HAL
open science

Exome-Wide Association Study Identifies New Low-Frequency and Rare UGT1A1 Coding Variants and UGT1A6 Coding Variants Influencing Serum Bilirubin in Elderly Subjects

Abderrahim Oussalah, Paolo Bosco, Guido Anello, Rosario Spada, Rosa-Maria Gueant-Rodriguez, Celine Chery, Pierre Rouyer, Thomas Josse, Antonino Romano, Maurizzio Elia, et al.

To cite this version:

Abderrahim Oussalah, Paolo Bosco, Guido Anello, Rosario Spada, Rosa-Maria Gueant-Rodriguez, et al.. Exome-Wide Association Study Identifies New Low-Frequency and Rare UGT1A1 Coding Variants and UGT1A6 Coding Variants Influencing Serum Bilirubin in Elderly Subjects. Medicine, 2015, 94 (22), pp.e925. 10.1097/MD.0000000000000925 . hal-01679021

HAL Id: hal-01679021

https://hal.univ-lorraine.fr/hal-01679021v1

Submitted on 9 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

OPEN

Exome-Wide Association Study Identifies New Low-Frequency and Rare *UGT1A1* Coding Variants and *UGT1A6* Coding Variants Influencing Serum Bilirubin in Elderly Subjects

A Strobe Compliant Article

Abderrahim Oussalah, MD, PhD, Paolo Bosco, MD, PhD, Guido Anello, PhD, Rosario Spada, MD, Rosa-Maria Guéant-Rodriguez, MD, PhD, Céline Chery, PhD, Pierre Rouyer, BS, Thomas Josse, MS, Antonino Romano, MD, PhD, Maurizzio Elia, MD, PhD, Jean-Pierre Bronowicki, MD, PhD, and Jean-Louis Guéant, MD, DSc, AGAF

Abstract: Genome-wide association studies (GWASs) have identified loci contributing to total serum bilirubin level. However, no exome-wide approaches have been performed to address this question. Using exome-wide approach, we assessed the influence of protein-coding variants on unconjugated, conjugated, and total serum bilirubin levels in a well-characterized cohort of 773 ambulatory elderly subjects from Italy. Coding variants were replicated in 227 elderly subjects from the same area. We identified 4 missense rare (minor allele frequency, MAF $< 0.5\%$) and low-frequency (MAF, 0.5%–5%) coding variants located in the first exon of the *UGT1A1* gene, which encodes for the substrate-binding domain (rs4148323 [MAF = 0.06%; p.Gly71Arg], rs144398951 [MAF = 0.06%; p.Ile215Val], rs35003977 [MAF = 0.78%; p.Val225Gly], and rs57307513 [MAF = 0.06%; p.Ser250Pro]). These variants were in strong linkage disequilibrium with 3 intronic *UGT1A1* variants (rs887829, rs4148325, rs6742078), which were significantly associated with total bilirubin level ($P = 2.34 \times 10^{-34}$, $P = 7.02 \times 10^{-34}$, and $P = 8.27 \times 10^{-34}$), as well as unconjugated, and conjugated bilirubin levels. We also identified *UGT1A6* variants in association with total (rs6759892, p.Ser7Ala, $P = 1.98 \times 10^{-26}$; rs2070959, p.Thr181Ala, $P = 2.87 \times 10^{-27}$; and rs1105879, p.Arg184Ser, $P = 3.27 \times 10^{-29}$), unconjugated, and conjugated bilirubin levels. All *UGT1A1* intronic variants (rs887829, rs6742078, and rs4148325) and *UGT1A6* coding variants (rs6759892, rs2070959, and rs1105879) were significantly associated with gallstone-related cholecystectomy risk. The *UGT1A6* variant rs2070959 (p.Thr181Ala) was associated with the highest risk of gallstone–related cholecystectomy (OR, 4.58; 95% CI, 1.58–13.28; $P = 3.21 \times 10^{-3}$). Using an exome-wide approach we identified coding variants on *UGT1A1* and *UGT1A6* genes in association with serum bilirubin level and hyperbilirubinemia risk in elderly subjects. *UGT1A1* intronic single-nucleotide polymorphisms (SNPs) (rs6742078, rs887829, rs4148324) serve as proxy markers for the low-frequency and rare *UGT1A1* variants, thereby providing mechanistic explanation to the relationship between *UGT1A1* intronic SNPs and the UGT1A1 enzyme activity. *UGT1A1* and *UGT1A6* variants might be potentially associated with gallstone-related cholecystectomy risk.

(*Medicine* 94(22):e925)

Abbreviations: 3D = three-dimensional, 95% CI = 95% confidence interval, EM = expectation-maximization, GWAS = genome-wide association study, HWE = Hardy–Weinberg equilibrium, IBD = identity by descent, LD = linkage disequilibrium, MAF = minor allele frequency, MRP2 = multidrug resistance-associated protein 2, OR = odds ratio, PCA = principal-component analysis, SNP = single nucleotide polymorphism, UGT1A1 = UDP-glucuronosyltransferase 1 family, polypeptide A1, UGT1A6 = UDP-glucuronosyltransferase 1 family, polypeptide A6.

Bilirubin is the major metabolite of heme, the iron-binding tetrapyrrole ring found in hemoglobin, myoglobin, and cytochromes.[1] The straight-chain compound biliverdin is produced through the oxidation of heme porphyrin ring by microsomal heme oxygenase. Biliverdine reductase, a nicotinamide adenine dinucleotide phosphate (NADPH)-dependent enzyme, reduces biliverdin to produce bilirubin.[1] After being captured by the hepatocyte through its membrane surface in contact with the sinusoids, bilirubin is transported to the smooth endoplasmic reticulum and becomes the substrate of the UDP-glucuronosyltransferase 1 family, polypeptide A1 enzyme (UGT1A1), which catalyzes the esterification of the propionic acid side chains of bilirubin with glucuronic acid (present as uridine diphosphoglucuronic acid) to form mainly the diglucuronide conjugate, a water-soluble conjugated molecule.[1] Conjugated bilirubin is then actively secreted into the bile canaliculi by a membrane

Editor: Chaeyoung Lee.
Received: October 20, 2014; revised: May 4, 2015; accepted: May 6, 2015.
From the Inserm, NGERE – Nutrition, Genetics, and Environmental Risk Exposure (AO, R-MG-R, CC, PR, J-PB, J-LG); Faculty of Medicine of Nancy, University of Lorraine (AO, R-MG-R, CC, J-PB, J-LG); University Hospital of Nancy, Department of Molecular Medicine and Personalized Therapeutics, Department of Biochemistry, Molecular Biology, Nutrition, and Metabolism (AO, R-MG-R, CC, TJ, J-LG); Reference Centre for Inherited Metabolic Diseases (ORPHA67872), Vandoeuvre-lès-Nancy, France (AO, R-MG-R, CC, TJ, J-LG); IRCCS, Oasi Maria SS–Institute for Research on Mental Retardation, Troina (PB, GA, RS, AR, ME); Department of Internal Medicine and Geriatrics, UCSC, CI Columbus, Roma, Italy (AR); and Department of Gastroenterology and Hepatology, University Hospital of Nancy, Vandoeuvre-lès-Nancy, France (J-PB).
Correspondence: Jean-Louis Guéant, INSERM U954, NGERE – Nutrition, Genetics, and Environmental Risk Exposure, U954, Vandoeuvre-lès-Nancy F-54511, France (e-mail: jean-louis.gueant@univ-lorraine.fr.).
The study was supported by grants from the Region of Sicily, the French National Institute for Health and Medical Research (INSERM) and the French National Agency for Research (ANR, Nutrivigène program)
The authors declare no conflicts of interest.
Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.
This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0, where it is permissible to download, share and reproduce the work in any medium, provided it is properly cited. The work cannot be changed in any way or used commercially.
ISSN: 0025-7974
DOI: 10.1097/MD.0000000000000925

ATP-dependent transporter, designated as multidrug resistance-associated protein 2 (MRP2).[2]

It is now well established that unconjugated hyperbilirubinemia is a risk factor for gallstones.[3,4] In patients with sickle cell disease, it has been shown that genetic variation in the promoter of *UGT1A1* may be a risk factor for symptomatic gallstones in older people.[5] However, no studies are available in the healthy population, especially among older subjects.

Genome-wide association studies (GWASs) systematically evaluate common genetic variants, typically with a minor allele frequency (MAF) >5% and have been extensively used to dissect the genetic architecture of complex diseases and quantitative traits.[6] The high number of genetic variations identified in GWASs contrasts with their low effect on disease risk or quantitative trait variation. Thus, GWASs generally fail to translate into functional understanding or clinical practice. The ''missing heritability'' observed in GWASs could be explained by the fact that they do not assess low-frequency (MAF, 0.5%–5%) and rare (MAF < 0.5%) genetic variants that play a major role in human pathology.[6] Recent evidence suggest that low-frequency and rare variants are associated with complex diseases.[6–8]

It is estimated that the protein-coding regions of the human genome constitute about 85% of the disease-causing mutations.[9] The Illumina Human Exome BeadChip provides coverage of functional exonic variants. Around 250,00 markers on this BeadChip represent SNPs in *RefSeq* genes, non-synonymous SNPs, and SNPs in coding regions (including untranslated regions, UTRs). It has been demonstrated that exome-wide genotyping identified additional medically actionable variant calls and helped resolve ambiguous single-nucleotide variants in comparison with genome-wide approaches.[10]

To date, there are no available data regarding the exome-wide association study approach that evaluated potential-associated variants with unconjugated, conjugated, and total serum bilirubin levels in physiological conditions. To determine the role of rare and low-frequency coding variants in traits reflecting unconjugated, conjugated, and total serum bilirubin level, we evaluated putative functional coding variants using the Illumina HumanExome BeadChip on a well-characterized cohort of ambulatory elderly subjects from Italy.[11–13] Furthermore, we assessed the potential influence of bilirubin-related variants on hyperbilirubinemia risk and the risk of gallstone-related cholecystectomy.

