

ORIGINAL ARTICLE

Common and rare single nucleotide polymorphisms in the *LDLR* gene are present in a black South African population and associate with low-density lipoprotein cholesterol levels

Tertia van Zyl¹, Johann C Jerling¹, Karin R Conradie¹ and Edith JM Feskens²

The LDL receptor has an essential role in regulating plasma LDL-C levels. Genetic variation in the *LDLR* gene can be associated with either lower or moderately raised plasma levels of LDL-C, or may cause familial hypercholesterolemia. The prevalence of single-nucleotide polymorphisms (SNPs) in the *LDLR* in the black South African population is not known and therefore, we aimed to determine the genotypic variation of the *LDLR* in the study population as well as to define the association of the different genotypes with plasma LDL-C levels. A random selection of 1860 apparently healthy black South African volunteers aged 35–60 years was made in a cross-sectional study. Novel SNPs were identified in a subset of 30 individuals by means of automated sequencing before screening the entire cohort by means of the Illumina VeraCode GoldenGate Genotyping Assay on a BeadXpress Reader system. Twenty-five SNPs were genotyped, two of which were novel. A very rare SNP, rs17249141, in the promoter region was significantly associated with lower levels of LDL-C. Four other SNPs (rs2738447, rs14158, rs2738465 and rs3180023) were significantly associated with increased levels of LDL-C. We can conclude that some of the various SNPs identified do indeed associate with LDL-C levels.

Journal of Human Genetics (2014) 59, 88–94; doi:10.1038/jhg.2013.123; published online 28 November 2013

Keywords: African population; haplotypes; LDL-C; LDLR

INTRODUCTION

The LDL receptor is a cell surface glycoprotein, which has a focal role in regulating low-density lipoprotein cholesterol (LDL-C) levels in humans, as it is the major protein involved in cholesterol transport. The receptor mediates the specific binding, uptake and degradation of plasma lipoproteins containing either apolipoprotein B or apoE. Remnants of triglyceride-rich lipoprotein metabolism (also known as intermediate-density lipoprotein), which are the precursors of plasma LDL and LDL itself, are recognized by the LDL receptor. It regulates plasma cholesterol by mediating the endocytosis of LDL-C by means of clathrin-coated pits.¹ An elevated LDL-C level is a major risk factor for the development of coronary artery disease through the role it has in the etiology of atherosclerosis.² The LDL-C levels of the black South African population are thought to be historically low³ and the prevalence of coronary artery disease is minimal.⁴ Genetic variation influences LDL-C levels in a population and 50% of the interindividual variability is due to this genetic component.⁵ A large number of genetic variants have been identified that cause a range of LDL-C levels in various populations.⁶ Some variants, such as the single-nucleotide polymorphism (SNP) found in the promoter region of the *LDLR* gene, are able to increase the promoter activity by

~50% and cause lower LDL-C levels.⁷ The question arises whether there are genetic variants present in the South African Prospective Urban and Rural Epidemiological (PURE) study population that contribute to the different levels of LDL-C, and to what degree these variants associate with LDL-C levels. The aim of this study was thus to determine whether single-nucleotide polymorphisms (SNPs) in the *LDLR* gene are present in the study population and to what extent these variants associate with LDL-C.

MATERIALS AND METHODS

Study population

The study population is part of the South African PURE study and consists mainly of self-identified Tswana-speaking individuals. The PURE study is a prospective cohort study that follows changes in lifestyle, risk factors and chronic disease by means of periodic standardized data collection in two urban and two rural areas in the NorthWest Province of South Africa over a period of 12 years. The central hypothesis of the study states that lifestyle (physical activity and nutrition) and psychosocial transitions secondary to urbanization and industrialization lead to behaviors that cause the development of chronic noncommunicable disease, such as CVD, diabetes, high blood pressure and obesity. Two thousand eligible volunteers between the ages of 35 and 60 years

¹Centre of Excellence for Nutrition, North-West University, Potchefstroom, South Africa and ²Division of Human Nutrition, Wageningen University, Wageningen, The Netherlands
Correspondence: T van Zyl, Centre of Excellence for Nutrition, North-West University, 11 Hoffman Street, Potchefstroom 2531, South Africa.
E-mail: tertia.vanzyl@nwu.ac.za

Received 3 October 2013; revised 18 October 2013; accepted 30 October 2013; published online 28 November 2013

were randomly selected. To avoid bias from population stratification, the urban and rural cohorts were matched for ethnicity, gender and age. The design of the PURE study has been described previously in greater detail.⁸ DNA samples for 1860 of the volunteers were available for analysis.

Biochemical analysis

Blood was sampled after 8 h of fasting. Serum lipid (high-density lipoprotein and triglyceride) measurements were carried out by a Sequential Multiple Analyzer Computer using a Konelab™ auto analyzer (Thermo Fisher Scientific, Vantaa, Finland). Plasma LDL-C levels were determined by using the Friedewald equation ($LDL-C = TC - HDL-C - (TG)/2.17$)⁹ and missing values were assigned to individuals with a triglyceride measurement of $>4.5 \text{ mmol l}^{-1}$.

DNA analysis

Informed consent was obtained from each participant in the PURE study for the collection of blood samples and ethical approval for the study was given by the Ethics Committee of the North-West University (Ethics number: NWU-00016-10-A1). DNA isolation was performed by using Qiagen Flexigene DNA extraction kits (QIAGEN, Valencia, CA, USA; catalog number 51206).

