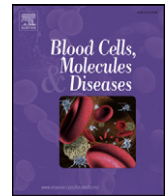




Contents lists available at SciVerse ScienceDirect

Blood Cells, Molecules, and Diseases

journal homepage: www.elsevier.com/locate/ybcm

Impact of *UGT1A1* gene variants on total bilirubin levels in Gilbert syndrome patients and in healthy subjects

Carina Rodrigues^{a,b}, Emília Vieira^c, Rosário Santos^c, João de Carvalho^d, Alice Santos-Silva^{a,e}, Elísio Costa^{e,f}, Elsa Bronze-da-Rocha^{a,e,*}

^a Departamento de Ciências Biológicas, Laboratório de Bioquímica, Faculdade de Farmácia da Universidade do Porto, Portugal

^b Departamento das Tecnologias de Diagnóstico e Terapêutica, Escola Superior de Saúde do Instituto Politécnico de Bragança, Portugal

^c Unidade de Genética Molecular, Centro de Genética Médica Dr. Jacinto Magalhães INSARJ, Porto, Portugal

^d Serviço de Gastrenterologia, Centro Hospitalar de Vila Nova de Gaia, Portugal

^e Instituto de Biologia Molecular e Celular da Universidade do Porto, Portugal

^f Instituto de Ciências da Saúde da Universidade Católica do Porto, Portugal

ARTICLE INFO

Article history:

Submitted 7 December 2011

Revised 9 January 2012

Available online 9 February 2012

Communicated by M. Lichtman, M.D.,
11 January 2012

Keywords:

UGT1A1 variants

Bilirubin levels

Polymorphism phenotype prediction

Gilbert syndrome

SNPs

ABSTRACT

The Gilbert syndrome is a benign form of unconjugated hyperbilirubinemia, mainly associated with alterations in *UGT1A1* gene. This work investigated the effect of *UGT1A1* variants on total bilirubin levels in Gilbert patients ($n = 45$) and healthy controls ($n = 161$). Total bilirubin levels were determined using a colorimetric method; molecular analysis of exons 1–5 and two *UGT1A1* promoter regions were performed by direct sequencing and automatic analysis of fragments. Five *in silico* methods predicted the effect of new identified variants.

A significant different allelic distribution, in Gilbert patients and in controls, was found for two promoter polymorphisms. Among patients, 82.2% were homozygous and 17.8% heterozygous for the $c.-41_-40dupTA$ allele; in control group, 9.9% were homozygous and 43.5% heterozygous for this promoter variant, while 46.6% ($n = 75$) presented the $[A(TA)_6TAA]$. For the T>G transition at $c.-3279$ promoter region, in patients, 86.7% were homozygous and 13.3% heterozygous; in control group, 33.5% were homozygous for the wild type allele, 44.1% were heterozygous and 22.4% homozygous for the mutated allele. The two polymorphisms were in Hardy–Weinberg equilibrium in both groups. Sequencing of *UGT1A1* coding region identified nine novel variants, five in patients and four in controls. *In silico* analysis of these amino acids replacements predicted four of them as benign and three as damaging.

In conclusion, we demonstrated that total bilirubin levels are mainly determined by the TA duplication in the TATA-box promoter and by the $c.-3279T>G$ variant. Alterations in the *UGT1A1* coding region seem to be associated with increased bilirubin levels, and, therefore, with Gilbert syndrome.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Gilbert syndrome (GS, OMIM 606785) is an autosomal recessive condition characterized by unconjugated hyperbilirubinemia in the absence of hemolysis or underlying liver disease [1], due to the reduced activity of uridine diphosphate-glucuronosyltransferase (*UGT1A1*). This enzyme is mainly expressed in the liver and has an important role in the glucuronidation of bilirubin, 17β -estradiol [2], some therapeutic drugs [3] and mutagenic xenobiotics [4]. Absence or severe reductions of *UGT1A1* activity are associated with Crigler-Najjar syndrome type I [5] and type II [6], respectively. Crigler Najjar syndrome is an autosomal recessive disease; so a subject must be

carrying at least two mutations in *trans* to manifest the disease. Only one mutation should not affect the TBL levels. However, if such nucleotide substitution is in *trans* with an allele carrying the $A(TA)_7TAA$ allele ($c.-41_-40dupTA$) it is possible that the bilirubin levels could be compatible with the GS or with Crigler Najjar type II syndrome [7].

The *UGT1A1* gene locus has been mapped to chromosome 2q37 [8], has nine potential functional exon 1 sequences at the 5' end of the gene that can be joint with the constant exons 2–5 at the 3' end [3]. One of the most common genetic variants is a repeat polymorphism in the TATA box region of the *UGT1A1* promoter. This allele variant consists of seven thymine adenine (TA) repeats, in the $A(TA)_nTAA$ motif, while the common allele is characterized by the presence of six TA repeats $[A(TA)_6TAA]$. Homozygous individuals carrying the $A(TA)_7TAA$ allele ($c.-41_-40dupTA$) have higher levels of unconjugated bilirubin, caused by a reduction of 30% in *UGT1A1* transcription [9]. The estimated frequency of this allele is 0.35 in Caucasians,

* Corresponding author at: Departamento de Ciências Biológicas, Laboratório de Bioquímica, Faculdade de Farmácia da Universidade do Porto, Rua Anibal Cunha, 164, 4050-047 Porto, Portugal. Fax: +351 222 003 977.