METHODS

Study Populations

Between January 1, 1999 and December 31, 2006, ambulatory subjects were included from 3 communes from Sicily, Italy, in the same Province of Enna: Troina, Cerami, and Gagliano Castelferrato (the Oasi cohort Sicily) (Supplemental Digital Content: Figure 1, http://links.lww.com/MD/A287).[13] Institutional review board approval was obtained from the ethical committees of the Medical Centre of Troina (*Instituto di Ricovero e Cura a Carattere Scientifico*, the Oasi Maria Institute, Institute for Research on Mental Retardation, Troina, Sicily, Italy) and the University Hospital of Nancy (Vandoeuvre-lès-Nancy, France).[13] All subjects involved in the study gave valid informed consent in accordance with approval from the relevant local ethical committees or institutional review boards and received a systematic clinical evaluation to rule out cancer, cardiovascular, renal, hepatic, or genetic disease, and any vitamin supplementation before inclusion in the study.

Blood Sampling

Phlebotomy was performed and fasting venous blood was collected in ethylenediaminetetraacetic acid-containing tubes. Samples were centrifuged immediately at 2500*g* for 15 minutes at room temperature. Aliquots were stored at −70°C until analysis.

Bilirubin Assays

All specimens were tested in a single laboratory (Vandoeuvre-lès-Nancy, France). Quantitative estimation of total and conjugated fraction of serum bilirubin was assayed by the Synermed (Rosny-Sous-Bois, France) reagent using the Olympus AU2700 chemistry-immuno analyzer (Rungis, France) according to manufacturer's protocol. Direct bilirubin reagent measures the absorbance of an oxidation product of bilirubin at 700 nm. The absorbance of the cholecyanin oxidation product is directly proportional to bilirubin concentration. Conjugated bilirubin is oxidized in the presence of the electron transfer agent, KHC3, to a blue-green infrared light absorbing chromophore. Unconjugated bilirubin is not significantly detected due to the fact that it is not soluble under the conditions of the direct reaction mixture and remains bound to albumin. Serum total bilirubin is reported in mg/dL. The intra-assay coefficients of variation for total bilirubin were 0.00% and 0.99% for level 1 and level 2, respectively. The corresponding figures for conjugated bilirubin were 0.00% and 1.07%, respectively.

Genotyping and Quality Controls

Genomic DNA was extracted from the buffy coat and genotyping was performed at the National Institute of Health and Medical Research Unit-954 (Vandoeuvre-lès-Nancy, Lorraine, France) using the Illumina Infinium HumanExome BeadChip v1.2 (Illumina Inc, San Diego, CA) according to the manufacturer's recommendations. The Illumina HumanExome Beadchip includes 247,870 markers focused on protein-altering variants selected from >12,000 exome and genome sequences representing multiple ethnicities and complex traits.[14] Nonsynonymous genetic variants had to be observed ≥3 times in at least 2 studies, and splicing and stop-altering variants had to be observed ≥2 times in at least 2 studies.[14] Additional, array content includes variants associated with complex traits in previous GWAS, HLA tags, ancestry-informative markers, markers for identity-by-descent estimation, and random synonymous SNPs.[14] Data were analyzed using Illumina's GenomeStudio Genotyping Module.

Bead intensity data were processed and normalized for each sample in GenomeStudio. The Genome Reference Consortium GRCh37 (hg19; Feb. 2009) map was used (Illumina manifest file Immuno_BeadChip_11419691_B.bpm), and normalized probe intensities were extracted for all samples that passed standard laboratory quality-control thresholds. Genotype calling was performed with the GenomeStudio GenTrain 2.0 algorithm (Illumina's GenomeStudio data analysis software; Illumina Inc.).

Sample quality control measures included sample call rate, overall heterozygosity, relatedness testing. All patients included in this study also had a genotyping success rate >90%. Duplicated samples and those showing cryptic relatedness were assessed by calculating identity by descent (IBD). We estimated IBD sharing and carried out PCA on an linkage disequilibrium (LD)-pruned set of 28,509 autosomal variants obtained by removing large-scale high-LD regions,[15,16] variants with a call rate <0.99, variants with a minor allele frequency <5% or variants with Hardy-Weinberg equilibrium (HWE) *P*-value $<10^{-4}$. By estimating the probability of sharing zero, 1, or 2

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

alleles for any 2 individuals as P(Z = 0), P(Z = 1), and P(Z = 2), respectively, a proportion of IBD was calculated using the following formula: PI-HAT = PI = P(Z = 1)/2 + P(Z = 2). Eleven pairs of individuals with a relatedness measure (PI-HAT) value >0.99 were considered to be a duplicate sample. From these pairs, one of the individuals was randomly removed from the data. A PI-HAT threshold >0.2 was used to suggest cryptic relatedness. Three samples had high positive inbreeding coefficients *f* suggesting truly inbred samples and were removed before analysis. Principal-component analysis (PCA) was performed on the samples merged with 7 HapMap phase 3 populations (CEU, TSI, YRI, MEX, JPT, CHD and CHB) as reference populations to identify ancestry outliers. Inspecting the first 10 principal components, we identified 3 population outliers that were excluded from subsequent analyses (Supplemental Digital Content: Figure 2, http://links.lww.com/MD/A287). All genotyping data sets underwent the same rigorous quality checks before cases and controls were compared. Genetic variants were removed from the primary analysis if they had a call rate <95%, a significant differential missingness in cases and controls ($P < 0.05$), a significant departure from Hardy–Weinberg equilibrium (exact HWE $P < 10^{-4}$), or a minor allele frequency (MAF) <0.05%.

Replication Study

The replication study was performed in an independent population of 227 Italian elderly subjects originating from Gagliano Castelferrato commune (Sicily, Enna). Genomic DNA was isolated from peripheral leukocytes using Qiagen kit (Qiagen-France, Courtaboeuf cedex, France). Determination of the *UGT1A6* gene polymorphisms rs6759892 (p.Ser7Ala), rs2070959 (p.Thr181Ala), and rs1105879 (p.Arg184Ser) were performed by LightCycler 480 instrument (Roche Molecular Biochemicals, Lyon, France) using commercial LightSNiP assays (SimpleProbe, TIB MolBiol, Berlin, Germany) according to the manufacturer's recommendations. Samples were set up in a final volume of 6 μL, containing 10 ng of DNA, 0.6 μL of LightCycler FastStart DNA MasterPLUS HybProbe Mix (Roche Molecular Biochemicals, Lyon, France), 3 mmol/L of $MgCl_2$, and 0.3 μL of LightSNiP reagent mix (Tib-MolBiol, Berlin, Germany). The cycling conditions were as follows: initial denaturation at 95°C for 10 min, followed by 45 cycles of denaturation at 95°C for 10 s, annealing at 60°C for 10 s, and extension at 72°C for 15 s. Fluorescence was monitored at the end of each annealing phase at 60°C. After completion of the PCR, a melting curve of the amplification products was plotted by denaturation at 95°C for 30 s, 40°C for 2 min, and then slowly heating the sample to 75°C with a ramp rate of 2 °C/s and continuous fluorescence acquisition.

Statistical Analysis

All quantitative variables are described as medians and percentiles (Interquartile range [IQR], 25–75th percentile). All proportions are expressed as percentages with 95% confidence intervals (95% CI). In the exome-wide stage, the association between Log transformed total, unconjugated, and conjugated bilirubin and genetic variants was determined by applying linear regression analysis and correlation/trend test using an allelic model with Bonferroni adjustment for multiple tests. The correlation/trend test tests the significance of any correlation between 2 numeric variables (or 2 variables that have been encoded as numeric variables). This test may also be thought of as any "trend," which either one of the numeric variables may have taken against the other one. The association between binary outcomes and genetic variants was determined by applying logistic regression analysis. We visually inspected the cluster plots for the most associated genetic variants to confirm genotyping quality. Linkage disequilibrium pairwise analysis was performed using a matrix output for both the expectation-maximization (EM) algorithm and composite haplotype method and both R2 and D' values. The comparison of bilirubin values across genotypes subgroups was performed using the Kruskal–Wallis test. All quality controls and genotypic analyses were performed using SNP & Variation Suite (SVS) 7.7.8 (Golden Helix Inc, Bozeman, MT).