A sub-sample of 30 randomly chosen individuals was used to identify known and novel mutations in the exons and flanking regions of the *LDLR* by means of Sanger sequencing. Thirty individuals would give a total sample size of 60 chromosomes and, as communicated by Collins and Schwartz¹⁰, using 40–65 chromosomes will detect polymorphisms at a frequency of 0.05 with a power of 0.80–0.95. Polymerase chain reaction amplification of the proximal promoter region, the exons and the exon–intron boundaries of the *LDLR* gene was performed on an iCycler (Bio-Rad, Hercules, CA, USA). Primers were designed to amplify and sequence the various regions of the *LDLR*, and these are described in supplementary documentation (Supplementary Table 2). NucleoFast 96 PCR Clean-Up Kits from Machery-Nagel (Düren, Germany) were used for the clean-up of the polymerase chain reaction product prior to performing the sequencing process. Automated cycle sequencing was performed according to the basic principles of Sanger sequencing.¹¹ The standard ABI Prism, BIGDye Terminator version 3 Ready Reaction Cycle Sequencing Kit supplied by Applied Biosystems (Carlsbad, CA, USA) was used as a platform for the sequencing reactions. The purified product underwent electrophoresis on a 50 cm capillary either on an ABI3730xl DNA analyzer or an ABI3120xl genetic analyzer according to standard protocols.

Genotyping

The sequences that were generated were examined for the presence of variants using the CLC DNA Workbench version 6.0.2. (Aarhus, Denmark). Variants were confirmed by repeat sequencing in the opposite direction. Seventy variants were identified within both the coding and noncoding areas of the *LDLR* gene, including one insertion (a 1 bp insertion, c.*1218_1219Gins, in the 3' untranslated region (3'UTR) not previously reported), one deletion (rs17249078) and 68 SNPs (these are available in Supplementary Table 1). Five of the SNPs were novel and not previously reported. Genotyping of the population was done by means of the Illumina VeraCode GoldenGate Genotyping Assay on a BeadXpress platform using the VeraCode technology. All 68 SNPs (rs numbers for known SNPs and sequences for novel SNPs) were sent through to Illumina Technical Support for evaluation using the Assay Design Tool. Each SNP was scored (varying from 0–1) by the Assay Design Tool based on the compatibility with successful GoldenGate genotyping. SNPs with a score above 0.4 were chosen for genotyping. Fourteen SNPs fell out for selection because they were located less than 60 bp from another SNP (this is a prerequisite for the assay design) and 10 SNPs had a design score of less than 0.4. Twenty-nine SNPs were then selected on the basis of their frequency in the sub-sample, novelty and whether they were previously reported to associate with LDL-C. A further four SNPs (rs3745677, novel 5, rs2304182, rs5742911) failed during the execution of the assay on the BeadXpress platform. The study population was finally genotyped for 25 SNPs (of which two were novel and not previously reported). SNPs with a gene call score of >0.5 and samples with a call rate of ≥ 0.9 were included in the analysis. The A of the initiating

ATG codon is designated as nucleotide c.1 and the reference sequence *LDLR*: NM_001195798.1 was used.

Haplotype frequencies

Haplotype frequencies were determined by means of the software package Haploview version 4.2.25 (Broad Institute, Cambridge, MA, USA) to determine the haploblocks. Markers with a Hardy–Weinberg *P*-value of ≥ 0.05 and a minor allele frequency of $\geq 1\%$ were included in the analysis. Association analyses were conducted using LDL-C of $\leq 3 \text{ mmol l}^{-1}$ for controls and $>3 \text{ mmol l}^{-1}$ for cases.

Measurements

The anthropometric measurements such as weight (kilograms), height (meters) and body mass index (body mass index = weight in kilograms per height² in meters) were supervised by a level 3 anthropometrist according to guidelines adopted at the Airlie Conference sponsored by the National Institutes of Health. Standardized and calibrated equipment was used for measurements on volunteers wearing minimal clothing.¹²

Diet

Dietary intakes of the volunteers were determined by means of previously developed¹³ and validated¹⁴ culturally sensitive quantified food frequency questionnaires. The nutrient intakes were analyzed by using the Foodfinder3 program (Medical Research Council, Tygerberg, South Africa), which is based on the South African food composition tables.¹⁵

Statistical analysis

Descriptive statistics on the study population were computed on the statistical program SPSS version 20 and are expressed in means and s.d. The observed genotype frequencies were tested for deviation from Hardy–Weinberg equilibrium by using the χ^2 test of independence. Analyses of variance were used to test association between different genotypes and LDL-C levels. Adjustments for the confounders, age, gender and body mass index were made in all analyses. The Bonferroni correction was used to deal with the issue of multiple testing. *P*-values of 0.05 or lower were considered statistically significant.

RESULTS

The characteristics of the study population are shown in Table 1. The mean LDL-C levels for the men were 2.70 mmol l^{-1} (s.d. 1.12) and for the women, 3.06 mmol l^{-1} (s.d. 1.17). The mean age for both men and women was similar at 49.9 years (s.d. 10.4) and 49.1 years (s.d. 10.4), respectively. Twenty-five of the 70 variants identified by bidirectional sequencing of polymerase chain reaction products were

Table 1 Population characteristics

Variables	n	Men mean (s.d.)	n	Women mean (s.d.)
Age (years)	696	49.9 (10.4)	1168	49.1 (10.4)
Total cholesterol (mmol l^{-1})	696	4.79 (1.31)	1168	5.13 (1.39)
LDL-C (mmol l^{-1})	696	2.70 (1.12)	1168	3.06 (1.17)
HDL-C (mmol l^{-1})	696	1.58 (0.65)	1168	1.48 (0.62)
TG (mmol l^{-1})	696	1.15 (0.58)	1168	1.33 (0.70)
Energy intake (kJ)	675	8039 (3762)	1133	7231 (3373)
%fat of TE	675	23.17 (7.25)	1133	24.55 (7.72)
%SFA of TE	675	5.53 (2.50)	1133	5.93 (2.78)
<i>BMI categories:</i>				
1 ($<19 \text{ kg m}^{-2}$)	196	17 (1.15)	130	16.8 (1.25)
2 ($19\text{--}24.99 \text{ kg m}^{-2}$)	373	20.9 (1.70)	390	22 (1.85)
3 ($25\text{--}29.99 \text{ kg m}^{-2}$)	68	26.8 (1.18)	246	27.4 (1.48)
4 ($>30 \text{ kg m}^{-2}$)	22	34.1 (3.63)	338	35.5 (4.81)

Abbreviations: BMI, body mass index; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; SFA, saturated fatty acids; TE, total energy; TG, triglycerides.

finally genotyped for this study population. Two of these variants were novel and have not been previously reported.