E-mail address: elsa.rocha@ff.up.pt (E. Bronze-da-Rocha).

leading to a homozygous genotype in about 10% of the population, but the frequency is highly variable in different ethnicities [10,11]. Homozygosity for the TA duplication was considered as the main cause of GS in Caucasian population [9,10], and justifies some of the inter-individual variations in bilirubin levels, even in the normal population [12]. However, individuals with normal bilirubin levels and without clinical diagnosis of GS can also show this polymorphism in homozygosity [3,9]. Thus, this polymorphism appears to be necessary for GS, but it is not sufficient for the complete manifestation of the disease [12]. Another promoter polymorphism, a T>G transition at position –3279, at the phenobarbital responsive enhancer module (gtPBREM), has also been associated with GS [13,14]. In Caucasian and Japanese patients this polymorphism was described and referred to be in linkage disequilibrium with the TA duplication (c.–41_–40dupTA) [13–16]. In the last years more than 100 single nucleotide polymorphisms (SNPs) have also been described within the *UGT1A1* gene [3].

The aim of the present work was the analysis of the impact of both promoter and/or coding region of *UGT1A1* variants on bilirubin levels, in GS patients and in healthy controls. Additionally, bioinformatics tools were used to study the new identified variants in order to evaluate their clinical impact.

Patients and methods

Subjects

This study included 45 GS Caucasian patients (36 males and 9 females, with a mean age of 45 ± 13.2 years) of the gastroenterology department of Vila Nova de Gaia/Espinho Hospital Center, EPE, with clinical diagnosis of GS, based on standard criteria [17,18]. Hyperbilirubinemia was defined as a concentration of total bilirubin above $17.1 \mu\text{mol/L}$. Data from GS patients were obtained from medical records. Total bilirubin levels (TBL) correspond to the median of three bilirubin measurements, and every GS patients presented normal physical examination, liver function tests and blood cell counts.

Healthy control volunteers ($n = 161$; 45 males and 116 females; mean age of 21.7 ± 5.6 years) were included in this work. Healthy controls were selected based on normal liver and/or hematological values, no history of inflammatory or malignancy diseases, and of regular medication. Control participants were also excluded when conjugated bilirubin was more than 20% of total bilirubin. GS patients and healthy controls, gave their informed consent to participate in this study.

Samples and measurements

Blood samples were collected in the morning after 8 h of fasting, using EDTA as anticoagulant, in order to obtain plasma and buffy coat for DNA extraction. TBL concentration was determined by using a colorimetric method (diazotized sulfanilic acid reaction; Roche Diagnostics Co, Indianapolis). Genomic DNA was extracted from blood samples using standard salting out method [19].

Genotyping

All patients and controls were screened for the presence of the TA insertion in the TATA box region of *UGT1A1* gene by PCR amplification, with the introduction of a fluorochrome label (NED™) in the reverse primer, as previously described [20,21]. Direct sequencing of the five exons, and the gtPBREM (c.–3576 to c.–3209) region of *UGT1A1* gene was done in independent sequencing reactions, using the respective forward and reverse primers (Table 1), and by the Dye Terminator Cycle Sequencing Kit (Perkin Elmer, Boston, USA). The protocol of cycling was performed on a DNA thermal cycler GeneAmp PCR System 9700 (Applied Biosystems, Foster City, USA). The

Table 1

Primers used for the amplification of the *UGT1A1* exons and promoter's regions.

<i>UGT1A1</i> region	Primer name	Sequence (5' → 3')	Reference
Promotor	UGT1A1-TAF	5'-TAACCTGGTGTATCGATGGTTTTC-3'	21
	UGT1A1-TAR	5'-ACAGCCATGGCCCTTTGCT-3'	
	UGT1A1-Reg	5'-CTA GCC ATT CTG GAT CCC TTG-3'	16
	UGT1A1-Reg	5'-TTTTGAGATCTGAGTTCTTTCACCTC-3'	
Exon 1	UGT1A1-1aF	5'-TATAAGTAGGAGAGGGCGAACC-3'	This study
	UGT1A1-1aR	5'-TCAAATTCAGGCTGCATG-3'	
	UGT1A1-1bF	5'-GGCCTCCCTGGCAGAAAG-3'	
Exon 2	UGT1A1-1bR	5'-ATGCCAAAGACAGACTCAAACC-3'	
	UGT1A1-2 F	5'-TCTATCTCAAACACGCATGCC-3'	20
	UGT1A1-2R	5'-GGCAGGGAAAAGCCAAATCTA-3'	
Exon 3	UGT1A1-3 F	5'-TTGCCAGTCTCAGAAGCCTT-3'	20
	UGT1A1-3R	5'-ATGCCCTTGCAGAAATTTGC-3'	
Exon 4	UGT1A1-4 F	5'TGCAAGGGCATGTGAGTAACA-3'	20
	UGT1A1-4R	5'-AAGCCAAGATTGCACCACTG-3'	
Exon 5	UGT1A1-5 F	5'-GAGGATTGTTTCATACCACAGG-3'	20
	UGT1A1-5R	5'-GCACCTCGGGGCTGATTAAT-3'	

resulting products were subsequently analyzed by automated capillary electrophoresis (ABI Prism 3130xl® Genetic Analyser, Applied Biosystems, Foster City, USA), using the GeneMapper v4.0 along with the molecular weight marker GeneScan ROXTM 500 Size Standard (Applied Biosystems, Foster City, USA).

Bioinformatics tools

For primers design, we used the Primer Express® software (Applied Biosystems, Foster City, USA). The analysis of sequence results were performed on SeqScape™ version 2.5, and Finch TV version 1.3 (<http://www.geospiza.com/finchtv>). Sequence analyses were carried out using the reference sequence (rs) from GenBank (accession number NM_000463.2) and amino acid reference sequence from GenePret (accession number P22309). The annotation of the mutations was done according to the recommendation of the Human Genome Variation Society. The homozygous and heterozygous individuals for the TA polymorphism c.–41_–40dupTA were categorized as [TA]₇/[TA]₇ and [TA]₇/[TA]₆, respectively, and those that were homozygous for the wild type allele were referred as [TA]₆/[TA]₆. The homozygous and heterozygous for the polymorphism c.3279T>G were categorized as G/G and T/G, respectively, and the homozygous for the wild type allele as T/T.