Gene-Based Analysis

We performed gene-based tests to further investigate the influence of rare and low-frequency variants on serum bilirubin level. Gene-based tests offer an alternative to single-variant tests, which are often underpowered to detect association with rare variants.[14,17] Accordingly, exome-wide data were analyzed using the combined multivariate and collapsing (CMC) method with regression analysis as described by Li and Leal.[17] The CMC approach uses regions such as genes. The difference is that within each region, CMC first bins variants according to a criterion such as minor allele frequency, then collapses the variants within each bin, and finally performs multivariate testing on the counts across the various bins.[17] We collapsed variants with MAF <0.01 and variants with larger MAFs were investigated separately for each gene. In total, we tested 14,87 genes using the CMC method.

Model Building and Structural-Based Analysis

Three-dimensional (3D) modeling of the human wild-type and variant proteins was performed using SWISS-MODEL, an automated homology modeling program.[18,19] The homology-modeling server SWISS-MODEL was accessed at http://swiss-model.expasy.org/.[18,19] The 3D models were visualized by the DeepView/Swiss-PdbViewer 4.1.0 and OpenAstexViewer 3.0 softwares.[20,21] For each 3D model of the native and variant proteins, narrow grooves, hydrophobic patches, and electrostatic potential on charged residues using the Coulomb method were studied using the DeepView/Swiss-PdbViewer 4.1.0 software.[21]

RESULTS

In the initial study, we successfully genotyped 400 subjects originating from the Troina commune (Sicily, Enna) in the population-based OASI study for 242,901 variants on the Illumina HumanExome Beadchip and performed in silico replication study on 373 subjects originating from the Cerami commune (Sicily, Enna). Clinical characteristics of the 773 participants are summarized in Supplemental Digital Content: Table 1, http://links.lww.com/MD/A287. A total of 239,059 variants passed quality controls and were tested for association with unconjugated, conjugated and total serum bilirubin level, assuming additive allelic effects.

Genetic Variants Associated With Total Serum Bilirubin Level

Exome-Wide Analysis in the Initial Study (n = 400, Troina, Sicily, Enna)

The Exome-wide association study established a significant association between 6 genetic variants on the chromosome

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

2 locus 2q37.1 and total serum bilirubin level. Among these 6 variants, 3 are located on the *UGT1A1* gene (UDP glucuronosyltransferase 1 family, polypeptide A1; Entrez ID: 54658) and 3 are located in the *UGT1A6* gene (UDP glucuronosyltransferase 1 family, polypeptide A6; Entrez ID: 54578) (Table 1 and Figures 1 and 2). Functional annotation and quality controls of the *UGT1A1* and *UGT1A6* variants associated with serum total bilirubin level are reported in Supplemental Digital Content: Table S2, http://links.lww.com/MD/A287. The 3 variants in the *UGT1A1* gene (rs887829, rs4148325, rs6742078) are part of the same haplotype block (Supplemental Digital Content: Table 3, http://links.lww.com/MD/A287 and Figure 3). The 3 variants on *UGT1A6* are non-synonymous (rs6759892, c.19T>G, p.Ser7Ala, $P = 8.34 \times 10^{-11}$; rs2070959, c.541A>G, p.Thr 181Ala, $P = 1.36 \times 10^{-11}$; and rs1105879, c.552A > C, p.Arg1 84Ser, $P = 1.54 \times 10^{-11}$) and are part of the same haplotype block on the first exon of *UGT1A6* (Supplemental Digital Content: Table and Figures 4 and 5, http://links.lww.com/MD/A287).

In Silico Replication Study (n = 373, Cerami, Sicily, Enna)

In the in silico replication study, the 6variants that reached exome-wide significance for the association with serum total bilirubin level in the initial study were significantly associated with total serum bilirubin level (Table 1). The pooled analysis of initial and in silico replication studies reported the same figures for the 6 variants on *UGT1A1* and *UGT1A6* genes with the following *P* values for the 3 non-synonymous variants on *UGT1A6*: (rs6759892, $P = 1.98 \times 10^{-26}$; rs2070959, $P = 2.87 \times 10^{-27}$; and rs1105879, $P = 3.27 \times 10^{-29}$) (Table 1). Results were similar after adjusting regression analysis for age and sex (Supplemental Digital Content: Table 4, http://links.lww.com/MD/A287). Using the pooled data from both initial and in silico replication studies, for each of the 6 variants, the median total serum bilirubin values differed significantly between the 3 subgroups of genotypes (homozygous major, heterozygous, and homozygous minor) on the Kruskal–Wallis test after Bonferroni correction with higher total serum bilirubin level observed in subjects with homozygous minor genotype (Supplemental Digital Content: Table 5 and Figure 6, http://links.lww.com/MD/A287).

Genetic Variants Associated With Unconjugated Serum Bilirubin Level

Exome-Wide Analysis in the Initial Study (n = 400, Troina, Sicily, Enna)

Consistently with the exome-wide analysis on total serum bilirubin level, the same top variants on *UGT1A1* and *UGT1A6* genes were significantly associated with serum unconjugated bilirubin level (Supplemental Digital Content: Table 6 and Figure 7, http://links.lww.com/MD/A287) with the following *P* values for the non-synonymous variants on *UGTA16*: (rs6759892, $P = 4.75 \times 10^{-11}$; rs2070959, $P = 9.17 \times 10^{-12}$; and rs1105879, $P = 9.02 \times 10^{-12}$).

In Silico Replication Study (n = 373, Cerami, Sicily, Enna)

In the in silico replication study, the same 6 variants that reached exome-wide significance for the association with serum unconjugated bilirubin level in the initial study were also significantly associated with unconjugated serum bilirubin level (Supplemental Digital Content: Table 6, http://links.lww.com/MD/A287). The pooled analysis of initial and in silico replication studies reported the same figures for the 6 variants with the following *P* values: (rs6759892, $P = 4.19 \times 10^{-23}$; rs2070959,

TABLE 1. Genetic Variants Associated at the Exome-wide Level With Serum Total Bilirubin Level in the Troina Cohort Subjects

Marker	MAF	Regression *P*	Regression Bonf. *P*	Regression Slope	Regression Slope SE	Corr/ Trend *P*	Corr/Trend Bonf. *P*	Corr/ Trend R
Initial study (n = 400)								
rs6759892	0.363	8.34×10^{-11}	4.59×10^{-6}	0.0923	0.0138	2.34×10^{-10}	1.29×10^{-5}	0.3177
rs2070959	0.327	1.36×10^{-11}	7.46×10^{-7}	0.0982	0.0141	4.56×10^{-11}	2.51×10^{-6}	0.3301
rs1105879	0.359	1.54×10^{-11}	8.49×10^{-7}	0.0955	0.0138	5.15×10^{-11}	2.84×10^{-6}	0.3300
rs887829	0.328	5.62×10^{-13}	3.09×10^{-8}	0.1027	0.0138	2.70×10^{-12}	1.49×10^{-7}	0.3505
rs6742078	0.331	1.53×10^{-12}	8.41×10^{-8}	0.1003	0.0137	6.51×10^{-12}	3.58×10^{-7}	0.3438
rs4148325	0.331	1.53×10^{-12}	8.41×10^{-8}	0.1003	0.0137	6.51×10^{-12}	3.58×10^{-7}	0.3438
In silico replication study (n = 373)								
rs6759892	0.383	1.44×10^{-18}	1.01×10^{-17}	0.1296	0.0140	5.69×10^{-17}	3.98×10^{-16}	0.4340
rs2070959	0.345	2.25×10^{-18}	1.58×10^{-17}	0.1306	0.0142	8.21×10^{-17}	5.75×10^{-16}	0.4318
rs1105879	0.370	1.01×10^{-20}	7.07×10^{-20}	0.1381	0.0139	1.15×10^{-18}	8.07×10^{-18}	0.4622
rs887829	0.341	3.01×10^{-25}	2.11×10^{-24}	0.1558	0.0139	3.34×10^{-22}	2.33×10^{-21}	0.5024
rs6742078	0.340	3.33×10^{-25}	2.33×10^{-24}	0.1551	0.0139	3.66×10^{-22}	2.56×10^{-21}	0.5026
rs4148325	0.341	3.01×10^{-25}	2.11×10^{-24}	0.1558	0.0139	3.34×10^{-22}	2.33×10^{-21}	0.5024
Pooled analysis from initial and replication studies (n = 773)								
rs6759892	0.373	1.98×10^{-26}	1.38×10^{-25}	0.1095	0.0099	9.73×10^{-25}	6.81×10^{-24}	0.3698
rs2070959	0.336	2.87×10^{-27}	2.01×10^{-26}	0.1133	0.0101	1.84×10^{-25}	1.29×10^{-24}	0.3756
rs1105879	0.364	3.27×10^{-29}	2.29×10^{-28}	0.1157	0.0099	4.24×10^{-27}	2.97×10^{-26}	0.3908
rs887829	0.334	2.34×10^{-34}	1.64×10^{-33}	0.1273	0.0099	1.94×10^{-31}	1.36×10^{-30}	0.4201
rs6742078	0.336	8.27×10^{-34}	5.79×10^{-33}	0.1259	0.0099	5.51×10^{-31}	3.85×10^{-30}	0.4169
rs4148325	0.336	7.02×10^{-34}	4.91×10^{-33}	0.1262	0.0099	4.76×10^{-31}	3.34×10^{-30}	0.4170

Bonf = Bonferroni, Corr = Correlation, MAF = minor allele frequency, SE = Standard error.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