The European and African populations from the 1000 Genomes Project were used to compare the minor allele frequency of the study population (Table 2). The following rare variants were absent in the European population: rs17249141, rs148698650, rs116405216 and rs72658862. All four previously mentioned variants were present in the study population but, according to the 1000 Genomes database, rs148698650 is not present in the African populations. Rare and common variants such as rs17242759, rs10423288, rs11669576, rs5929, rs4508523, rs6413503, rs3826810 and rs2738464 have a minor allele frequency that is greater in the study population and the African population than in the European population. Variants with a minor allele frequency greater than that of the study population and the other African populations that are present in the European population are as follows: rs2228671, rs5930, rs2738447, rs688, rs5925, rs14158, rs1433099 and rs3180023. Three common variants, rs1569372, rs5927 and rs2738465, have similar minor allele frequencies across the three different populations.

Variant c.-314 C→T (rs369850745) was found in the promoter region of the gene but the association with LDL-C levels was not statistically significant ($P=0.85$) (Table 3). The other novel SNP found in intron 15, c.2311+9 T→G (rs369402076), was also not associated with LDL-C levels. For both variants, only heterozygote carriers were detected and both variants were in HWE. Both these

Table 2 Minor allele frequency of the various SNPs in *LDLR* in different populations

SNP	Base pair	Minor (major) allele	Pure population frequency	MAF 1000 genomes AFR ^a	MAF 1000 genomes EUR ^a
rs369850745 ^b	11199911	T (C)	0.01	NA	NA
rs17249141	11200008	T (C)	0.01	0.02	0.00
rs17242759	11200309	A (C)	0.09	0.06	0.01
rs2228671	11210912	T (C)	0.03	0.02	0.09
rs10423288	11215846	C (T)	0.14	0.16	0.02
rs148698650	11218079	A (G)	0.01	0.00	0.00
rs116405216	11221324	A (G)	0.06	0.02	0.00
rs11669576	11222300	A (G)	0.16	0.24	0.05
rs72658862	11223929	T (C)	0.06	0.02	0.00
rs5930	11224265	A (G)	0.14	0.21	0.41
rs1569372	11224491	G (A)	0.42	0.29	0.47
rs5929	11226800	T (C)	0.10	0.12	0.04
rs4508523	11226944	T (C)	0.44	0.46	0.12
rs2738447	11227480	A (C)	0.12	0.16	0.43
rs688	11227602	T (C)	0.04	0.06	0.47
rs5925	11230881	C (T)	0.19	0.17	0.45
rs5927	11233941	A (G)	0.27	0.35	0.27
rs369402076 ^b	11234029	G (T)	0.02	NA	NA
rs6413503	11241904	A (G)	0.11	0.16	0.01
rs14158	11242044	A (G)	0.11	0.12	0.22
rs3826810	11242133	A (G)	0.13	0.13	0.03
rs2738464	11242307	G (C)	0.32	0.38	0.13
rs2738465	11242469	A (G)	0.26	0.25	0.26
rs1433099	11242658	C (T)	0.48	0.40	0.71
rs3180023	11243209	A (C)	0.001	0.01	0.08

Abbreviations: AFR, African; EUR, European; MAF, minor allele frequency; NA, not available; SNP, single-nucleotide polymorphisms.

^aMAF in individuals from African (AFR) and European (EUR) descent taken from <http://browser.1000genomes.org>.

^bNovel single-nucleotide polymorphisms.

novel SNPs had a very low minor allele frequency of 0.01 and 0.02 for novel rs369850745 and rs369402076, respectively. A very rare variant in the promoter region of the *LDLR* gene, rs17249141 (c.-217 C→T), had a significant association with decreased LDL-C levels ($P=0.05$). For variant rs5930, detected in exon 10, the mean LDL-C level (3.29 mmol l^{-1}) for the AA genotype was higher than for the other two genotypes, although this was not statistically significant ($P=0.09$); the mean increase in LDL-C level for the AA genotype is 13.4%. In exon 11, variant rs5929 had a 3.5–5% reduction in LDL-C levels for each mutant allele even though there was no significant association with LDL-C levels. Intron 11 harbored the variant rs2738447, where the AA genotype was significantly associated with elevated LDL-C levels with a 23.7% increase in mean LDL-C levels when the individual had both the mutant alleles ($P=0.004$). SNP rs5925 was detected in exon 13, where the CC genotype had a mean LDL-C level greater than the TT and the TC genotypes, although the association did not reach statistical significance ($P=0.06$). In the 3'UTR, rs14158 was significantly associated ($P=0.05$) with mean LDL-C levels with an increase of 27.9% when both of the mutant alleles (AA) were present. Variants rs2738465 and rs3180023, also in the 3'UTR, were both significantly associated with elevated LDL-C levels ($P=0.05$ and $P=0.006$, respectively). All the SNPs were in Hardy–Weinberg equilibrium except for rs6413503.

