detection level for this assay is 0.099 mg/dl. The reproducibility of the assay measured by the coefficient of variation between and within assays is between 0.45 and 0.8%. Serum total bilirubin is reported in mg/dl.

**Genotyping**

DNA was extracted using buffy coat samples following the manufacturer’s instructions using a Gentra PUREGENE DNA Isolation Kit (QIAGEN, Valencia, CA, USA). After sample processing, genome-wide genotyping was performed using the Affymetrix Genome Wide Human SNP Array 6.0.22 Genotypes were called using Birdshead, version 2.2. All samples passed a sample success rate of 95%. SNPs were excluded if they had a success rate of <95% (41 885 SNPs excluded), or had a p-value for the Hardy-Weinberg test of equilibrium <10^-3 (6317 SNPs excluded). The current analysis focused on the 808 465 autosomal SNPs that passed these filters. The sample genotyping rate for this set of SNPs in these individuals was 99.5%. The concordance of blind duplicates was 99.74%. In addition, imputation was performed as previously reported.23 Briefly, we successfully imputed 1 506 100 autosomal SNPs using the YRI reference panel and an additional 52 291 SNPs using the CEU reference panel, for a total of 2 366 856 experimentally determined and imputed SNPs.

**Check for population stratification**

EIGENSTRAT24 was used to detect and correct for population stratification. This method is based on principal components analysis and the resulting correction is specific to a candidate marker’s variation in frequency across ancestral populations, minimizing spurious associations whereas maximizing power to detect true associations. The first two principal components were significant and were included as covariates in the regression analysis.

**Association and replication analyses**

All association and conditional association testing was performed using PLINK.25 Serum bilirubin levels were natural log transformed to follow a normal distribution. Association between an autosomal marker and serum bilirubin was assessed using linear regression assuming an additive genetic model with adjustments for age, sex and first two principal components from EIGENSTRAT. We estimated the genomic inflation factor (λGC) based on the median chi-squared test of all study participants.26 Conditional analyses to examine specific SNP associations were performed using PLINK by specifying the associated SNP as a covariate along with the previously mentioned covariates in the linear regression model.

We attempted to replicate published GWAS of serum bilirubin16–17 in the HUFS samples in two stages: direct (ie, same SNPs as reported) and local (ie, SNPs in reference ancestry linkage disequilibrium (LD) with reported SNPs) replication as previously reported.27 Of the 13 reported SNPs, we were able to directly test 7 in this study; the remaining 6 were either not present on the Affymetrix 6.0 chip, were not successfully imputed or failed our quality control filters. Local replication was performed for SNPs not directly replicated. The SNPs in LD at a r^2 ≥ 0.3 with discovery variants were queried. HapMap CEU reference data for European-ancestry populations and the HapMap CHB data for the East Asian-ancestry populations were used in the local replication analysis. To adequately account for multiple testing, we estimated the effective degrees of freedom for the spectrally decomposed covariance matrix for the block of SNPs using the HUFS genotyped SNP data as previously described.27,28

Power calculations were done using the Quanto software package.29 We used a mean (SD) of serum bilirubin=0.58 (0.31) mg/dl, MAF values obtained from HUFS genotype dataset, and assumed an additive genetic model with α=3×10^-8 in two tail tests to calculate study power.

**RESULTS**

Clinical characteristics for the 619 participants included in this study are presented in Table 1. Consistent with the national data from NHANES III, male participants had significantly higher levels of bilirubin than the females.30 In contrast, female participants were heavier and had higher levels of insulin, and male participants had higher systolic blood pressure, fasting glucose and alanine amino transferase levels. Only a small proportion (3.8%) of the HUFS participants had elevated (ie, >1.3 mg/dl) serum total bilirubin level. The heritability (h^2) of serum total bilirubin in HUFS is 0.42 (SE 0.064) with a P-value of 6.63×10^-13.

The distribution of genome-wide association P-values for total serum bilirubin is displayed in a Manhattan plot (Supplementary Figure 1) and in a quantile-quantile plot (Supplementary Figure 2). The value of genomic control inflation factor (λGC) was 1.008, which indicated that inflation of association test statistics because of population stratification was negligible. Out of the 100 top scoring SNPs 39 (Supplementary Table S1) were located within a 78 kb region in the UGT1A1 gene on 2q37; all 39 SNPs displayed P-values lower than the predetermined genome wide threshold of 5×10^-8 (Supplementary Figure 3 and Supplementary Table S2). The lowest P-value corresponded to rs887829 (1.97×10^-22). After conditioning on the most significant SNP (rs887829), none of the remaining 38 SNPs in the UGT1A1 gene maintained genome-wide significance (Supplementary Table S2). This finding is consistent with the observation that the 38 SNPs are in moderate to strong LD with rs887829 (r^2 ranged from 0.28 to 1.0 in this study). Of note, previous studies identified rs887829 as a modifier of bilirubin levels with each copy of the T allele resulting in an average increase of 1.77 mg/dl in serum bilirubin in Europeans,17,31 the corresponding estimate of effect size in this study of African Americans is 1.25 mg/dl per copy of the T allele. SNP rs887829 explained 12.4% of the variance in serum total bilirubin in this study of African Americans.