The potential pathogenicity of the new variants and the correspondent altered sequences were analyzed by Alamut software v.1.5 (Interactive Biosoftware, Rouen, France) that performs an *in silico* analysis using methods based on sequence homology, protein structure information and amino acid physicochemical properties. The methods used in this analysis were: Polymorphism Phenotyping (PolyPhen-2) [22]; Sorting Intolerant from Tolerant (SIFT) [23]; Align Grantham Variance-Grantham Difference (A-GVGD) [24]; Grantham score [25]; BLOSUM 62 matrix score (BLOSUM62) [26].

Statistical analysis

For statistical analysis, we used the Statistical Package for Social Sciences-SPSS version 18.0 for Windows (SPSS Inc., Chicago, USA). χ^2 or Fisher's exact tests were used to compare genotype allele frequencies between groups. Bilirubin, the dependent variable, was Log transformed to obtain a normalized data. Independent Student *t* test was used to compare total bilirubin values between male and female subjects. Multiple comparisons between groups were performed by one-way ANOVA supplemented with Tukey's HSD post hoc test. The pairwise (Lewontin's) *D'* and correlation coefficient (*r*²) values of linkage disequilibrium for the variation of the c.–3279T>G with the polymorphism c.–41_–40dupTA was

Table 2
Genotypes frequencies and TBL values of *UGT1A1* c.-41_–40dupTA and c.-3279T>G polymorphisms in GS patients and in controls.

Polymorphism	GS patients			Controls	
	% (n)	TBL ($\mu\text{mol/L}$)	% (n)	TBL ($\mu\text{mol/L}$)	p^*
c.-41_–40dupTA					
[TA] ₆ /[TA] ₆	0%(0)	–	46.6% (75)	7.5 \pm 3.2	<0.0001
[TA] ₆ /[TA] ₇	17.8%(8)	27.7 \pm 12.5	43.5% (70)	9.6 \pm 4.6 ^a	
[TA] ₇ /[TA] ₇	82.2%(37)	37.9 \pm 13.3	9.9% (16)	19.5 \pm 9.9 ^{b,c}	
c.3279T>G					
TT	0%(0)	–	33.5% (54)	7.7 \pm 3.1	<0.0001
TG	13.3%(6)	28.3 \pm 5.8	44.1% (71)	8.9 \pm 4.5	
GG	86.7%(39)	36.8 \pm 13.3 ^b	22.4% (36)	13.8 \pm 8.8 ^{b,c}	

TBL: total bilirubin levels that are presented as mean \pm standard deviation (SD); * p -value for differences in allelic distribution between controls and GS patients. p -value <0.05 for TBL between different genotypes: ^awild type vs. heterozygous; ^bhomozygous for the mutated alleles vs. heterozygous; ^chomozygous for the mutated alleles vs. wild type.

calculated using the web available tool: <http://www.oege.org/software/cubex/> [27]. Significance was accepted at $p < 0.05$.

Results

In this work, we first analyzed the impact of promoter's polymorphisms in TBL, considering the nine possible diplotypes. After that, we also evaluated the influence of coding region of *UGT1A1* variants in TBL. The impact of new variants in the *UGT1A1* gene was estimated by *in silico* methods.

Promoter *UGT1A1* variants

Results from genotyping revealed the presence of two promoter polymorphisms, c.-41_–40dupTA and c.-3279T>G, in both groups (GS patients and controls) with significant different allelic distribution (Table 2). Among GS patients, 82.2% ($n = 37$) were homozygous and 17.8% ($n = 8$) were heterozygous for the c.-41_–40dupTA allele. In the control group, we found that 9.9% ($n = 16$) were homozygous and 43.5% ($n = 70$) were heterozygous for this promoter variant, and 46.6% ($n = 75$) presented the six TA repeats [A(TA)₆TAA]. For the T>G transition located at c.-3279 nucleotide, we found that 86.7% ($n = 39$) were homozygous and 13.3% ($n = 6$) were heterozygous for the c.-3279T>G variant in GS patients. In the control group, we found that 33.5% ($n = 54$) were homozygous for the wild type allele, 44.1% ($n = 71$) were heterozygous and 22.4% ($n = 36$) were homozygous for the c.-3279T>G variant. The two polymorphisms were in Hardy-Weinberg equilibrium in control group. Results of linkage disequilibrium, for the association of the two promoter polymorphisms, showed that D' and r' values were 1.0 and 0.57, respectively, confirming existence of the close linkage the disequilibrium between them.

From the nine possible diplotype combinations, four were found in GS patients and six in controls. The frequencies of combined diplotypes in GS patients and controls are shown in Table 3.

Table 3
GS patients and controls diplotype frequencies and its relationship with TBL.

Diplotype (c.-41_–40dupTA/c.-3279T>G)	GS patients		Controls		p^a
	% (n)	TBL ($\mu\text{mol/L}$)	% (n)	TBL ($\mu\text{mol/L}$)	
[TA] ₆ /[TA] ₆ -TT	0% (0)	–	33.5% (54)	7.7 \pm 3.1	<0.0001
[TA] ₆ /[TA] ₆ -TG	0% (0)	–	11.2% (18)	6.9 \pm 3.4	
[TA] ₆ /[TA] ₇ -TG	8.9% (4)	27.8 \pm 7.4	32.9% (53)	9.6 \pm 4.9	
[TA] ₇ /[TA] ₇ -TG	4.4% (2)	29.1 \pm 0.5	0% (0)	–	
[TA] ₆ /[TA] ₆ -GG	0% (0)	–	1.9% (3)	8.4 \pm 3.8	
[TA] ₆ /[TA] ₇ -GG	8.9% (4)	34.8 \pm 7.6	10.6% (17)	9.5 \pm 4.3	
[TA] ₇ /[TA] ₇ -GG	77.8% (35)	37.9 \pm 13.3	9.9% (16)	19.5 \pm 9.9 ^b	

TBL: total bilirubin levels that are presented as mean \pm standard deviation (SD); ^a p -value for difference in total bilirubin levels between the several diplotypes; ^b p -value <0.05 for [TA]₇/[TA]₇-G/G TBL and other diplotypes.