Haplotype analysis

Four blocks were identified from the haplotype analysis performed with the software package Haploview version 4.2.25 (results in Table 4). From the association analysis, all four blocks contained haplotypes that were associated with either elevated or lower LDL-C levels. Block 1 revealed two haplotypes, namely GCGGC and GCAAC, where the first haplotype was associated with lower LDL-C ($P=0.03$) and the second haplotype was associated with increased LDL-C ($P<0.01$). Block 2 also contained two haplotypes where the CT haplotype was associated with lower LDL-C as more controls presented with the haplotype. The other haplotype in Block 2 was the AT haplotype, which was associated with elevated LDL-C. Three haplotypes were calculated in Block 3 and two of these were associated with elevated LDL-C levels, namely AGG ($P=0.03$) and GAG ($P=0.01$). The third haplotype (GGG) in Block 3 was more prevalent among the controls and is therefore associated with lower levels of LDL-C. Only one of the haplotypes (GC) identified in Block 4 was associated with lower levels of LDL-C.

DISCUSSION

For this study, we sequenced the exons and flanking intron regions of the *LDLR* gene to determine the genotypic variation of the population within the gene. Of the 70 SNPs identified by the bidirectional sequencing, 25 were genotyped within the entire study population, only two of which were novel (rs369850745 and rs369402076). We also performed haplotype analysis to determine whether there were any associations with mean LDL-C levels.

There were no variants found in exons three, four and five in the sub-sample of 30 individuals. Together with exons two and six, the above-mentioned exons code for the ligand-binding domain which is of great importance in binding apoE-containing lipoproteins such as LDL-C. Mutations within repeats four and five of the ligand-binding domain (present in exon 4) are common in FH patients.¹⁶

We found that four rare variants (rs17249141, rs148698650, rs116405216 and rs72658862) were present within the study population as well as in the African populations from the 1000 Genomes database but they were absent from the European population. SNP

Table 3 *LDLR* genotype distribution and association with LDL-C

SNP number	Nucleotide substitution		Position	<i>p</i>		LDL-C Frequency (mmol l ⁻¹)	ANCOVA P-value ^a
				(HWE)			
rs369850745	c.-314 C>T	Promoter	0.91				
CC	(novel)			1501	2.94		
CT				34	2.90	0.852	
TT				0	—		
rs17249141	c.-217 C>T	Promoter	0.98				
CC				1521	2.95 ^b		
CT				16	2.4 ^b	0.050	
TT				0	—		
rs17242759	c.67 + 18 C>A	Intron 1	0.42				
CC				1270	2.94		
CA				258	2.91	0.120	
AA				8	3.73		
rs2228671	c.81 C>T	Exon 2	0.19				
CC	(C27C)			1457	2.95		
CT				77	2.81	0.474	
TT				3	2.59		
rs10423288	c.314-50 T>C	Intron 3	0.99				
TT	(IVS3-50)			1132	2.93		
TC				373	2.97	0.444	
CC				30	2.71		
rs148698650	c.829 G>A	Exon 6	0.93				
GG	(E277K)			1507	2.94		
GA				29	3.05	0.590	
AA				0	—		
rs116405216	c.941-4 G>A	Intron 6	0.90				
GG				1364	2.94		
GA				164	2.92	0.945	
AA				6	3.04		
rs11669576	c.1171 G>A	Exon 8	0.77				
GG	(A391T)			1082	2.91		
GA				412	3.01	0.320	
AA				34	3.02		
rs72658862	c.1187-25 C>T	Intron 8	0.69				
CC				1346	2.94		
CT				163	2.94	0.653	
TT				7	3.33		
rs5930	c.1413 A>G	Exon 10	0.98				
AA	(R471R)			29	3.29		
AG				357	3.04 ^b	0.023	
GG				1146	2.90 ^b		
rs1569372	c.1586+53 A>G	Intron 10	0.99				
AA				518	2.99		
AG				747	2.90	0.376	
GG				266	2.95		
rs5929	c.1617 C>T	Exon 11	0.97				
CC	(P539)			1239	2.97		
CT				281	2.82	0.077	
TT				17	2.72		
rs4508523	c.1705 + 56 C>T	Intron 11	0.81				
CC				493	2.99		
CT				739	2.92	0.446	
TT				296	2.90		
rs2738447	c.1706-55A>C	Intron 11	0.64				
CC				1188	2.9 ^b		
CA				322	3.03 ^b	0.001	
AA				27	3.59 ^b		
rs688	c.1773 C>T	Exon 12	0.82				
CC	(N591N)			1381	2.95		
CT				132	2.83 ^b	0.051	
TT				2	4.62 ^b		

Table 3 (Continued)

SNP number	Nucleotide substitution		Position	<i>p</i>		LDL-C Frequency (mmol l ⁻¹)	ANCOVA P-value ^a
				(HWE)			
rs5925	c.1959 T>C	Exon 13	0.92				
TT	(V653V)				1001	2.97	
TC					475	2.84 ^b	0.020
CC					60	3.19 ^b	
rs5927	c.2232 A>G	Exon 15	0.18				
AA	(R744R)				96	2.90	
AG					629	3.00	0.212
GG					806	2.94	
rs369402076	c.2311 + 9 T>G	Intron 15	0.67				
TT	(novel)				1466	2.94	
TG					68	2.96	0.854
GG					0	—	
rs6413503	c.2548-42 G>A	Intron 17	<0.01				
GG					1267	2.91	
GA					166	3.10	0.089
AA					88	3.01	
rs14158	c.*52 G>A	3'UTR	0.83				
GG					1220	2.9 ^b	
GA					282	3.05 ^b	0.001
AA					19	3.71 ^b	
rs3826810	c.*141 G>A	3'UTR	0.13				
GG					1120	2.94	
GA					360	2.94	0.899
AA					17	2.82	
rs2738464	c.*315 G>C	3'UTR	0.59				
GG					165	2.87	
GC					650	2.93	0.620
CC					722	2.96	
rs2738465	c.*504 G>A	3'UTR	0.71				
GG					855	2.90 ^b	
GA					573	2.95	0.051
AA					107	3.18 ^b	
rs1433099	c.*666 T>C	3'UTR	0.20				
TT					403	2.92	
TC					800	2.95	0.917
CC					330	2.95	
rs3180023	c.*1217 C>A	3'UTR	1.00				
CC					1535	2.94 ^b	
CA					2	5.09 ^b	0.006
AA					0	—	