### Table 1 Clinical characteristics of African American participants included in this study

|                          | Male    | Female   | P-values |
|--------------------------|---------|----------|----------|
| N (%)                    | 251 (40.55%) | 368 (59.45%) |          |
| Age (years)              | 40.74±11.54 | 41.19±11.83 | 0.641649 |
| Height (cm)              | 176.44±6.99 | 163.76±7.18 | 1.75E-78 |
| Body mass index (Kg/m²)  | 27.35±6.14  | 30.64±8.37  | 2.76E-08 |
| Systolic B.P. (mmHg)     | 118.41±10.48 | 116.51±10.77 | 0.028991 |
| Diastolic B.P. (mmHg)    | 75.26±8.69  | 74.31±7.5    | 0.160415 |
| Fasting glucose (mg/dl)  | 85.29±11.08 | 82.99±9.14  | 0.006826 |
| Insulin (uU/ml)          | 9.18±8.65   | 11.24±18.04 | 0.061149 |
| Bilirubin (mg/dl)        | 0.7±0.34     | 0.5±0.28    | 7.01E-14 |
| Alanine AMINO Transferase (U/L) | 15.43±18.48 | 10.35±11.11 | 0.000129 |
We directly replicated the top scoring SNP (rs11891311, P-value = 4.78 × 10^{−150}) reported in the Korean GWAS.\textsuperscript{15} This SNP was also reported in the GWAS conducted in African-ancestry populations;\textsuperscript{16–17} however, SNP rs11891311 is in moderate LD (r^2 > 0.66) with the top scoring UGT1A1 SNP (rs887829) in European-ancestry populations (Table 2). This is in sharp contrast to the weaker LD (0.26 in HapMap YRI and 0.31 in HUFS) observed between these SNPs in African-ancestry populations (Supplementary Table S3, Supplementary Figure 4). Despite the weaker LD in our African-ancestry population, we were able to directly replicate all three SNPs in African Americans, perhaps indicating the robustness of previous findings.

As described in the materials and methods section, we also conducted local replication analyses of SNPs in LD (r^2 ≥ 0.3) with the reported SNPs in European- and East-Asian-ancestry populations. This approach was implemented when the reported SNP was not genotyped or successfully imputed. This approach yielded additional significant associations after Bonferroni correction for multiple testing. For example, SNP rs4148323 in the UGT1A1 gene with a MAF of 0.21 in East Asians was successfully replicated by SNPs in LD with it using the local approach despite being monomorphic in CEU, YRI and African Americans (Table 2, Figure 1). Notably, we successfully replicated all four previously reported UGT1A1 variants from GWAS of serum bilirubin in European- or East-Asian-ancestry populations using either the direct or local replication approaches in our African American study (Table 2).

In addition, we replicated reported association between variants in the semaphorin 3C (SEMA3C) gene and bilirubin levels.\textsuperscript{16} The discovery SNP (rs4236644) located in the SEMA3C gene was locally replicated by two LD based SNPs (rs1358503 and rs10251680, P-values 0.0461 and 0.0447, respectively, Figure 2). Each SNP explained ~1% of the variance in serum bilirubin levels in this study of African Americans.

### DISCUSSION

The present study has confirmed the important role of the UGT1A1 gene in serum total bilirubin levels in an African-ancestry population. In this sample of African Americans, the most significantly associated variant (rs887829) is located in the core promoter region of the UGT1A1 gene. Specifically, rs887829 is 211 bp upstream of the short thymine adenine (TA)_n repeat sequence covering the TATA box of the UGT1A1 gene. The same SNP was previously reported to be associated with bilirubin levels in populations of East Asians and Europeans.\textsuperscript{15–17} Of note, we also replicated the most significant SNP (rs11891311) reported in East Asians and Europeans.\textsuperscript{15–17} This SNP is 28 kb upstream from and is in moderate LD (r^2 = 0.31) with rs887829 in African Americans. However, the two SNPs show stronger LD (r^2 = 0.73) in East Asians and Europeans (r^2 = 0.66) as estimated using the HapMap CHB and CEU reference data (Supplementary Table S3 and Supplementary Figure 4). These differences in LD and greater association of rs887829 in HUFS (Table 2) taken together with its functional proximity (ie, near promoter elements) support the likelihood that rs887829 is a more refined proxy for the causative candidate SNP influencing serum total bilirubin levels. Our ability to refine this GWAS signal is because of the lower LD and shorter haplotypes present in African-ancestry populations compared with European- and Asian-ancestry populations; We note that we have successfully refined GWAS signals for fasting plasma glucose and serum uric acid using African ancestry samples in previous studies.\textsuperscript{27,32}