Coding region of *UGT1A1* variants

Sequencing analysis of the coding regions of the *UGT1A1* gene allowed the identification of nine additional SNPs, all of them in a heterozygous state. From these, five were identified in GS patients and four in controls (Table 4). Seven of these variants were new c.538G>C, c.643A>G, c.814A>G, c.1156G>A; c.1211T>C; c.1423C>T; c.1475T>C (Fig. 1), and two were already described in literature, c.674T>G [6] and c.923G>A [28]. Moreover, three of these new variants were detected in GS patients (c.538G>C; c.1211T>C; c.1423C>T) and four in controls (c.643A>G; c.814A>G; c.1156G>A; c.1475T>C).

Amino acid conservation analysis was performed by using non-redundant alignment of homologous sequences, and the results from amino acid changes as well as the association with c.-41_40dupTA and c.-3279T>G, are described in Table 4. The new sequence variants detected were located at the intraluminal part of the C-terminal half of the UGT1A1 protein that includes the domain that binds the UDP glucuronic acid (UDPGA) donor substrate, in both GS patients (p.E180Q; p.V225G) and controls (p.I215V; p.M272V); at the donor-binding region (residues: 292–409), in GS patients (p.G308E; p.M404T; p.R475C) and in controls (p.V386I); and one at reticulum endoplasmic-anchoring transmembrane domain (residues: 491–507), in a healthy control (p.I492T). Only one exchanged amino acid, the glycine 308, was highly conserved among homologous sequences. The majority of others residues were moderately conserved and three were weakly conserved (Table 5). Data regarding the new variants showed that four of them were anticipated to be benign (p.E180Q; p.I215V; p.M272V; p.V386I), while three were expected to have an effect on protein function (p.M404T; p.475 C; p.I492T). Functional prediction of the consequence of these amino acid replacements in the UGT1A1 protein, performed by the different *in silico* methods, was in agreement for the variants p.M404T, p.R475C, and p.I215V (Table 5).

Additionally, we also detected four previously described polymorphic variants in GS patients and controls: c.864+89C>T (rs.17868341); c.997-86C>T (rs.34082659); c.997-82T>C (rs.12471326); and c.997-37T>C (rs.2471326).

Table 4
UGT1A1 gene variants and TBL in GS patients and controls.

	TBL (μmol/L)	cDNA variation	Effect on protein	Polymorphism c.-41_-40dupTA	Polymorphism c.-3279T>G	References
Patients						
1/GS	29.4	c.538G>C	p.E180Q	[TA] ₇ /[TA] ₇	T/G	This study
2/GS	27.7	c.674T>G	p.V225G	[TA] ₆ /[TA] ₇	T/G	[6]
3/GS	50.1	c.674T>G	p.V225G	[TA] ₇ /[TA] ₇	G/G	[6]
4/GS	54.2	c.923G>A	p.G308E	[TA] ₆ /[TA] ₇	G/G	[28]
5/GS	36.6	c.1211T>C	p.M404T	[TA] ₇ /[TA] ₇	G/G	This study
6/GS	51.3	c.1423C>T	p.R475C	[TA] ₆ /[TA] ₇	G/G	This study
Controls						
1/C	5.1	c.643A>G	p.I215V	[TA] ₆ /[TA] ₇	T/G	This study
2/C	10.1	c.814A>G	p.M272V	[TA] ₇ /[TA] ₇	G/G	This study
3/C	4.3	c.1156G>A	p.V386I	[TA] ₆ /[TA] ₆	T/T	This study
4/C	5.8	c.1475T>C	p.I492T	[TA] ₆ /[TA] ₇	T/G	This study

The annotation of the mutations was done according to the recommendation of the Human Genome Variation Society. The first A in the UGT1A1 start codon was considered as position 1. TBL: total bilirubin levels.

Impact of the UGT1A1 variants in total bilirubin levels

The median of TBL (mean ± SD) presented for GS patients was 36.0 ± 12.7 μmol/L (33.4 ± 15.8 μmol/L for females and 35.6 ± 13.1 μmol/L for males, *p* = 0.667), and 9.5 ± 5.8 μmol/L (9.2 ± 6.1 μmol/L for females, and 9.9 ± 4.5 μmol/L for males, *p* = 0.529) for controls. No significant differences were found in TBL, according to gender, between GS patients and controls.

In GS patients, the presence of homozygosity for the two promoter polymorphisms, c.-41_-40dupTA and c.-3279T>G, were

associated with significant increased TBL. Analyzing TBL between the different dyplotypes of GS patients, we didn't find statistically significant differences between the four combinations (Table 3). Increments in TBL were observed in GS patients with a diplotype heterozygous for c.-41_-40dupTA and homozygous for c.3279T>G, but without statistical significance (34.8 ± 7.6 vs. 27.8 ± 7.43 mol/L; *p* = 0.254). The same was observed for diplotype [TA]₇/[TA]₇-T/G, as compared to diplotype [TA]₇/[TA]₇-G/G (29.1 ± 0.5 vs. 37.9 ± 13.3 μmol/L; *p* = 0.275). Only one GS patient (2/GS, Table 2) was identified in heterozygosity for both promoters polymorphism; the