Abbreviations: ANCOVA, analysis of covariance; BMI, body mass index; HWE, Hardy-Weinberg equilibrium; LDL-C, low-density lipoprotein cholesterol; UTR, untranslated region.
^aP-value after adjustment for age, gender and BMI.
^bMeans with the same vowels differ significantly.

rs148698650 was not detected in the 1000 Genomes African samples but it was identified in a woman (heterozygote) in the 1000 Genomes population of Mexican ancestry from Los Angeles. The SNP was originally detected in individuals of Spanish, Cuban and Swedish descent.¹⁷⁻¹⁹ This could possibly be the first report of rs148698650 in an African population. The minor allele frequencies of the SNPs in the study population are more comparable to the African populations than the European populations. Variants such as rs688 and rs5925, which were used in a genetic risk score to define the predictive value of genetics in determining cholesterol levels,²⁰ would not be an obvious choice in the study population as their frequencies are much lower than in the European population. The frequencies and associations of variants in different populations should be investigated before they are used in further analyses to ensure their relevance.

Table 4 Haplotypes associated with LDL-C within the PURE population

Haplotype	Frequency	Case, control frequencies	P-value
<i>Block 1—rs11669576; rs72658862; rs5930; rs1569372; rs5929</i>			
GCGGC	0.36	0.34, 0.37	0.03
GCGAC	0.19	0.18, 0.20	0.16
ACGAC	0.16	0.17, 0.15	0.22
GCAAC	0.14	0.16, 0.12	<0.01
GCGAT	0.11	0.10, 0.11	0.21
GCGAT	0.06	0.06, 0.05	0.29
<i>Block 2—rs2738447; rs5925</i>			
CT	0.69	0.66, 0.70	0.01
CC	0.19	0.19, 0.20	0.42
AT	0.12	0.15, 0.10	<0.01
<i>Block 3—rs5927; rs14158; rs3826810</i>			
GGG	0.50	0.46, 0.52	0.03
AGG	0.27	0.29, 0.25	0.03
GGA	0.13	0.13, 0.14	0.34
GAG	0.10	0.12, 0.09	0.01
<i>Block 4—rs2738465; rs1433099</i>			
GT	0.53	0.54, 0.52	0.34
AC	0.25	0.27, 0.25	0.20
GC	0.22	0.20, 0.23	0.01

Abbreviations: LDL-C, low-density lipoprotein cholesterol; PURE, Prospective Urban and Rural Epidemiological.

The two novel SNPs (rs369850745 and rs369402076) identified had no meaningful effect on the function of the LDL receptor and there was no association with LDL-C levels. Variant rs369850745 is located more than 100 bp from the SREBP1 binding site, which could indicate that the SNP is too far away to have any effect on the transcription of the protein. In intron 15, the variant rs369402076 is most likely located just after the region that would influence the splicing of the messenger RNA (mRNA) or the base change does not have any effect on the splice function. Splice donor mutations 2311 + 2T>G,²¹ 2311 + 1G>C²² and 2311 + 1G>A²³ are located close to rs369402076 (2311 + 9T>G) and these variants are associated with hypercholesterolemia.

Sixteen heterozygote carriers of variant rs17249141 (c.-217C>T) were identified in the study population. This variant was first reported by Scholtz *et al.*⁷ in 1999 in a subject of mixed ancestry. The promoter activity of this variant increased to 160% of normal in cells depleted of sterol and 50% in cells supplemented with a sterol-containing medium in the transfection study. In this study, the effect of one allele was an 18.6% reduction in LDL-C levels, which is in accordance with an increased transcriptional rate of the gene.

In exon 11, the silent variant rs5929 (c.1617C>T; P539P) was detected and the base change from a cytosine to a thymine causes a synonymous mutation. As previously reported, the SNP is not associated with LDL-C levels²⁴ and the small lowering effect of the mutant allele in the study population might be caused by a variant with which the SNP is in linkage or it might be a specific phenomenon related to the study population.

The intronic variant rs2738447 (c.1706-55A>C) had a statistically significant association with elevated LDL-C where the homozygote mutants (genotype AA) had a mean LDL-C level of 3.59 mmol l⁻¹. No genotype-phenotype information could be found for this variant.

Analyses of this variant in the Human Splicing Finder indicate that the SNP could form part of various enhancer motifs and therefore influence the splicing efficiency of the pre-mRNA, which leads to alternative splicing.²⁵ We can hypothesize that the function of the different proteins (receptors) formed by the alternative splicing could be less effective in internalizing LDL-C than receptors synthesized by normal splicing, which would then lead to increased plasma levels of LDL-C.²⁵

Variants rs688 (c.1773 C>T) and rs5925 (c.1959 T>C) are both synonymous mutations and have been identified as exon-splicing enhancers associated with increased LDL-C levels by Rafiq *et al.*²⁶ These variants alter the efficiency of LDLR splicing and therefore have an elevated effect on LDL-C.²⁶ As previously mentioned, these two variants have been used in a genetic risk score to determine the predictive value of genetics on plasma total cholesterol levels where the results demonstrated that it could be beneficial in identifying individuals with elevated plasma cholesterol levels.²⁰ In the study done by Lu *et al.*,²⁰ the minor allele frequencies of both these SNPs were 0.43 in their European population. The minor allele frequencies of the SNPs in the PURE population were much lower (rs688 = 0.04; rs5925 = 0.19). Rs5929 and rs688 was in strong linkage disequilibrium (LD) with $r^2 = 0.99$;²⁰ however, the LD of these two SNPs in the PURE population were much weaker, with $r^2 = 0.20$. The lower r^2 in the study population is also an indication that the two SNPs have been separated through recombination.²⁷ LD between most of the analyzed SNPs (data not shown) is weak when r^2 is used as a measure of LD, which is a characteristic of the African genome where more recombination events have taken place.²⁸ A multiallelic D' of 0.35 computed between haplotype Block 2 and Block 3 also indicates that historical recombination has taken place between the different blocks.²⁷ Rs688 and rs5925 are situated within haploblock 2 (Table 4) as identified by Haploview analyses using Gabriel *et al.*²⁸ confidence intervals. Differences in LD among populations are possibly attributable to differences in demographic history,²⁹ as the biological determinants of LD (rates of recombination, mutation, gene conversion) are estimated to be constant across groups.²⁸