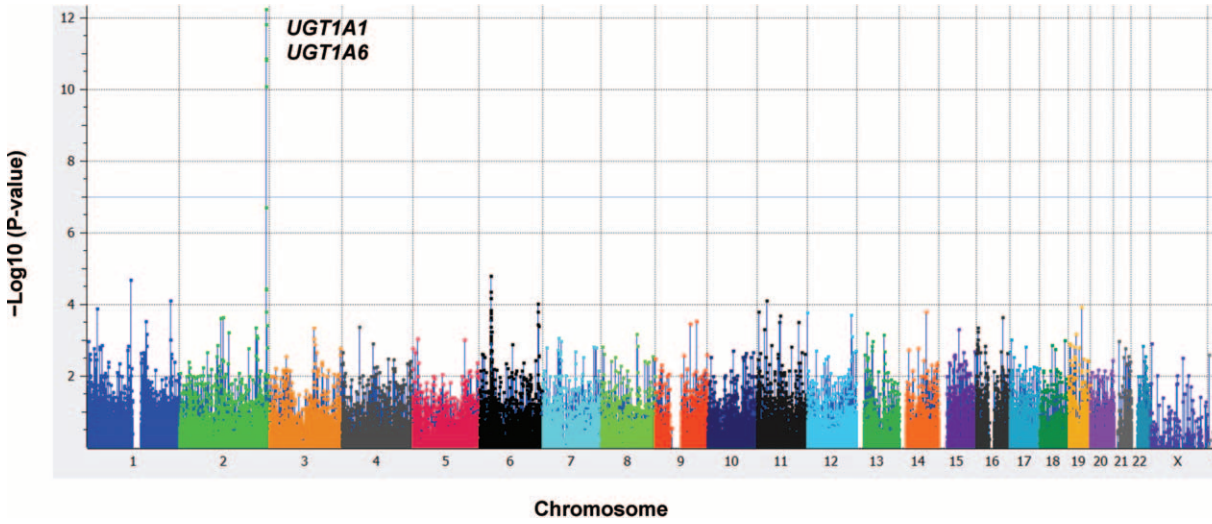

FIGURE 1. (A) Manhattan plot for total serum bilirubin level. Association results of the single-variant analysis (−log10 P) are plotted against genomic position (NCBI build 37). The blue horizontal line indicates the significance threshold for genome-wide association.

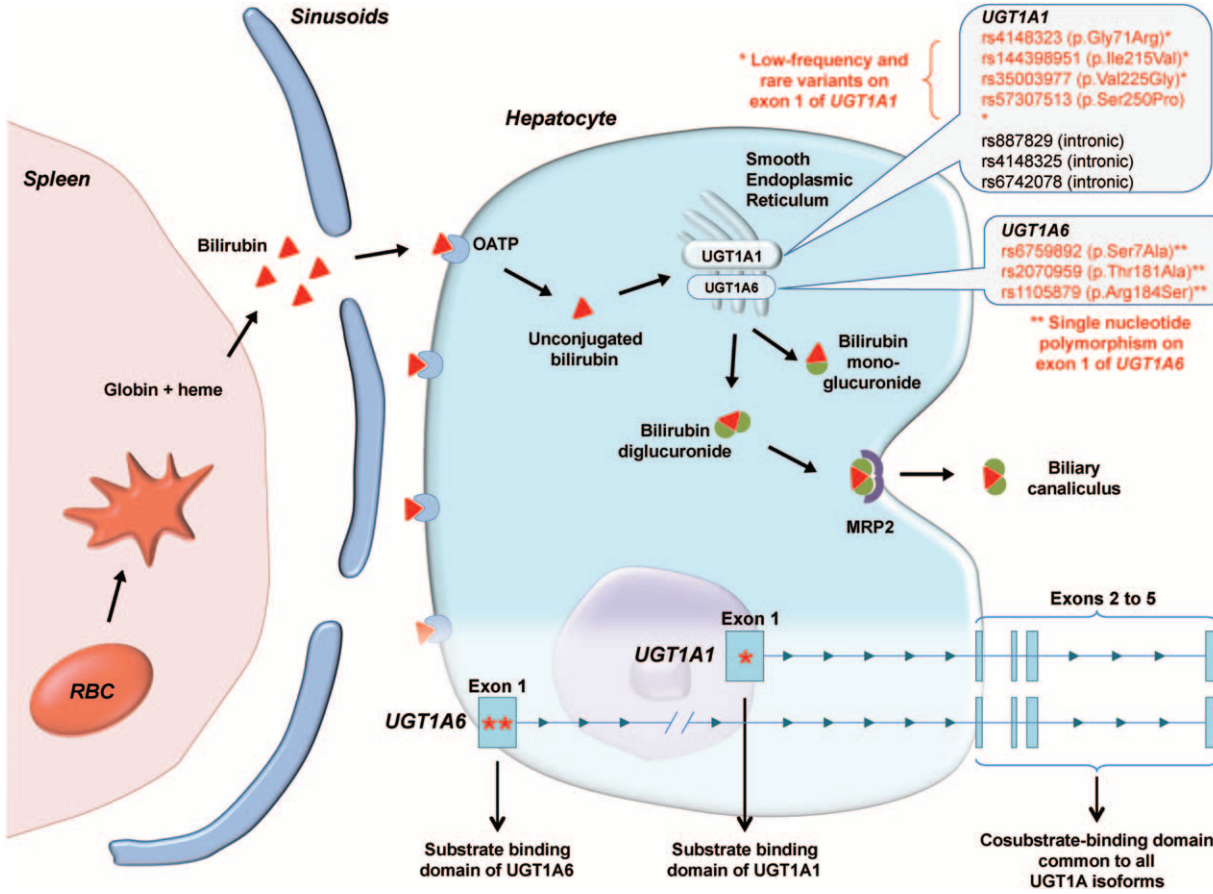

FIGURE 2. Bilirubin metabolism and *UGT1A1* and *UGT1A6* variants associated with unconjugated, conjugated, and total serum bilirubin level in physiological conditions. Coding variants are denoted in red font. *UGT1A1*: UDP-glucuronosyltransferase 1 family, polypeptide A1; *UGT1A6*: UDP-glucuronosyltransferase 1 family, polypeptide A6; OATP: organic anion transporter; MRP2: hepatic multidrug resistance protein.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

$P = 1.26 \times 10^{-24}$; rs1105879, $P = 1.41 \times 10^{-26}$) (Supplemental Digital Content: Table 6, http://links.lww.com/MD/A287). Results were similar after adjusting regression analysis for age and sex (Supplemental Digital Content: Table 4, http://links.lww.com/MD/A287). As reported for total bilirubin level, using the pooled data from both initial and *1* replication studies, for each 1 of the 6 variants, the median unconjugated serum bilirubin values differed significantly between the 3 subgroups of genotypes (homozygous major, heterozygous, and homozygous minor) on the Kruskal–Wallis test after Bonferroni correction with higher unconjugated serum bilirubin level observed in subjects with homozygous minor genotype (Supplemental Digital Content: Table 5 and Figure 6, http://links.lww.com/MD/A287).

Genetic Variants Associated With Conjugated Serum Bilirubin Level

Exome-Wide Analysis in the Initial Study (n = 400, Troina, Sicily, Enna)

Consistently with the exome-wide analysis on both total and unconjugated serum bilirubin level, the same top variants on *UGT1A1* and *UGT1A6* genes were significantly associated with serum conjugated bilirubin level (Supplemental Digital Content: Table 7 and Figure 8, http://links.lww.com/MD/A287) with the following *P* values for the non-synonymous *UGT1A6* variants: (rs2070959, $P = 4.37 \times 10^{-8}$, rs1105879, $P = 4.49 \times 10^{-8}$, rs6759892, $P = 1.33 \times 10^{-7}$).

In Silico Replication Study (n = 373, Cerami, Sicily, Enna)

In the in silico replication study, the 6 variants were significantly associated with conjugated serum bilirubin level (Supplemental Digital Content: Table 8, http://links.lww.com/MD/A287). The pooled analysis of initial and in silico replication studies reported the same figures for the 6 *UGT1A1* and *UGT1A6* variants with the following *P* values: (rs6759892, $P = 6.51 \times 10^{-21}$; rs2070959, $P = 1.11 \times 10^{-20}$; rs1105879, $P = 1.66 \times 10^{-21}$) (Supplemental Digital Content: Table 7, http://links.lww.com/MD/A287). Results were similar after adjusting regression analysis for age and sex (Supplemental Digital Content: Table 4, http://links.lww.com/MD/A287). As reported for both total and unconjugated serum bilirubin level, using the pooled data from both initial and in silico replication studies, for each of the 6 variants, the median conjugated serum bilirubin values differed significantly between the 3 subgroups of genotypes (homozygous major, heterozygous, and homozygous minor) on the Kruskal–Wallis test after Bonferroni correction with higher conjugated serum bilirubin level observed in subjects with homozygous minor genotype (Supplemental Digital Content: Table 5 and Figure 6, http://links.lww.com/MD/A287).

