Zinc Finger Protein 202, genetic variation, and HDL cholesterol in the general population

Maria C. Stene,* Ruth Frikke-Schmidt,* Børge G. Nordestgaard,†,§ and Anne Tybjærg-Hansen[1,*,§]

Department of Clinical Biochemistry,* Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; Department of Clinical Biochemistry,† Herlev University Hospital, Copenhagen, Denmark; and Copenhagen City Heart Study,§ Bispebjerg University Hospital, Copenhagen, Denmark

Abstract Zinc Finger Protein 202 (ZNF202) is a transcriptional repressor that binds elements found predominantly in genes involved in HDL metabolism. We tested the following hypotheses: *1*) frequencies of single-nucleotide polymorphisms (SNPs) and haplotypes in *ZNF202* differ between individuals with low and high HDL cholesterol; and *2*) SNPs in *ZNF202* affect HDL cholesterol levels in the general population. We screened the promoter and protein-coding exons of *ZNF202* in individuals with the highest 1% (n = 95) and lowest 1% (n = 95) HDL cholesterol among 9,259 Danish adults. None of the 10 SNPs identified differed in frequency as single sites or as haplotypes between low and high HDL cholesterol groups. In accordance with this, seven mutations were equally frequent (4–5%) in individuals with low or high HDL cholesterol. Finally, for all five SNPs identified in the coding region, we determined the association of genotype with HDL cholesterol in 9,259 individuals from the general population. Four SNPs were not associated with variation in HDL cholesterol, although c.*2T>G homozygosity was associated with a discrete effect on HDL cholesterol in men. We show that genetic variation in *ZNF202* is common in the general population. However, SNPs in the protein-coding region of *ZNF202* do not make a major contribution to HDL cholesterol levels.—Stene, M. C., R. Frikke-Schmidt, B. G. Nordestgaard, and A. Tybjærg-Hansen. ***Zinc Finger Protein 202*, genetic variation, and HDL cholesterol in the general population.** ***J. Lipid Res.*** **2006.** 47: **944–952.**

Supplementary key words apolipoproteins • genetic epidemiology • large-scale genotyping • lipids • lipoproteins • molecular biology • molecular medicine • reverse cholesterol transport • transcription factors • high density lipoprotein cholesterol

Levels of HDL cholesterol are inversely related to risk of ischemic heart disease in the general population (1, 2). The HDL particle is responsible for the delivery of cellular cholesterol from peripheral tissues to the liver and is thus a key component in reverse cholesterol transport (3).

Twin and family studies suggest that approximately half of the variation in HDL cholesterol is genetically determined (4–7). A new susceptibility locus for familial hypoalphalipoproteinemia (Online Mendelian Inheritance in Man 604091) on chromosome 11q23 was identified in Utah pedigrees (8); this region contains the *Zinc Finger Protein 202* (*ZNF202*) gene. ZNF202 is functionally characterized by a SRE-ZBP, CT-finS1, AW-1, Number 18 (SCAN) oligomerization domain, a Krüppel-associated box (KRAB) repression domain, and eight zinc finger (Cys_2His_2) DNA binding motifs, a typical domain architecture for transcription factors (9). ZNF202 target genes are mainly involved in lipid and, particularly, HDL cholesterol metabolism (10–12), suggesting that this transcriptional repressor might be important in the determination of HDL cholesterol levels in the general population. However, it is largely unknown to what extent *ZNF202* varies genetically in the general population and whether such genetic variation influences HDL cholesterol levels.

We tested the following hypotheses: *1*) frequencies of single-nucleotide polymorphisms (SNPs) and haplotypes in *ZNF202* differ between individuals with low and high HDL cholesterol levels; and *2*) SNPs in the protein-coding region of *ZNF202* affect HDL cholesterol levels in the general population. To increase the likelihood of identifying genetic variation with significant effects on HDL cholesterol levels, we screened the promoter and protein coding regions of *ZNF202* (~2 kb) in 95 individuals with the 1% lowest and in 95 individuals with the 1% highest HDL cholesterol levels for age and gender from a general population sample, the Copenhagen City Heart Study (n = 9,259). All SNPs identified in and around the protein-coding region were genotyped in the entire general population sample, and the effect on HDL cholesterol and apolipoprotein A-I (apoA-I) levels was determined.

Manuscript received 30 November 2005 and in revised form 27 January 2006.

Published, JLR Papers in Press, February 7, 2006.
DOI 10.1194/jlr.M500521-JLR200

Abbreviations: apoA-I, apolipoprotein A-I; BMI, body mass index; KRAB, Krüppel-associated box; LD, linkage disequilibrium; SCAN, SRE-ZBP, CT-finS1, AW-1, Number 18; SNP, single-nucleotide polymorphism; UTR, untranslated region; ZNF202, Zinc Finger Protein 202.