The 27.9% increase in LDL-C levels in individuals homozygous for rs14158 is indicative of the role the SNP has in the expression of the gene, as variants found in the 3'UTR of a gene have been proven to affect the regulation of mRNA.³⁰ LDLR gene expression is controlled mainly by regulatory sequences in the 3'UTR as it regulates changes in mRNA stability at the transcriptional and the post-transcriptional level.^{31,32} Three mRNA destabilizing elements, known as AU-rich elements, responsible for the rapid turnover of LDLR mRNA, have been identified.^{32,33} Variant, rs14158 (c.*52G>A) is present in the 2.5-kb-long 3'UTR of the LDLR gene and, together with rs3826810 (c.*141G>A), rs2738464 (c.*315G>C), rs2738465 (c.*504G>A), rs1433099 (c.*666T>C) and rs2738466 (c.*773A>G), it forms a common haplotype reported by Muallem *et al.*³⁴ Rs2738466 was not genotyped in the PURE study population (SNP is too close to rs1433099 to be analyzed on the BeadXpress platform) and therefore no comparison could be made between the haplotypes formed within the current study population and those used by Muallem *et al.*³⁴ However, analysis of the five SNPs that were genotyped revealed seven haplotypes, of which only two had significant associations with LDL-C (Table 4). Haplotype GCGGC was associated with lower LDL-C levels with a mean level of 2.69 mmol l⁻¹ and haplotype AGCAC was associated with elevated LDL-C levels with a mean value of 3.71 mmol l⁻¹ ($P < 0.01$).

The AA genotype (homozygote mutant) of variant rs2738465 was associated with increased levels of LDL-C ($P = 0.05$). The SNP is also

found in the 3'UTR that forms part of the above-mentioned haplotype and therefore has a role in the regulation of mRNA. When the mutant allele is included in a haplotype (AGCAC), it is associated with increased LDL-C levels. Two heterozygotes were detected for the very rare SNP, rs3180023, and, despite the small number, the association with LDL-C did reach statistical significance ($P=0.006$). The variant is also more prevalent in the European population than in the study population ($MAF=0.001$ vs $MAF=0.08$). Another rare variant, rs2228671, found in exon 2, is a synonymous mutation where the base change of a cytosine to a thymine does not cause the cysteine at position 27 to change to a different amino acid. Even though the association of the SNP in this population is not statistically significant ($P=0.47$), it has been reported on in various other populations such as a German population where there was a significant association with lower LDL-C and the T allele of the variant. LDL-C was 0.19 mmol l^{-1} lower for every T allele carried by the German population.³⁵ In the PURE population, such an association was not present but for the TT genotype, LDL-C was 22% lower than for the CC genotype. Rs11669576 (A391T original amino acid nr A370T) was also detected in this study population but with no significant association with LDL-C as determined in various other studies.^{36–38}

The haplotype analysis according to the confidence intervals by Gabriel *et al.*²⁸ revealed four haploblocks. Interestingly, the haplotypes from each block with the highest frequency (GCGGC=0.36; CT=0.69; GGG=0.50; GC=0.22) were associated with lower levels of LDL-C except for Block 4 (Table 4). These haplotypes include the alleles associated with lower LDL-C, as can be seen from the first haplotype in Block 1, where the G alleles from rs5930 and rs1569372 have lower mean LDL-C levels. Of the haplotypes in Block 4, the GC haplotype is the only haplotype for which the association with lower LDL-C reached statistical significance, where the G allele in rs2738465 has a lower mean LDL-C level. The T and C alleles in rs1433099 have similar mean LDL-C levels.

The only variant that was not in HWE was rs6413503, where there was an excess of homozygote mutants as a result of the genotyping assay and a large number of heterozygotes were not included in the analysis. This is also the only variant that is not in HWE and therefore we do not suspect that it is due to genetic drift, non-random mating, selection or population structure.

There are few studies on the association of rare and common SNPs with LDL-C in an African population and these results give us a greater insight into how genetics have a role in LDL-C at population level in this particular ethnic group. This was, however, a cross-sectional study and functional information for the novel and lesser-known SNPs is not available. Further research into the specific function of each of the SNPs would be of great value in determining to what extent they regulate LDL-C levels. All the SNPs identified by the automated sequencing could not be genotyped within the entire study population because of restriction within the genotyping assay and it would be of great value to genotype these SNPs as well in the study population as 22 of these variants were detected with in the 3'UTR. The predictive value of the significantly associated SNPs can be determined at a later stage, however, as the study is prospective with a follow-up of 12 years, it would also be of great value to determine the prevalence and frequency of SNPs in the intron regions of the LDLR that were not analyzed as current research indicates the importance of variants found within these regions.