Combined Multivariate and Collapsing Method

Exome-wide data analysis using the CMC analysis approach on the pooled data from the 773 subjects retrieved 6 loci in the 2q37.1 region (position from 2:234590584 to 2:234681945), namely: *UGT1A4* ($P = 9.99 \times 10^{-19}$), *UGT1A3* ($P = 1.27 \times 10^{-18}$), *UGT1A1* ($P = 1.28 \times 10^{-18}$), *UGT1A5* ($P = 1.02 \times 10^{-13}$), *UGT1A6* ($P = 6.35 \times 10^{-10}$), *UGT1A7* ($P = 7.67 \times 10^{-10}$). All the 6 loci *P* values were Bonferroni significant (Supplemental Digital Content: Table 8 and Figure 9, http://links.lww.com/MD/A287). From the Illumina Infinium HumanExome BeadChip, 54 variants corresponded to the 6 loci spanning from position 2:234590584 to 2:234681945. Among them, only the 6 above-mentioned *UGT1A1* and *UGT1A6* variants were significantly associated with serum total bilirubin level (rs887829, rs4148325, rs6742078 on the *UGT1A1* locus and rs1105879, rs2070959, and rs6759892 on the *UGT1A6* locus) and 37 variants were monomorphic.

The LD analysis revealed that the 3 top intronic SNPs rs887829, rs4148325, and rs6742078 on *UGT1A1* locus were in strong LD with 4 missense rare and low-frequency coding variants located in the first exon of the *UGT1A1* gene, namely: rs4148323 (MAF = 0.06%; p.Gly71Arg), rs144398951 (MAF = 0.06%; p.Ile215Val), rs35003977 (MAF = 0.78%; p.Val225Gly), and rs57307513 (MAF = 0.06%; p.Ser250Pro) (Supplemental Digital Content: Table 9 and Figure 10, http://links.lww.com/MD/A287). Due to the low MAF of these 4 *UGT1A1* exonic variants, only rs35003977 (p.Val225Gly) was studied in exploratory post-hoc analysis in the initial study and was associated with an elevated risk of having a serum bilirubin level beyond the upper normal limit (total bilirubin>1.2 mg/dL [odds ratio, OR 4.82; 95% CI, 1.29–18.11; $P = 4.31 \times 10^{-2}$]; unconjugated bilirubin >0.7 mg/dL [OR = 4.52, 95% CI, 1.42–14.38; $P = 1.78 \times 10^{-2}$]; and conjugated bilirubin >0.3 mg/dL [OR = 7.45, 95% CI, 1.58–35.11; $P = 3.99 \times 10^{-2}$] Supplemental Digital Content: Table 10, http://links.lww.com/MD/A287).

Model Building and Structure-Based Analysis of UGT1A1 and UGT1A6 Proteins Variants

The 3D models of the UGT1A1 proteins harboring the rare variants amino acid substitutions (p.Gly71Arg, p.Ile215Val, p.Val225Gly, and p.Ser250Pro) show alterations in the 3D protein structure, narrow grooves, hydrophobic patches, and electrostatic potentials (Figure 3). Similarly, the 3D models of the UGT1A6 proteins harboring the amino acid substitutions (p.Ser7Ala, p.Thr181Ala, and p.Arg184Ser) show alterations in the 3D protein structure, narrow grooves, hydrophobic patches, and electrostatic potentials (Figure 4).

Replication Analysis on the 3 Top Genetic Variants on *UGT1A6* Gene (n = 227, Gagliano Castelferrato, Sicily, Enna)

The independent replication study was performed on 227 Italian subjects from the Gagliano Castelferrato commune (Sicily, Enna). As the 3 intronic variants of *UGT1A1* have been previously reported and validated for their association with serum bilirubin level,[22–29] we report here replication data on the 3 exonic missense variants on *UGT1A6* (rs6759892, p.Ser7Ala; rs2070959, p.Thr181Ala; and rs1105879, p.Arg184Ser). These variants were significantly associated with total serum bilirubin level, as well as unconjugated and conjugated serum bilirubin levels (Table 2).

Cholecystectomy Risk According to *UGT1A1* and *UGT1A6* Variants

Among the 773 OASI cohort participants from the Troina (Sicily, Enna) and Cerami (Sicily, Enna) communes, standardized collected data regarding gallstone-related cholecystectomy was available on 669 participants. Among them, 8 underwent cholecystectomy for cholelithiasis (1.20%; 95% CI, 0.37–2.02). All *UGT1A1* intronic variants (rs887829, rs6742078, and rs4148325) and *UGT1A6* exonic variants

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

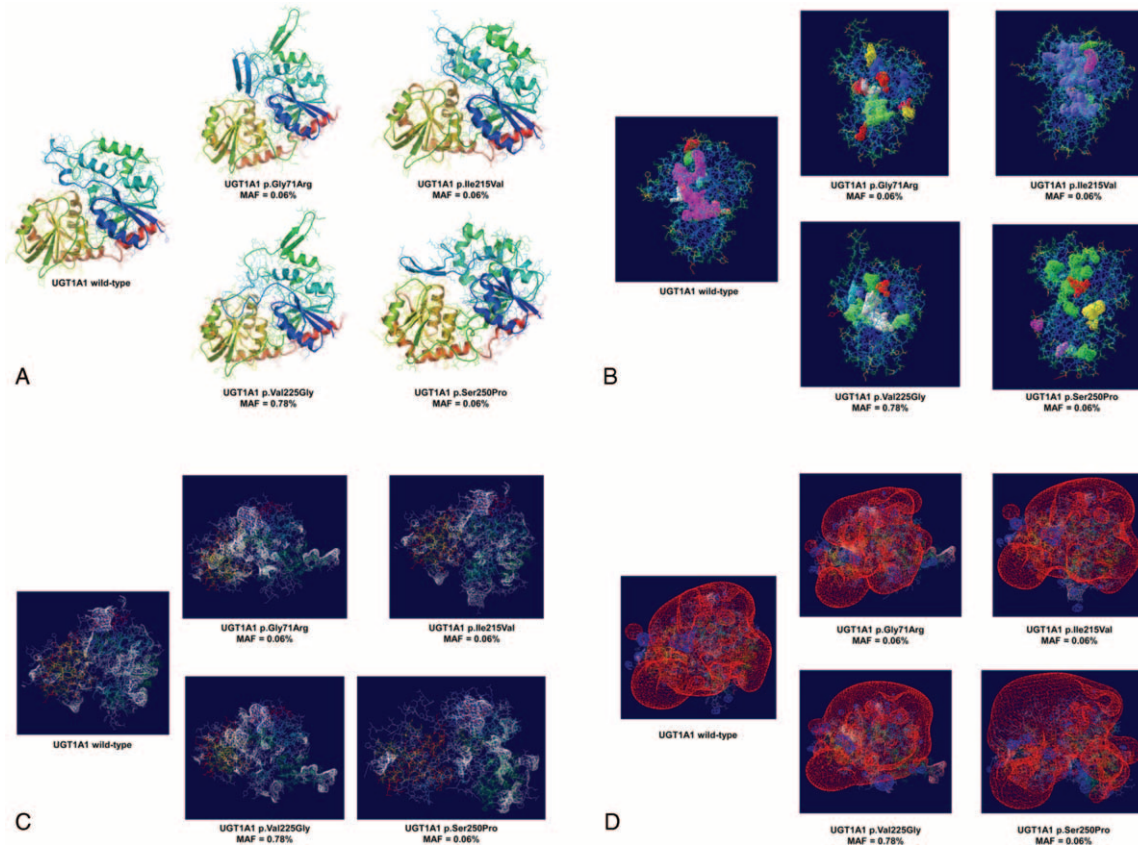

FIGURE 3. (A) Three-dimensional modeling of the human wild-type UGT1A1 protein and the low-frequency variants (p.Gly71Arg [MAF = 0.06%], p.Ile215Val [MAF = 0.06%], p.Val225Gly [MAF = 0.78%], and p.Ser250Pro [MAF = 0.06%]) performed using OpenAstexViewer 3.0. Secondary structure succession was colored using the 'rainbow' method. Elements of secondary structure are colored from the N-terminal to the C-terminal end in the order violet, blue, green, yellow, orange, and red. (B) Three-dimensional modeling of the human wild-type UGT1A1 protein and the low-frequency variants (p.Gly71Arg [MAF = 0.06%], p.Ile215Val [MAF = 0.06%], p.Val225Gly [MAF = 0.78%], and p.Ser250Pro [MAF = 0.06%]) performed using DeepView/Swiss-PdbViewer 4.1.0. Secondary structure succession was colored using the "rainbow" method. Narrow grooves in each model are depicted as a 3-dimensional network in a different color than the underlying secondary structure of the protein. (C) The negative electrostatic potential map is displayed as a red-colored 3-dimensional network surrounding the protein. (D) The negative electrostatic potential map is displayed as a red-colored three-dimensional network surrounding the protein.

(rs6759892, rs2070959, and rs1105879) were significantly associated with gallstone-related cholecystectomy risk (Table 3). The *UGT1A6* variant rs2070959 (p.Thr181Ala) was associated with the highest risk of gallstone–related cholecystectomy (OR, 4.58; 95% CI, 1.58–13.28; $P = 3.21 \times 10^{-3}$) (Figure 5). None of the 4 low-frequency and rare exonic variants of *UGT1A1* (rs4148323, p.Gly71Arg; rs144398951, p.Ile215-Val; rs35003977, p.Val225Gly; and rs57307513, p.Ser250Pro) has been replicated due to their low MAF.