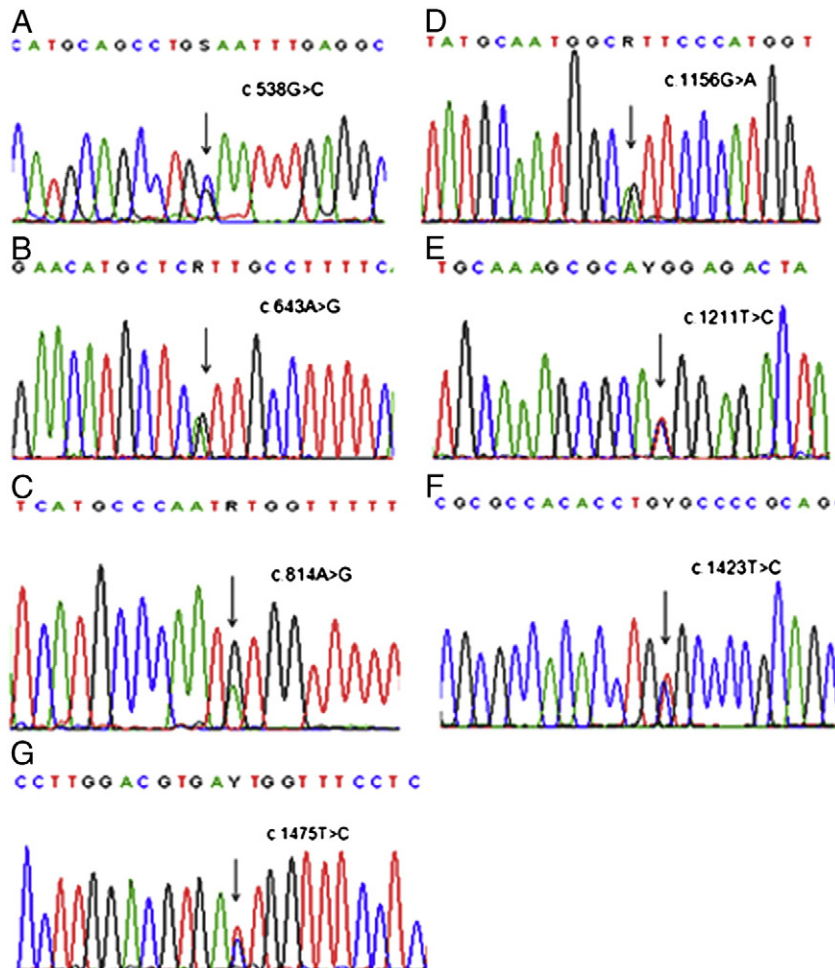


Fig. 1. Partial sequences of exons 1, 2, 4 and 5. Arrows indicate the position of nucleotide substitution at the seven novel identified variants. (A) – p.E180Q; (B) – p.I215V; (C) – p.M272V; (D) – p.V386I; E – p.M404T; (F) – p.R475C; (G) – p.I492T.

Table 5
Functional predictions of *UGT1A1* variants by three algorithms based in homology (Polyphen-2; SIFT; A-GVGD) and two based in amino acid physiochemical properties (Grantham Distance and BLOSUM62 matrix score).

Variant	Exon	Domain	Amino acid conservation	<i>In silico</i> evaluation (prediction and scores)				
				Polyphen-2 ^a	SIFT ^b	A-GVGD ^c	Grantham Distance ^d	BLOSUM62 ^e
GS patients								
p.E180Q	1	Aglycone-recognition	Weakly conserved	Possibly damaging	Tolerated (0.25)	C0	Tolerated (29)	Tolerated (2)
p.V225G ^f	1	Aglycone-recognition	Weakly conserved	Benign	Deleterious (0.03)	C0	Deleterious (109)	Deleterious (−3)
p.G308E ^f	2	Donor-binding	Highly conserved	Probably damaging	Deleterious (0.0)	C0	Deleterious (98)	Deleterious (−2)
p.M404T	4	Donor-binding	Moderately conserved	Probably damaging	Deleterious (0.03)	C0	Deleterious (81)	Deleterious (−1)
p.R475C	5	Donor-binding	Moderately conserved	Probably Damaging	Deleterious (0.02)	C0	Deleterious (180)	Deleterious (−3)
Controls								
p.I215V	1	Aglycone-recognition	Weakly conserved	Benign	Tolerated (1.0)	C0	Tolerated (29)	Tolerated (3)
p.M272V	1	Aglycone-recognition	Moderately conserved	Probably damaging	Tolerated (0.27)	C0	Tolerated (21)	Tolerated (1)
p.V386I	4	Donor-binding	Moderately conserved	Probably damaging	Tolerated (0.47)	C0	Tolerated (29)	Tolerated (3)
p.I492T	5	Transmembrane segment	Moderately conserved	Probably damaging	Tolerated (0.24)	C0	Deleterious (89)	Deleterious (−1)

^a Qualitative classification given by Polyphen: probably damaging, possibly damaging and benign.

^b Sort Intolerant from Tolerant, a tolerance index ≤ 0.05 is considered to be deleterious.

^c Align GVGD classifiers range from C65 (most likely deleterious) to C0 (most likely neutral).

^d Grantham Distance classified amino acid substitutions as either tolerated (score < 60) or deleterious (score ≥ 60).

^e BLOSUM62 (Blok Substitution Matrix), scores ≥ 0 were considered tolerated.

^f Variants already described in literature.

others were homozygous for one or both mutated alleles in these two promoter regions, beyond the variants identified in the coding regions, but all of them presented high values of bilirubin (Table 4). The polymorphic variants already described in literature (c.864+89C>T, c.997-86C>T, c.997-82T>C, and c.997-37T>C) seem to have no influence in TBL.