In summary, we have shown that there are SNPs present in the study population that associate with LDL-C levels. For many of these SNPs the prevalence was low and only the individuals harboring these SNPs had significantly higher or lower LDL-C levels. The study

population appears to be healthy with normal mean LDL-C levels and the effect of some variants does not raise LDL-C to pathological levels. Unfortunately, the clinical significance of the results is absent because of the observational nature of the study. The SNPs that significantly associated with LDL-C levels were situated mostly within regulatory regions such as the promoter area and the 3'UTR, which affects either transcriptional rate or gene expression (regulatory SNPs). Several of the SNPs, which were situated in the coding areas (exons) of the gene, were also prevalent in a few individuals and either had an association with elevated LDL-C or had no effect on the levels. In this population it appears, therefore, that the regulatory SNPs make a greater contribution toward LDL-C levels than do SNPs in the coding regions. Interestingly, there were no variants present in exons 3, 4, 5, 9 and 17 of the gene. The absence of variants in these exons could be a reason for the low mean LDL-C as well as the lower prevalence of FH in this population. Exons 4 and 9 have been identified as hot spots for mutations causing FH whereas exons 13 and 15 contain a smaller number of mutations and are classified as FH cold spots.³⁹ In agreement with previous evidence,³⁹ we found only one SNP in each of exon 13 and 15.

Haplotypes were also identified with significant associations mostly with lower LDL-C, and the low mean LDL-C levels of the study population could possibly be attributed to the presence of these haplotypes in a large percentage of the population (Table 4). This need to be further investigated, however, to determine whether other confounding factors exist that could result in the low mean LDL-C levels of the study population. Minimal linkage disequilibrium between the various SNPs exists, which is an indication of historical recombination, a characteristic of the African genome.

We also showed that the minor allele frequencies of the identified SNPs in the black South African study population are different from those of the European population, with only a handful of SNPs sharing frequencies across the populations; conversely, the frequencies are very similar to those of other African populations. This indicates that genetic information should be population-specific when used in downstream analysis.

Further research on the other 45 variants identified in the sub-sample of 30 individuals is warranted in order to give us even greater insight into the role that the LDL receptor has in determining LDL-C levels in this study population.

ACKNOWLEDGEMENTS

We thank the organizers of the South African PURE study, especially Prof A Kruger and her team, for providing the opportunity to work with the data and samples from the study population. This study would not have been possible without their contributions. Wayne Towers is thanked for his support with regard to the genetic aspects of the study as well as for obtaining funding for the project. Funding was received from the NorthWest University, and we would like to thank the institution for all of its contributions.

- 1 Brown, M. S., Kovanen, P. T. & Goldstein, J. L. Regulation of plasma cholesterol by lipoprotein receptors. *Science* **212**, 628–635 (1981).
- 2 Burnett, J. R. & Hooper, A. J. Common and rare gene variants affecting plasma LDL cholesterol. *Clin. Biochem. Rev.* **29**, 11–26 (2008).
- 3 Steyn, K., Jooste, P. L., Bourne, L., Fourie, J., Badenhorst, C. J., Bourne, D. E. *et al.* Risk factors for coronary heart disease in the black population of the Cape Peninsula. The BRISK study. *S. Afr. Med. J.* **79**, 480–485 (1991).
- 4 Walker, A. R. P. & Sareli, P. Coronary heart disease: outlook for Africa. *J. R. Sc. Med.* **90**, 23–27 (1997).
- 5 Heller, D. A., de Faire, U., Pedersen, N. L., Dahlen, G. & McClearn, G. E. Genetic and environmental influences on serum lipid levels in twins. *N. Engl. J. Med.* **328**, 1150–1156 (1993).