DISCUSSION

The UGT1A enzymes family is derived from a single gene locus (*UGT1A*) spanning about 210 kb on chromosome 2 (2q37), which is composed of 17 exons including 13 unique alternate first exons followed by 4 common exons.[30] Because of alternative splicing, only one of 13 different exon-1 sequences on the locus is associated with the 4 downstream exons, which are common to all *UGT1A* gene isoforms.[30] Among the 13 exon-1 sequences, 9 code for functional proteins (UGT1A1, UGT1A3 to UGT1A10) and 4 correspond to pseudogenes (UGT1A2p, UGT1A11p, UGT1 A12p, and UGT1A13p).[30,31] UGT1A proteins are composed of 265 (UGT1A6) to 534 amino-acid residues (UGT1A3, 4, and 5) with a molecular weight of 50 to 57 kDa.[30] The exon-1 sequence of UGTs codes for the substrate-binding domain (N-terminal half of the protein), whereas the 4 common exons code for the cosubstrate-binding domain (C-terminal half of the protein).[30]

Several SNPs in the *UGT1A1* locus have been reported in previous studies to be significantly associated with serum bilirubin.[22–29,32–34] The reported *UGT1A1* SNPs are intronic (rs6742078, rs887829, rs4148324, and rs4148325) and are in complete linkage disequilibrium within the same haplotype block.[24] To date, outside of published polymorphisms on *UGT1A1* gene isoform, no data have been reported regarding potential variants on other isoforms of *UGT1A* gene complex in physiological conditions.

We identified 4 low-frequency and rare missense variants located in the first exon substrate-binding domain coding region of the *UGT1A1* gene (rs4148323, p.Gly71Arg; rs144398951, p.Ile215Val; rs35003977, p.Val225Gly; and rs57307513, p.Ser250Pro). These coding variants are in strong linkage disequilibrium with 3 intronic *UGT1A1* SNPs, which are highly associated with serum bilirubin concentration. The *UGT1A1*

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

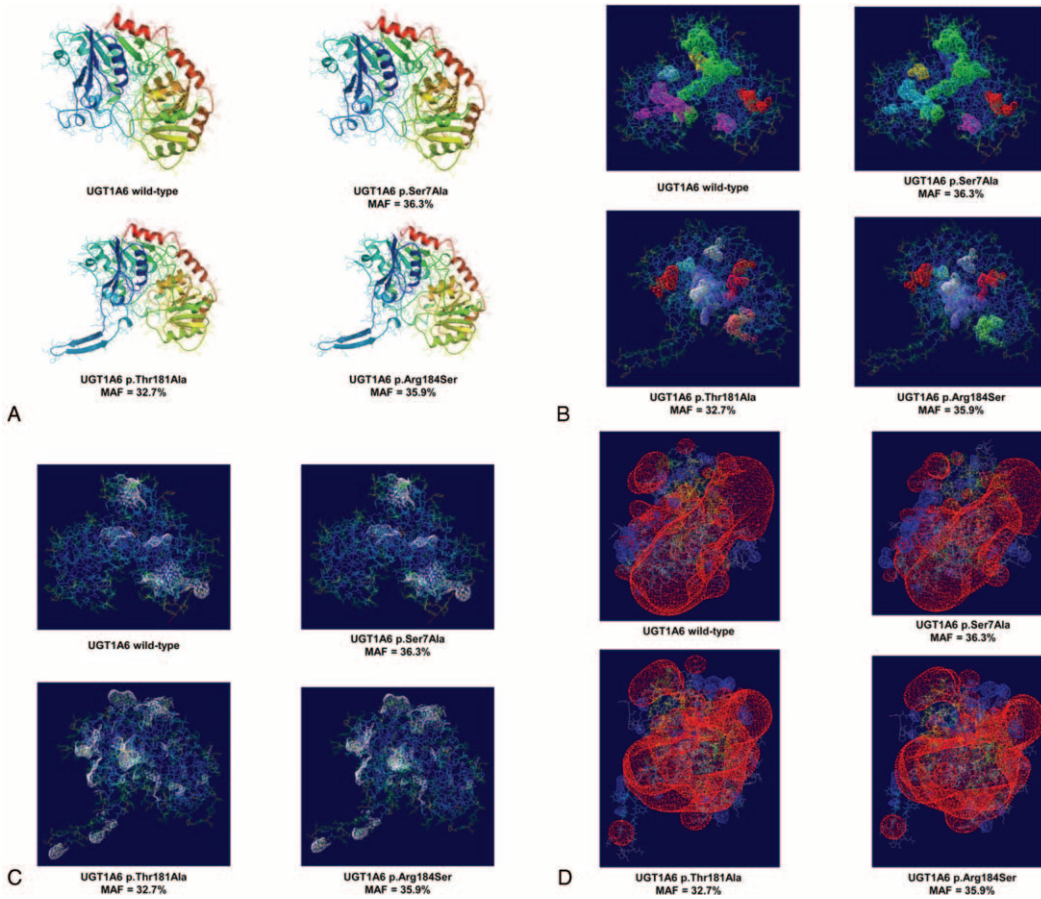

FIGURE 4. (A) Three-dimensional modeling of the human wild-type UGT1A6 protein and the protein variants p.Ser7Ala (MAF = 36.3%), p.Thr181Ala (MAF = 32.7%), and p.Arg184Ser (MAF = 35.9%), and p.Ser250Pro (MAF = 0.06%) performed using OpenAstexViewer 3.0. Secondary structure succession was colored using the ''rainbow'' method. Elements of secondary structure are colored from the N-terminal to the C-terminal end in the order violet, blue, green, yellow, orange, and red. (B) Three-dimensional modeling of the human wild-type UGT1A6 protein and the protein variants p.Ser7Ala (MAF = 36.3%), p.Thr181Ala (MAF = 32.7%), and p.Arg184Ser (MAF = 35.9%) performed using DeepView/Swiss-PdbViewer 4.1.0. Secondary structure succession was colored using the ''rainbow'' method. Narrow grooves in each model are depicted as a 3-dimensional network in a different color than the underlying secondary structure of the protein. (C) The negative electrostatic potential map is displayed as a red-colored three-dimensional network surrounding the protein. (D) The negative electrostatic potential map is displayed as a red-colored three-dimensional network surrounding the protein.

TABLE 2. Association Between UGT1A6 Missense Variants and Serum Bilirubin (Total, Unconjugated, and Conjugated) in the Replication Study (Gagliano Castelferrato, Sicily, Enna)

Marker	MAF	Regression *P*	Regression Bonf. *P*	Regression Slope	Regression Slope SE	Corr/ Trend *P*	Corr/Trend Bonf. *P*	Corr/ Trend R
Total serum bilirubin (allelic model)								
rs6759892 (p.Ser7Ala)	0.453	7.59×10^{-10}	2.28×10^{-9}	0.1808	0.0287	1.67×10^{-9}	5.01×10^{-9}	0.2944
rs2070959 (p.Thr181Ala)	0.410	1.19×10^{-11}	3.56×10^{-11}	0.1950	0.0280	3.53×10^{-11}	1.06×10^{-10}	0.3125
rs1105879 (p.Arg184Ser)	0.434	1.02×10^{-11}	3.07×10^{-11}	0.1918	0.0274	3.15×10^{-11}	9.46×10^{-11}	0.3162
Unconjugated bilirubin (allelic model)								
rs6759892 (p.Ser7Ala)	0.453	7.76×10^{-11}	2.33×10^{-10}	0.1485	0.0222	2.08×10^{-10}	6.25×10^{-10}	0.3105
rs2070959 (p.Thr181Ala)	0.410	6.23×10^{-13}	1.87×10^{-12}	0.1599	0.0216	2.49×10^{-12}	7.47×10^{-12}	0.3305
rs1105879 (p.Arg184Ser)	0.434	5.58×10^{-13}	1.67×10^{-12}	0.1578	0.0212	2.31×10^{-12}	6.93×10^{-12}	0.3340
Conjugated bilirubin (allelic model)								
rs6759892 (p.Ser7Ala)	0.453	1.64×10^{-5}	4.92×10^{-5}	0.0322	0.0074	1.96×10^{-05}	5.89×10^{-5}	0.2086
rs2070959 (p.Thr181Ala)	0.410	1.83×10^{-6}	5.48×10^{-6}	0.0352	0.0073	2.37×10^{-06}	7.10×10^{-6}	0.2227
rs1105879 (p.Arg184Ser)	0.434	2.18×10^{-6}	6.55×10^{-6}	0.0340	0.0071	2.82×10^{-06}	8.45×10^{-6}	0.2230

Bonf = Bonferroni, Corr = Correlation, MAF = minor allele frequency, SE = standard error.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 3. Effect-sizes of Gallstone-related Cholecystectomy Risk Among the OASI Cohort Participants According to UGT1A1 and UGT1A6 Genetic Variants