Interestingly, rs887829 is in strong LD (r^2 ≥ 0.74) with rs10929302, which was associated with serum total bilirubin with a P-value of 1.37 × 10^{-11} in the present study. SNP rs10929302 is located in the phenobarbital response enhancer module and known to be associated with irinotecan response.\textsuperscript{23} This implies that this small (~ 3 kb region) UGT1A1 region is pleiotropic, containing variants that influence: (1) serum bilirubin levels, (2) phenobarbital induction of UGT1A1 expression and (3) irinotecan response.
Two families of uridine-diphosphate glucuronosyltransferase (UGT) enzymes are essential to the proper metabolism of a number of exogenous and endogenous compounds. The UGT1 family includes UGT1A1, which has a critical role in the detoxification of neurotoxic bilirubin. Reduced UGT1A1 activity is associated with unconjugated hyperbilirubinemia and loss of function has been observed in rare inherited disorders (eg, Gilbert’s syndrome and Crigler–Najjar syndrome). Recently, Horsfall et al reported that variation in the short (TA)n repeat sequence in the UGT1A1 gene promoter is associated with hyperbilirubinaemia. There is also considerable continental and ethnic variation in the distribution of the (TA)n repeat in the UGT1A1 gene.19,34 For example, there are more (TA)n alleles in African populations compared with European-ancestry populations. Alleles (TA)5, (TA)6, (TA)7 and (TA)8 have all been observed in most African populations studied to date; in contrast, alleles (TA)5 and (TA)6 have not been observed in Europeans.19,34 Also, large difference in allele frequency has been observed across African ethnic groups. The low affinity alleles (TA)7 and (TA)8 are highly prevalent in African populations around the equator.19

Our study emphasizes the importance of conducting GWAS in populations with ancestry from different parts of the world. For example, rs4148323 in UGT1A1 was reported in East Asian ancestry with a P-value of $1.22 \times 10^{-82}$. In European (CEU), West African (YRI) and African American (HapMap) samples, this variant is monomorphic compared with a MAF of 0.211 in East Asian (CHB and JPT) HapMap data. However, we were able to replicate this finding by querying a 500 kb-window centered on rs4148323 and performing association analysis on all 61 SNPs in LD using a cutoff $r^2 \geq 0.3$.

GWAS of serum bilirubin in European-ancestry populations also reported significant association with SNP rs4236644 located in the SEMA3C gene; this gene has been linked to increased expression of severe steatotic livers.35 After Bonferroni correction, local replication using rs4236644 as the query SNP, we observed significant association between SNPs (rs1358503 and rs10251660) and bilirubin level (P-values = 0.0461 and 0.0447, respectively). In African-ancestry populations, the locally replicated SNPs have higher $r^2$ values with the discovery SNP when compared with European ancestry (Figure 3). Using HapMap CHB data or JPT data, LD is even weaker, which may explain why this SNP was not observed in the GWAS of bilirubin in Koreans.

In summary, we showed that UGT1A1 is a major locus influencing serum total bilirubin levels in African Americans. Our findings may also contribute to the understanding of the etiology and the treatment of hyperbilirubinaemia in African-ancestry populations. Taking advantage of the lower LD and shorter haplotypes present in African-ancestry populations, we refined previously reported association of...
rs887829 with serum bilirubin thereby demonstrating the utility of using ethnically diverse populations for replication studies.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
The study was supported by grants S06GM008016-320107 to CR and S06GM008016-380111 to AA, both from the National Institute of General Medical Sciences MBRS/SCORE Program. Participant enrollment was carried out at the Howard University General Clinical Research Center supported by National Center for Research Resources grant 2M01RR010284 at the National Institutes of Health (NIH). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official view of NIH or the Department of Health and Human Services. This research was supported in part by the Intramural Research Program of the Center for Research on Genomics and Global Health, which is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for Information Technology and the Office of the Director at the NIH (Z01HG200362). Genotyping support was provided by the Coriell Institute for Medical Research.

1 Arias IM: The excretion of conjugated bilirubin by the liver cell. Medicine (Baltimore) 1966; 45: 513–515.
2 McDonagh AF: Bilirubin toxicity to human erythrocytes: a more sanguine view. Pediatrics 2007; 120: 175–178.