- 6 Soutar, A. K. & Naoumova, R. P. Mechanisms of disease: genetic causes of familial hypercholesterolemia. *Nat. Clin. Pract. Card.* **4**, 214–225 (2007).
- 7 Scholtz, C., Peeters, A., Hoogendijk, C., Thiar, R., Villiers, J., Hillermann, R. *et al*. Mutation -59c->t in repeat 2 of the LDL receptor promoter: reduction in transcriptional activity and possible allelic interaction in a South African family with familial hypercholesterolaemia. *Hum. Mol. Genet.* **8**, 2025–2030 (1999).
- 8 Teo, K., Chow, C. K., Vaz, M., Rangarajan, S. & Yusuf, S. PURE Investigators-Writing Group. The Prospective Urban Rural Epidemiology (PURE) study: examining the impact of societal influences on chronic non-communicable diseases in low-, middle-, and high-income countries. *Am. Heart J.* **158**, 1–7, e1 (2009).
- 9 Friedewald, W. T., Levy, R. I. & Fredrickson, D. S. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* **18**, 499–502 (1972).
- 10 Collins, J. S. & Schwartz, C. E. Detecting polymorphisms and mutations in candidate genes. *Am. J. Hum. Genet.* **71**, 1251 (2002).
- 11 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
- 12 Lohman, T. G., Roche, A. F. & Martorell, R. *Anthropometric Standardization Reference Manual*. 1st edn (Human kinetic books, Champaign, Illinois, 1988).
- 13 MacIntyre, U., Venter, C. & Vorster, H. A culture-sensitive quantitative food frequency questionnaire used in an African population: 1. Development and reproducibility. *Public Health Nutr.* **4**, 53 (2001).
- 14 MacIntyre, U., Venter, C. & Vorster, H. A culture-sensitive quantitative food frequency questionnaire used in an African population: 2. Relative validation by 7-day weighed records and biomarkers. *Public Health Nutr.* **4**, 63 (2001).
- 15 Langenhoven, M., Kruger, M., Gouws, E. & Faber, M. *Medical Research Council Food Composition Tables* (1998).
- 16 Soutar, A. K. in *Low-density Lipoprotein Receptors* (eds Betteridge, D. J., Illingworth, D. R. & Shepherd, J.) 303–322 (Arnold Publishers, London, 1999).
- 17 Cenarro, A., Jensen, H. K., Casao, E., Civeira, F., González-Bonillo, J., Rodríguez-Rey, J. C. *et al*. Identification of recurrent and novel mutations in the LDL receptor gene in Spanish patients with familial hypercholesterolemia. *Hum. Mutat.* **11**, 413–413 (1998).
- 18 Pereira, E., Ferreira, R., Hermelin, B., Thomas, G., Bernard, C., Bertrand, V. *et al*. Recurrent and novel LDL receptor gene mutations causing heterozygous familial hypercholesterolemia in La Habana. *Hum. Genet.* **96**, 319–322 (1995).
- 19 Ekström, U., Abrahamson, M., Sveger, T., Lombardi, P. & Nilsson-Ehle, P. An efficient screening procedure detecting six novel mutations in the LDL receptor gene in Swedish children with hypercholesterolemia. *Hum. Genet.* **96**, 147–150 (1995).
- 20 Lu, Y., Feskens, E. J., Boer, J., Imholz, S., Verschuren, W., Wijmenga, C. *et al*. Exploring genetic determinants of plasma total cholesterol levels and their predictive value in a longitudinal study. *Atherosclerosis* **213**, 200–205 (2010).
- 21 Amsellem, S., Briffaut, D., Carrié, A., Rabès, J., Girardet, J., Fredenrich, A. *et al*. Intronic mutations outside of Alu-repeat-rich domains of the LDL receptor gene are a cause of familial hypercholesterolemia. *Hum. Genet.* **111**, 501–510 (2002).
- 22 Humphries, S. E., Whittall, R. A., Hubbart, C. S., Maplebeck, S., Cooper, J. A., Soutar, A. *et al*. Genetic causes of familial hypercholesterolaemia in patients in the UK: relation to plasma lipid levels and coronary heart disease risk. *J. Med. Genet.* **43**, 943–949 (2006).
- 23 Lelli, N., Garuti, R., Ghisellini, M., Tiozzo, R., Rolleri, M., Aimale, V. *et al*. Occurrence of multiple aberrantly spliced mRNAs of the LDL-receptor gene upon a donor splice site mutation that causes familial hypercholesterolemia (FHBenevento). *J. Lipid Res.* **36**, 1315–1324 (1995).
- 24 Zakharova, F., Damgaard, D., Mandelshtam, M., Golubkov, V., Nissen, P., Nilsen, G. *et al*. Familial hypercholesterolemia in St-Petersburg: the known and novel mutations found in the low density lipoprotein receptor gene in Russia. *BMC Med. Genet.* **6**, 6 (2005).
- 25 Desmet, F., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M. & Béroud, C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67–e67 (2009).
- 26 Rafiq, S., Ahmed, N., Soutar, A. & Qamar, R. The genetic characterization of familial hypercholesterolemia in Pakistan. *J. Basic Appl. Sci.* **7**, 21–25 (2011).
- 27 Bansal, A., Vignal, C. & McGinnis, R. in *Tools for Statistical Genetics*. 2nd edn (ed Barnes, M. R.) 217–246 (Wiley, West Sussex, England, 2008).
- 28 Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B. *et al*. The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
- 29 Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J. *et al*. Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
- 30 Chen, J., Férec, C. & Cooper, D. N. A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. *Hum. Genet.* **120**, 1–21 (2006).
- 31 Nakahara, M., Fujii, H., Maloney, P. R., Shimizu, M. & Sato, R. Bile acids enhance low density lipoprotein receptor gene expression via a MAPK cascade-mediated stabilization of mRNA. *J. Biol. Chem.* **277**, 37229–37234 (2002).
- 32 Wilson, G., Vasa, M. & Deeley, R. Stabilization and cytoskeletal-association of LDL receptor mRNA are mediated by distinct domains in its 3' untranslated region. *J. Lipid Res.* **39**, 1025–1032 (1998).
- 33 Knouff, C., Malloy, S., Wilder, J., Altenburg, M. K. & Maeda, N. Doubling expression of the low density lipoprotein receptor by truncation of the 3'-untranslated region sequence ameliorates type iii hyperlipoproteinemia in mice expressing the human apoE2 isoform. *J. Biol. Chem.* **276**, 3856–3862 (2001).
- 34 Muallem, H., North, K. E., Kakoki, M., Wojczynski, M. K., Li, X., Grove, M. *et al*. Quantitative effects of common genetic variations in the 3' UTR of the human LDL-receptor gene and their associations with plasma lipid levels in the Atherosclerosis Risk in Communities study. *Hum. Genet.* **121**, 421–431 (2007).
- 35 Linsel-Nitschke, P., Götz, A., Erdmann, J., Braenne, I., Braund, P., Hengstenberg, C. *et al*. Lifelong reduction of LDL-cholesterol related to a common variant in the LDL-receptor gene decreases the risk of coronary artery disease—A Mendelian Randomisation Study. *PLoS One* **3**, 1–9 (2008).
- 36 Gudnason, V., Patel, D., Sun, X., Humphries, S., Soutar, A. & Knight, B. Effect of the Stul polymorphism in the LDL receptor gene (Ala 370 to Thr) on lipid levels in healthy individuals. *Clin. Genet.* **47**, 68–74 (1995).
- 37 Vieira, J. R. S., Whittall, R. A., Cooper, J. A., Miller, G. J. & Humphries, S. E. The A370T variant (Stul polymorphism) in the LDL receptor gene is not associated with plasma lipid levels or cardiovascular risk in UK men. *Ann. Hum. Genet.* **70**, 697–704 (2006).
- 38 El Messal, M., Chihab, K. A., Chater, R., Vallvé, J. C., Bennis, F., Hafidi, A. *et al*. Familial hypercholesterolemia in Morocco: first report of mutations in the LDL receptor gene. *J. Hum. Genet.* **48**, 199–203 (2003).
- 39 Varret, M., Rabès, J., Thiar, R., Kotze, M. J., Baron, H., Cenarro, A. *et al*. LDLR Database: new additions to the database and the software, and results of the first molecular analysis. *Nucleic Acids Res.* **26**, 248–252 (1998).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)