Variant	Gene	HGVS (Amino Acid Change)	P^*	OR† (95% CI)	β 1	β 1 SE	MAF, "Chol." group	MAF, "No Chol." Group	dd n (%)	Dd n (%)	DD n (%)
rs6759892	*UGT1A6*	c.19T > G (p.Ser7Ala)	8.66×10^{-3}	3.88 (1.34–11.26)	1.357	0.543	0.688	0.362	241 (0.42)	261 (1.15)	84 (4.76)
rs2070959	*UGT1A6*	c.541A > G (p.Thr181Ala)	3.21×10^{-3}	4.58 (1.58–13.28)	1.522	0.543	0.688	0.324	269 (0.37)	248 (1.21)	69 (5.80)
rs1105879	*UGT1A6*	c.552A > C (p.Arg184Ser)	7.67×10^{-3}	3.97 (1.37–11.49)	1.378	0.543	0.688	0.357	246 (0.41)	254 (1.18)	84 (4.76)
rs887829	*UGT1A1*	c.-364C > T (intronic)	1.69×10^{-2}	3.38 (1.22–9.36)	1.217	0.520	0.625	0.330	270 (0.74)	240 (0.83)	76 (5.26)
rs6742078	*UGT1A1*	c.864+2842G > T (intronic)	1.78×10^{-2}	3.35 (1.34–11.26)	1.208	0.520	0.625	0.332	269 (0.74)	241 (0.83)	77 (5.19)
rs4148325	*UGT1A1*	c.865−2371C > T (intronic)	1.78×10^{-2}	3.35 (1.58–13.28)	1.208	0.520	0.625	0.332	269 (0.74)	241 (0.83)	77 (5.19)

% = proportion of subjects who underwent cholecystectomy, β 1 = beta 1 coefficient, 95% CI = 95% confidence interval, AA = amino acid, Chol = cholecystectomy, Corr = Correlation, Dd = heterozygous, DD = homozygous for the minor allele, dd = homozygous wild-type, HGVS = Human genome variation society, MAF = minor allele frequency, n = number of subjects in the genotype subgroup, OR = odds ratio, SE = standard error.

* Logistic regression analysis.

† Odds ratio for the minor allele.

intronic SNPs (rs6742078, rs887829, rs4148324) serve as proxy markers for the low-frequency and rare *UGT1A1*-coding variants, thereby providing mechanistic explanation to the relationship between *UGT1A1* intronic SNPs and the UGT1A1 enzyme activity.

Among the 4 low-frequency and rare variants that we found on the exon-1 of the *UGT1A1* gene, only 1 (rs4148323, p.Gly71Arg) has been reported to be associated with serum bilirubin level in Asian populations.[22,29,33,35,36] Using the 1000 genomes database, the MAF of the rs4148323 (p.Gly71Arg) variant was 0.9% in Europeans versus 17.3% in East Asians. Regarding the 3 other missense variants that we found on *UGT1A1* first exon, the MAFs in Europeans for rs144398951 (p.Ile215Val), rs35003977 (p.Val225Gly), and rs57307513 (p.Ser250Pro) were 0% (Exome Sequencing Project, ESP6500), 0.1% (ESP6500), and 0% (1000 genomes), respectively. None of these variants has been reported to date as potentially associated with serum bilirubin level.

It has been demonstrated that 2 alternative orientations for bilirubin binding to UGT1A1, either toward the loop 5 and the membrane or toward the loop 3 and the endoplasmic reticulum lumen were possible.[37] In the loop 5-binding configuration, bilirubin is in contact with residues Gln84, Asp87, Lys115, and Ile116, whereas in the loop 3-binding configuration, bilirubin is in contact with Lys115, Ile116, Ala174, Val193, and Asp224.[37] It should be mentioned, however, that the substrates in the known glucuronosyltransferase structures are found closer to loop 5.[37] It is noteworthy that 3 of the 4 variants on exon-1 of *UGT1A1* gene (rs144398951, p.Ile215Val; rs35003977, p.Val225Gly; and rs57307513, p.Ser250Pro) are close to the loop 5 on the corresponding UGT1A1 protein and therefore can potentially alter the binding site of the UGT1A1 protein to bilirubin.

In our study, we report for the first time a relationship between 3 nonsynonymous exonic SNPs of the *UGT1A6* gene (rs6759892, p.Ser7Ala; rs2070959, p.Thr181Ala; rs11058792, p.Arg184Ser) with unconjugated, conjugated, and total serum bilirubin level. To date, no study has reported association between *UGT1A6* gene variations and serum bilirubin in the general population. It has been demonstrated in whites that the 2 common variants, rs2070959 (p.Thr181Ala) and rs11058792 (p.Arg184Ser), were associated with 30% to 50% lower UGT1A6 enzyme activity.[38] In a recent study on subjects with Gilbert syndrome—characterized by intermittent unconjugated hyperbilirubinemia without structural liver damage, affecting about 10% of the white population—UGT1A6 protein variants (p.Ser7Ala, p.Thr181Ala, and p.Arg184Ser) were observed in 80%, 79%, and 80% of subjects, respectively versus 9% in healthy blood donors.[39] A humanized transgenic *UGT1A* mouse model containing a gene fragment from human chromosome 2 spanning *UGT1A1* up to a portion of *UGT1A10*, which encodes the most prevalent wild-type sequence (*htgUGT1A wild-type*) and a fragment of equal length containing *UGT1A6* SNPs (*htgUGT1A SNP*), showed lower *UGT1A6* messenger RNA (mRNA) expression and UGT1A6 protein synthesis in comparison with the corresponding wild-type mouse model.[39]

To date, 11 GWASs have reported the association with unconjugated, conjugated, and/or total bilirubin; however, no study reported significant association between *UGT1A6* locus and serum bilirubin the general population (Supplemental Digital Content: Table 11, http://links.lww.com/MD/A287).[22–29,32–34] In our study, using an exome-array approach, we demonstrated for the first time that the *UGT1A6* locus significantly influence total serum bilirubin as well as unconjugated and conjugated serum bilirubin in physiological conditions. Milton et al[28]

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

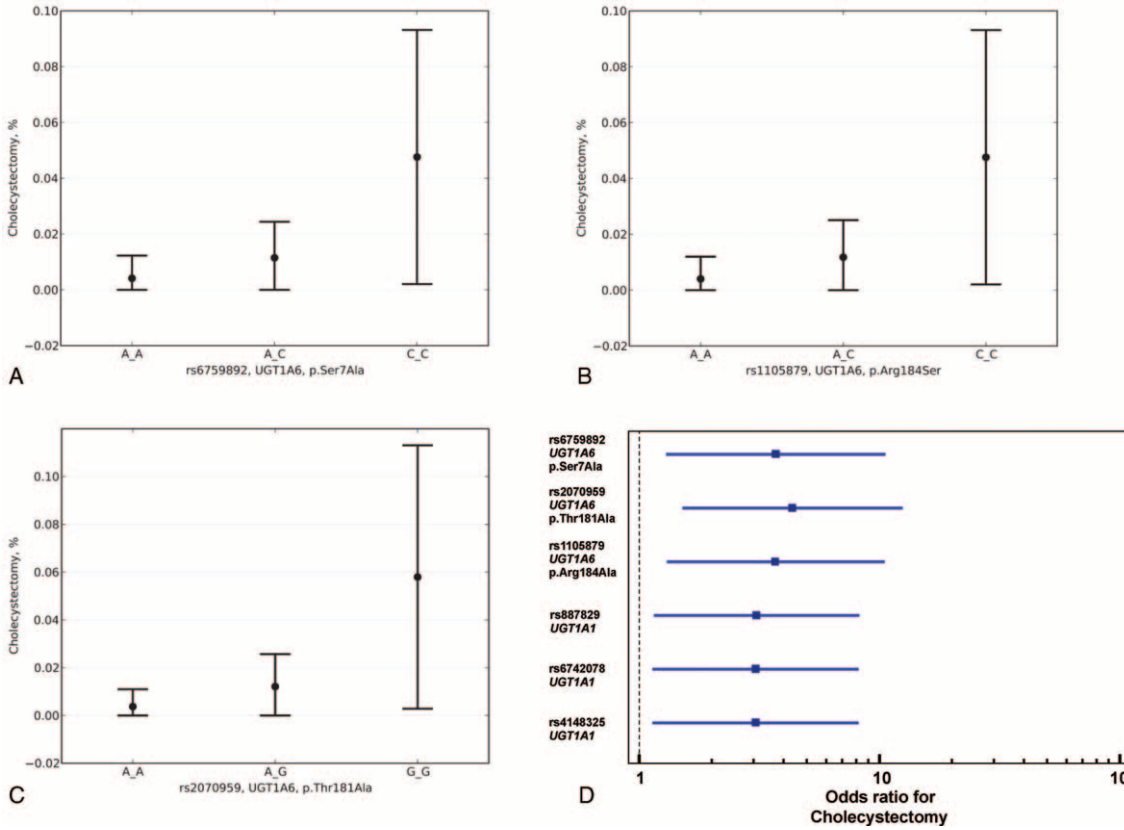

FIGURE 5. Gallstone-related cholecystectomy risk, according to *UGT1A6* variants: (A) (rs6759892, p. Ser7Ala), (B) (rs2070959, p. Thr181Ala), and (C) (rs1105879, p. Arg184Ser); (D) gallstone-related cholecystectomy risk effect sizes according to *UGT1A1* and *UGT1A6* variants.

performed a GWAS on total bilirubin levels and cholelithiasis risk in patients with sickle cell anemia. In this study, total bilirubin level and cholelithiasis risk were associated with *UGT1A6* locus, confirming its potential direct implication in serum bilirubin determinism.[28] However, results from sickle cell anemia patients' cohorts could not be inferred to general population and bilirubin metabolism in physiological conditions because the *UGT1A* alleles and haplotypes undergo protozoa-driven selective pressure.[40]

A major strength of the OASI cohort is that it is focused on the elderly and has assessed the phenotypic expression of genetic variants that require decades for their clinical translation. We reported, for the first time, an association between *UGT1A1* intronic variants and *UGT1A6* exonic variants (rs6759892, rs2070959, and rs1105879) and gallstone-related cholecystectomy risk. Indeed, using standardized collected data on a well-characterized cohort of ambulatory elderly subjects with a median age of 72 years, we showed for the first time high effect sizes for the significant association between *UGT1A1* and *UGT1A6* variants and the risk of gallstone-related cholecystectomy. In a meta-analysis on 2 GWASs, using hospital discharge diagnosis records Johnson et al[24] reported no association between *UGT1A1* intronic SNP (rs6742078) and cholelithiasis-related admission, history of gallbladder disease, or gallbladder removal. Milton et al[28] reported a small effect size significant association between *UGT1A1* SNPs and cholelithiasis risk in a young African American sickle cell anemia patients (median age around 33 years). In that study, cholelithiasis risk was defined by the presence of gallstones, non-functional gallbladder, or cholecystectomy and no data were reported separately on the risk of cholecystectomy as a hard criteria.[28] Moreover, results on sickle cell anemia patients' cohorts could not be inferred to general population and bilirubin metabolism in physiological conditions.