In healthy subjects, significant differences in TBL were obtained comparing individuals homozygous for the wild type alleles with those heterozygous for the c.-41_-40dupTA polymorphism ($p < 0.03$). However, when we compared individuals homozygous for the wild type alleles with those heterozygous for the c.-3279T>G transition, no significant differences were found ($p = 0.459$). The presence of homozygosity for the two promoter polymorphisms, c.-41_-40dupTA and c.-3279T>G, were associated with significantly increased TBL (Table 3). Nevertheless, from sixteen homozygous for the TA duplication, which were also homozygous for the c.-3279G allele, only six of them (37%) had normal bilirubin concentrations (TBL $< 17.1 \mu\text{mol/L}$).

Discussion

Several studies aimed to clarify the influence of genetic and non-genetic factors in bilirubin levels. GS is one of the most prevalent congenital metabolic disorders in the population, which is mainly associated with defects in *UGT1A1* gene [7,9,12,29]. The two major polymorphisms associated to increased TBL in GS Caucasian patients are the c.-41_-40dupTA and the c.-3279T>G polymorphisms, [9-16,30-33]. In this work, we analyzed the influence of *UGT1A1* gene variants on TBL in healthy subjects and in GS patients. This analysis was performed independently in the two groups, due to the differences in age between them.

According to our results and in agreement with literature [9,10,13], homozygosity for the c.-41_-40dupTA allele is strongly associated with higher bilirubin levels. The c.-3279T>G polymorphism has been described as another important factor for the development of hyperbilirubinemia [14-16,30-33]. Experimental observations suggested that this variant, located at the phenobarbital-responsive enhancer module, contributes to a decrease of 62% of the transcriptional activity of *UGT1A1* gene [13]. Moreover, a recent *in vitro* study suggests a possible combined effect of these two polymorphisms (c.-41_-40dupTA and c.-3279T>G) on transcriptional activity of *UGT1A1* gene, which further contributes for GS phenotype [34]. Our work showed that, both healthy individuals and GS patients who have double homozygosity for these two variants present increased TBL. In fact, we found a linkage disequilibrium between

these two polymorphisms [Lewontin's ($D' = 1$) and correlation coefficient ($r^2 = 0.57$)], which is in accordance with previous studies done by Borucki et al. ($D = 1$; $r^2 = 0.67$) [30], Innocenti et al. ($D = 1$; $r^2 = 0.65$) [31], Jirsa et al. ($D = 1$; $r^2 = 0.66$) [32], and Ferraris et al. ($D = 0.89$; $r^2 = 0.57$) [33]. This close linkage disequilibrium makes difficult the evaluation of the independent contribution of each polymorphism in TBL.

In this study, the impact of the two polymorphisms in TBL, in healthy individuals and GS patients, was first analyzed independently. The differences in TBL were more significant between the three genotypes of the c.-41_-40dupTA polymorphism than between the three genotypes of c.-3279T>G polymorphism (Table 2).

No significant differences were observed between homozygous and heterozygous for the mutated alleles, in GS patients. In four of our GS patients, heterozygous for one or both of the two promoter polymorphisms, we detected another variant at the exonic region (Table 4).

The low penetrance of these two variants was evident in our control group as we found sixteen homozygous for both variants and six of them had normal bilirubin levels. In GS patients and controls, we didn't find statistically significant differences in TBL between most of diplotypes, a result that may be a consequence of the small sample size of most of the diplotypes observed. However, we found that the diplotype [TA]₇/[TA]₇-G/G is associated with the highest TBL.

Variants in the promoter and coding regions of *UGT1A1* may co-exist in GS patients and healthy controls [6,7,9,28,35-38]. By direct sequencing of these regions, we identified for the first time seven variants, three in GS patients (c.538G>C; c.1211T>C; c.1423C>T) and four in controls (c.643A>G; c.814A>G; c.1156G>A; c.1475T>C). The sequence variant c.1211T>C (p.M404T) is positioned at the donor-binding site of the *UGT1A1* protein. Close to this position, two other altered residues, the c.1198A>G (N400D) [36] and c.1207 C>T (R403C) transitions [37], were described and associated with Crigler-Najjar syndrome type II. The variant c.1423 C>T (p.R475C), also located at donor-binding site, might alter the tertiary structure of the protein. The variant c.923 G>A (p.G308E), initially described by Erps and co-workers [28] is located in a highly conserved residue, and results in lack of enzyme activity. This substitution was described as the second most frequent variant of *UGT1A1* gene associated to GS in the Portuguese population [20].

In silico analysis, performed by four different methods, predicted in agreement that three of the amino acid changes (p.M404T, p.R475C and p.G308E) had a deleterious effect on *UGT1A1* activity. Only the A-GVGD score gave a different prediction, classifying these variants and all the others as most likely neutral, even for the

p.G308E transition, described as an inactivating alteration. Another variant with potentially negative impact in UGT1A1 activity was p.V225G, in exon 1, according to the consensual results of three *in silico* methods. This variant was described in association with Crigler-Najjar syndrome type II [6]. These divergences could be due to the fact that A-GVGD has better specificity than sensitivity (low risk of false positive pathogenic prediction), while SIFT and PolyPhen provide better sensitivity (low risk of false negative neutral prediction) [39]. We also identified p.V225G variant in two GS patients. The threonine instead of an isoleucine, at residue 492 (exon 5), was detected in a control individual that was also heterozygous [TA]₆/[TA]₇-T/G for the promoter polymorphisms. Three algorithms predict a deleterious outcome of this substitution; however, the individual presented normal levels of total bilirubin.