Our study was limited in its ability to look at very rare variants because of the content of the exome array. Although sequencing will still be required to completely assess variants associated with serum bilirubin, this study provides proof of concept that exome array genotyping is a powerful approach to identify low-frequency functional variants.[14]

In conclusion, this study demonstrates for the first time that exome-array genotyping is a valuable approach to identify protein-coding variants that contribute to bilirubin metabolism and sheds new light on the genetic architecture of bilirubin metabolism, thus providing new evidence on the relationship between bilirubin metabolism and the risk of gallstone disease.

ACKNOWLEDGMENTS

The authors thank all the data managers and hospitals for participating in the study and for their great efforts in entering data.

REFERENCES

1. Henry JB, McPherson RA, Pincus MR. Henry's clinical diagnosis and management by laboratory methods. 22nd ed. Philadelphia, PA: Elsevier/Saunders; 2011:http://www.clinicalkey.com/dura/browse/bookChapter/3-s2.0-C20090459154.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

2. Jedlitschky G, Leier I, Buchholz U, et al. ATP-dependent transport of bilirubin glucuronides by the multidrug resistance protein MRP1 and its hepatocyte canalicular isoform MRP2. *Biochem J.* 1997;327 (Pt 1):305–310.
3. Soloway RD, Trotman BW, Ostrow JD. Pigment gallstones. *Gastroenterology.* 1977;72:167–182.
4. Dutt MK, Murphy GM, Thompson RP. Unconjugated bilirubin in human bile: the nucleating factor in cholesterol cholelithiasis? *J Clin Pathol.* 2003;56:596–598.
5. Haverfield EV, McKenzie CA, Forrester T, et al. UGT1A1 variation and gallstone formation in sickle cell disease. *Blood.* 2005;105:968–972.
6. Lee S, Abecasis GR, Boehnke M, et al. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet.* 2014;95:5–23.
7. Rivas MA, Beaudoin M, Gardet A, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet.* 2011;43:1066–1073.
8. Wu L, Schaid DJ, Sicotte H, et al. Case-only exome sequencing and complex disease susceptibility gene discovery: study design considerations. *J Med Genet.* 2015;52:10–16.
9. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A.* 2009;106:19096–19101.
10. Lupski JR, Gonzaga-Jauregui C, Yang Y, et al. Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2 variant alleles causing Charcot-Marie-Tooth neuropathy. *Genome Med.* 2013;5:57.
11. Gueant-Rodriguez RM, Gueant JL, Debard R, et al. Prevalence of methylenetetrahydrofolate reductase 677T and 1298C alleles and folate status: a comparative study in Mexican, West African, and European populations. *Am J Clin Nutr.* 2006;83:701–707.
12. Gueant JL, Chabi NW, Gueant-Rodriguez RM, et al. Environmental influence on the worldwide prevalence of a 776C->G variant in the transcobalamin gene (TCN2). *J Med Genet.* 2007;44:363–367.
13. Oussalah A, Besseau C, Chery C, et al. Helicobacter pylori serologic status has no influence on the association between fucosyltransferase 2 polymorphism (FUT2 461 G->A) and vitamin B-12 in Europe and West Africa. *Am J Clin Nutr.* 2012;95:514–521.
14. Huyghe JR, Jackson AU, Fogarty MP, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet.* 2013;45:197–201.
15. Price AL, Weale ME, Patterson N, et al. Long-range LD can confound genome scans in admixed populations. *Am J Hum Genet.* 2008;83:132–135author reply 135–139.
16. Weale ME. Quality control for genome-wide association studies. *Methods Mol Biol.* 2010;628:341–372.
17. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet.* 2008;83:311–321.
18. Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics.* 2006;22:195–201.
19. Kiefer F, Arnold K, Kunzli M, et al. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* 2009;37 (Database issue):D387–D392.
20. Hartshorn MJ. AstexViewer: a visualisation aid for structure-based drug design. *J Comput Aided Mol Des.* 2002;16:871–881.
21. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* 1997;18:2714–2723.
22. Saito A, Kawamoto M, Kamatani N. Association study between single-nucleotide polymorphisms in 199 drug-related genes and commonly measured quantitative traits of 752 healthy Japanese subjects. *J Hum Genet.* 2009;54:317–323.
23. Sanna S, Busonero F, Maschio A, et al. Common variants in the SLCO1B3 locus are associated with bilirubin levels and unconjugated hyperbilirubinemia. *Hum Mol Genet.* 2009;18:2711–2718.
24. Johnson AD, Kavousi M, Smith AV, et al. Genome-wide association meta-analysis for total serum bilirubin levels. *Hum Mol Genet.* 2009;18:2700–2710.
25. Buch S, Schafmayer C, Volzke H, et al. Loci from a genome-wide analysis of bilirubin levels are associated with gallstone risk and composition. *Gastroenterology.* 2010;139:1942–1951e1942.
26. Bielinski SJ, Chai HS, Pathak J, et al. Mayo Genome Consortia: a genotype-phenotype resource for genome-wide association studies with an application to the analysis of circulating bilirubin levels. *Mayo Clin Proc.* 2011;86:606–614.
27. Chen G, Ramos E, Adeyemo A, et al. UGT1A1 is a major locus influencing bilirubin levels in African Americans. *Eur J Hum Genet.* 2012;20:463–468.
28. Milton JN, Sebastiani P, Solovieff N, et al. A genome-wide association study of total bilirubin and cholelithiasis risk in sickle cell anemia. *PLoS One.* 2012;7:e34741.
29. Dai X, Wu C, He Y, et al. A genome-wide association study for serum bilirubin levels and gene-environment interaction in a Chinese population. *Genet Epidemiol.* 2013;37:293–300.
30. Guillemette C. Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *Pharmacogenomics J.* 2003;3:136–158.
31. Gong QH, Cho JW, Huang T, et al. Thirteen UDPglucuronosyltransferase genes are encoded at the human UGT1 gene complex locus. *Pharmacogenetics.* 2001;11:357–368.
32. Lin JP, Schwaiger JP, Cupples LA, et al. Conditional linkage and genome-wide association studies identify UGT1A1 as a major gene for anti-atherogenic serum bilirubin levels–the Framingham Heart Study. *Atherosclerosis.* 2009;206:228–233.
33. Kang TW, Kim HJ, Ju H, et al. Genome-wide association of serum bilirubin levels in Korean population. *Hum Mol Genet.* 2010;19:3672–3678.
34. Datta S, Chowdhury A, Ghosh M, et al. A genome-wide search for non-UGT1A1 markers associated with unconjugated bilirubin level reveals significant association with a polymorphic marker near a gene of the nucleoporin family. *Ann Hum Genet.* 2012;76:33–41.
35. Lin R, Wang X, Wang Y, et al. Association of polymorphisms in four bilirubin metabolism genes with serum bilirubin in three Asian populations. *Hum Mutat.* 2009;30:609–615.
36. Lin R, Wang Y, Fu W, et al. Common variants of four bilirubin metabolism genes and their association with serum bilirubin and coronary artery disease in Chinese Han population. *Pharmacogenet Genom.* 2009;19:310–318.
37. Laakkonen L, Finel M. A molecular model of the human UDP-glucuronosyltransferase 1A1, its membrane orientation, and the interactions between different parts of the enzyme. *Mol Pharmacol.* 2010;77:931–939.
38. Ciotti M, Marrone A, Potter C, et al. Genetic polymorphism in the human UGT1A6 (planar phenol) UDP-glucuronosyltransferase: pharmacological implications. *Pharmacogenetics.* 1997;7:485–495.
39. Ehmer U, Kalthoff S, Fakundiny B, et al. Gilbert syndrome redefined: a complex genetic haplotype influences the regulation of glucuronidation. *Hepatology.* 2012;55:1912–1921.
40. Horsfall LJ, Zeitlyn D, Tarekegn A, et al. Prevalence of clinically relevant UGT1A alleles and haplotypes in African populations. *Ann Hum Genet.* 2011;75:236–246.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.