Four variants were predicted to be benign, one in a GS patient (p.E180Q) and three in controls (p.I215V, p.V386I, p.I492T). The p.E180Q transition could be a benign variant, according to four algorithms, with the exception of Polyphen-2, which classified this variant as possibly damaging, as it can affect the protein function or structure. Regarding the three variants found in the control sample, one was predicted to be benign by all algorithms (p.I215V) and the other two (p.V386I, p.I492T) were classified as benign by four algorithms. These variants were not associated to hyperbilirubinemia. The potential effect of 63 known UGTs non-synonymous (ns) SNPs on protein function was previously evaluated [40] and data obtained showed that SIFT and Polyphen had a correct prediction rate of 51.1% and 66.7%, respectively. However, we noticed, in the present study, that SIFTS and Polyphen were in divergence, comparing to the other methods, in five of the nine identified variants. This suggests the needs for additional *in silico* methods with a strong predictor of nsSNP pathogenicity. Nevertheless, together with the studies of *in silico* analysis, it is important to know the allelic phase of the nucleotide substitution with two promoter polymorphisms of the *UGT1A1* gene.

In the five patients carrying predicted deleterious variants identified in the coding region, the mean TBL concentrations was not significantly higher, when compared to the majority of the individuals with the [TA]₇/[TA]₇-G/G diplotype (42.1 ± 9.3 vs. 35.9 ± 12.4 $\mu\text{mol/L}$; $p = 0.230$), except three patients (3/SG, 4/SG and 6/SG) who had some of the highest TBL values observed. From the eight patients heterozygous for the [TA]₇ (four heterozygous for G/T and four homozygous for G/G), three of them had another variant, potentially deleterious, in the coding region. Regarding the alteration p.V225G, patient 2/SG was a compound heterozygous for the TA duplication and c.-3279T>G polymorphism, while patient 3/GS was homozygous for both mutated alleles [TA]₇/[TA]₇-G/G. In these two cases, we can observe that the presence of a homozygous diplotype increases bilirubin median values (patient 2/SG had a TBL value of 27.7 $\mu\text{mol/L}$ and patient 3/SG had 50.1 $\mu\text{mol/L}$). The higher TBL observed in 3/GS patient seems to be conditioned by the presence of the [TA]₇/[TA]₇-G/G diplotype.

In conclusion, our data showed that TBL are mainly determined by the presence of the TA duplication in the repetitive TATA-box sequence of the gene promoter, and that the c.-3279T>G variant is an additional factor for the development of hyperbilirubinemia. Moreover, our results also revealed that the presence of variants in the *UGT1A1* coding region seems to be associated with increased bilirubin levels, and therefore, linked with GS. Thus, the genetic diagnosis for GS should include the determination of both promoter polymorphisms, and the genetic analysis of the entire coding region namely when the homozygosity for both polymorphisms is not observed. Additionally, we described for the first time seven new variants in the coding region of the gene, expanding the spectrum of the known *UGT1A1* variants.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgments

A PhD grant (SFRH/BD/42791/2007) attributed to Carina Rodrigues, from Fundação para a Ciência e Tecnologia (FCT) and Fundo Social Europeu (FSE), supported this work.

References

- [1] M. Clementi, E. Di. Gianantonio, L. Fabris, et al., Inheritance of hyperbilirubinemia: evidence for a major autosomal recessive gene, *Dig. Liver Dis.* 4 (2007) 351–355.
- [2] S. Sinafi, D. Clarke, B. Burchell, 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 (1994) 233–240.
- [3] C.P. Strassburg, S. Kalthoff, U. Ehmer, Variability and function of family 1 uridine-5'-diphosphate glucuronosyltransferases (UGT1A), *Crit. Rev. Clin. Lab. Sci.* 45 (2008) 485–530.
- [4] M.A. Malfatti, E.A. Ubick, J.S. Felton, 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 (2005) 2019–2028.
- [5] S. Gantla, C.T. Bakker, B. Deocharan, et al., Splice-site mutations: a novel genetic mechanism of Crigler-Najjar syndrome type1, *Am. J. Hum. Genet.* 62 (1998) 585–592.
- [6] A. Iolascon, A. Meloni, B. Coppola, M.C. Rosatelli, Crigler-Najjar syndrome type II resulting from three different mutations in the bilirubin uridine 5'-diphosphate-glucuronosyltransferase (UGT1A1) gene, *J. Med. Genet.* 37 (2000) 712–713.
- [7] E. Costa, Hematologically important mutations: bilirubin UDP-glucuronosyltransferase gene mutations in Gilbert and Crigler-Najjar syndromes, *Blood Cells Mol. Dis.* 36 (2006) 77–80.
- [8] Q.H. Gong, J.W. Cho, T. Huang, et al., Thirteen UDPglucuronosyltransferase genes are encoded at the human UGT1 gene complex locus, *Pharmacogenetics* 11 (2001) 357–368.
- [9] P.J. Bosma, J.R. Chowdhury, C. Bakker, et al., The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome, *N. Engl. J. Med.* 333 (1995) 1171–1185.
- [10] E. Beutler, T. Gelbart, A. Bemina, Racial variability in the UDPglucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? *Proc. Natl. Acad. Sci. U. S. A.* 95 (1998) 8170–8174.
- [11] Y. Maruo, C. D'Addario, A. Mori, et al., Racial variability in haplotype frequencies of UGT1A1 and glucuronidation activity of a novel single nucleotide, *Drug Metab. Dispos.* 33 (2005) 458–465.
- [12] Y. Maruo, C. D'Addario, A. Mori, et al., Inheritance of hyperbilirubinemia: evidence for a major autosomal recessive gene, *Dig. Liver Dis.* 39 (2007) 351–355.
- [13] J. Sugatani, K. Yamakawa, K. Yoshinari, et al., Identification of a defect in the UGT1A1 gene promoter and its association with hyperbilirubinemia, *Biochem. Biophys. Res. Commun.* 292 (2002) 492–497.
- [14] Y. Maruo, C. D'Addario, A. Mori, et al., Two linked polymorphic mutations (A(TA)7TAA and T-3279G) of UGT1A1 as the principal cause of Gilbert Syndrome, *Hum. Genet.* 115 (2004) 525–526.
- [15] V. Servedio, M. d'Apolito, N. Maiorano, et al., A polymorphic mutation, c.-3279T>G, in the UGT1A1 promoter is a risk factor for neonatal jaundice in the Malay population, *Pediatr. Res.* 67 (2010) 401–406.
- [16] E. Costa, E. Vieira, R. dos Santos, The polymorphism c.-3279T>G in the phenobarbital-responsive enhancer module of the bilirubin UDP-glucuronosyltransferase gene is associated with Gilbert syndrome, *Clin. Chem.* 51 (2005) 2204–2206.
- [17] M. Sanpietro, A. Iolascon, Molecular pathology of Crigler-Najjar type I and II and Gilbert's syndromes, *Haematologica* 84 (1999) 150–157.
- [18] D.G. Nathan, S.H. Orking, N. Oski's, *Haematology of Infancy and Childhood*, 5th edition, RWB Sanders Company, Philadelphia, 1998, pp. 103–146.
- [19] S.A. Miller, D.D. Dykes, H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, *Nucleic Acids Res.* 16 (1988) 1215.
- [20] E. Costa, E. Vieira, M. Martins, et al., Analysis of the UDP-glucuronosyltransferase gene in Portuguese patients with a clinical diagnosis of Gilbert and Crigler-Najjar syndromes, *Blood Cells Mol. Dis.* 36 (2006) 91–97.
- [21] J.D. Bancroft, B. Kreamer, G.R. Gourley, Gilbert syndrome accelerates development of neonatal jaundice, *J. Pediatr.* 132 (1998) 656–660.
- [22] V. Ramensky, P. Bork, S. Sunyaev, Human non-synonymous SNPs: server and survey, *Nucleic Acids Res.* 30 (2002) 3894–3900.
- [23] P.C. Ng, S. Henikoff, SIFT. Predicting amino acid changes that affect protein function, *Nucleic Acids Res.* 31 (2003) 3812–3814.
- [24] S.V. Tavtigian, A.M. Deffenbaugh, L. Yin, et al., Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral, *J. Med. Genet.* 43 (2006) 295–305.
- [25] R. Grantham, Amino acid difference formula to help explain protein evolution, *Science* 185 (1985) 862–864.
- [26] S. Henikoff, J.G. Henikoff, Position-based sequence weights, *J. Mol. Biol.* 243 (1994) 574–578.
- [27] T.R. Gaunt, S. Rodriguez, I.N. Day, Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for linkage disequilibrium analyses and a web tool 'CubeX', *BMC Bioinform.* 8 (2007) 428–436.
- [28] T. Erps, J.K. Ritter, J.H. Hersh, et al., Identification of two single base substitutions in the UGT1 gene locus, which abolish bilirubin uridine diphosphate glucuronosyltransferase activity *in vitro*, *J. Clin. Invest.* 93 (1994) 564–570.

- [29] C. Rodrigues, E. Costa, E. Vieira, et al., in press. Bilirubin is mainly dependent on UGT1A1 polymorphisms, hemoglobin, fasting time and body mass index, *Am. J. Med. Sci.*
- [30] K. Borucki, C. Weikert, E. Fisher, et al., Haplotypes in the UGT1A1 gene and their role as genetic determinants of bilirubin concentration in healthy German volunteers, *Clin. Biochem.* 42 (2009) 1635–1641.
- [31] F. Innocenti, C. Grimsley, S. Das, et al., Haplotype structure of the UDP-glucuronosyltransferase 1A1 promoter in different ethnic groups, *Pharmacogenetics* 12 (2002) 725–733.
- [32] M. Jirsa, J. Petrasek, L. Vitek, Linkage between A(TA)₇TAA and –3279TNG mutations in UGT1A1 is not essential for pathogenesis of Gilbert syndrome, *Liver Int.* 26 (2006) 1302–1303.
- [33] A. Ferraris, G. D'Amato, V. Nobili, et al., Combined test for UGT1A1-3279TNG and A(TA)_nTAA polymorphisms best predicts Gilbert's syndrome in Italian pediatric patients, *Genet. Test.* 10 (2006) 121–125.
- [34] K. Matsui, Y. Maruo, H. Sato, Y. Takeuchi, Combined effect of regulatory polymorphisms on transcription of UGT1A1 as a cause of Gilbert syndrome, *BMC Gastroenterol.* 10 (2010) 57–68.
- [35] N. Kaniwa, K. Kurose, H. Jinno, et al., Racial variability in haplotype frequencies of UGT1A1 and glucuronidation activity of a novel single nucleotide polymorphism 686CNT (P229L) found in an African-American, *Drug Metab. Dispos.* 33 (2005) 458–465.
- [36] H. Sato, Y. Adachi, O. Koiwai, The genetic basis of Gilbert's syndrome, *Lancet* 34 (1996) 557–558.
- [37] P. Labrune, A. Myara, J. Chalas, et al., Association of a homozygous (TA)₈ promoter polymorphism and a N400D mutation of UGT1A1 in a child with Crigler-Najjar type II syndrome, *Hum. Mutat.* 20 (2002) 399–401.
- [38] V. Servedio, M. d'Apolito, N. Maiorano, et al., Spectrum of UGT1A1 mutations in Crigler-Najjar (CN) syndrome patients: identification of twelve novel alleles and genotype-phenotype correlation, *Hum. Mutat.* 25 (2005) 325–334.
- [39] S.V. Tavtigan, M.S. Greenblatt, F. Lesueur, et al., In silico analysis of missense substitutions using sequence-alignment based methods, *Hum. Mutat.* 29 (2008) 1327–1336.
- [40] Y.M. Di, E. Chan, M.Q. Wei, J.P. Liu, et al., Prediction of deleterious non-synonymous single-nucleotide polymorphisms of human uridine diphosphate glucuronosyltransferase genes, *AAPS J.* 11 (2009) 469–480.