[1] To whom correspondence should be addressed.
e-mail: at-h@rh.dk

METHODS

Subjects

The Copenhagen City Heart Study is a prospective cardiovascular population study of individuals selected based on the Central Population Register Code to reflect the adult Danish general population aged 20 to 80+ years. In 1991–1994, 9,259 participants (55% women) gave blood for DNA analyses (13, 14). More than 99% were white and of Danish descent. This study was approved by the local ethical committee: Nos. 100.2039/91 and 01-421/94, Copenhagen and Frederiksberg committee. All participants gave written informed consent.

For the genetic screening of *ZNF202* (GenBank accession number NM_003455), we selected individuals from the Copenhagen City Heart Study with the 1% lowest (n = 95) and 1% highest (n = 95) HDL cholesterol levels for age and gender (in 10 year age groups). Consequently, the cutoff levels for HDL cholesterol depend upon the seven age groups for each gender (15). We previously showed that by screening these groups with extreme phenotypes, we increased the likelihood of identifying mutations and SNPs (rare allele frequency of <1% and ⩾1%, respectively) with impact on HDL cholesterol levels in the general population (15). With the exception of a synonymous SNP [c.731A>G (p.V244V)], all five SNPs identified in or around the protein-coding region by screening *ZNF202* were genotyped in the entire general population sample (n = 9,259), and the effect of each SNP on variation in HDL cholesterol and apoA-I levels was determined as overall effects (regardless of variation at the other four sites) and as isolated single-site effects (five SNP genotypes differing only at the relevant SNP) using data from the third examination of the Copenhagen City Heart Study (1991–1994). In tables and figures, data for the general population sample are presented only for individuals in whom the five SNP genotypes for all SNPs were available (n = 9,103 of 9,259).

Gene screening

Genomic DNA was isolated from frozen whole blood (QiaAmp® DNA Blood Mini Kit; Qiagen, Hilden, Germany). The *ZNF202* gene contains ten exons; however, only six exons (exons 5–10) encode the ZNF202 protein (648 amino acids) (10). Sixteen PCR fragments were amplified covering 700 bp of the *ZNF202* promoter upstream of exon 1 (12), all six protein-encoding exons (exons 5–10) (accession number NM_003455) (10), and exon-intron boundaries. Mutational analysis of the amplicons was performed by denaturing HPLC. PCR fragments showing heteroduplex formation were subsequently sequenced on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Inc., Foster City, CA).

SNP genotyping in the general population

The ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Inc.) was used to genotype the general population sample for all five SNPs identified in and around the coding region. TaqMan-based assays were used. Combined genotypes for all five SNPs were available for 9,103 individuals of the total cohort of 9,259 individuals.

Biochemical assays

Colorimetric and turbidimetric assays (Hitachi autoanalyzer) were used to measure plasma levels of total cholesterol, HDL cholesterol, triglycerides, apoB, and apoA-I (all from Boehringer Mannheim, Mannheim, Germany).

Statistical analyses

Differences in allele frequencies between individuals with low and high HDL cholesterol were evaluated using Fisher's exact test. Pairwise disequilibrium statistics were calculated from D = h − pq, where h is the frequency of the rare estimated haplotype for a pair of sites and p and q are the frequencies, assuming no linkage, of the alleles in that haplotype (16). Linkage disequilibrium (LD) was expressed as D′ (17). Significance levels for pairwise LD were estimated by the likelihood-ratio test. Estimated haplotype frequencies were calculated using the expectation maximization algorithm (18). The association of *ZNF202* genotype with variation in HDL cholesterol and apoA-I levels was determined by ANOVA with Student's *t*-test as a posthoc test. $P < 0.05$ on a two-sided test was considered significant. *P* values were not corrected for multiple comparisons.

RESULTS

Characteristics of individuals with the lowest 1% and highest 1% HDL cholesterol levels and of the total general population sample are shown in **Table 1**. Individuals in the low HDL cholesterol group had lower apoA-I levels, higher triglyceride levels, and were more obese than individuals in the high HDL cholesterol group or in the general population.

Genetic variation in ZNF202

Six SNPs [c.IVS4-240A>T, c.IVS4-223T>C, c.461C>T (p.A154V), c.731A>G (p.V244V), c.775A>G (p.K259E), and c.*2T>G] and two mutations [c.820G>C (p.V274L)

TABLE 1. Characteristics of individuals from the general population with the lowest 1% and highest 1% HDL cholesterol levels and of the total general population sample

	Low HDL Cholesterol (n = 95)		High HDL Cholesterol (n = 95)		General Population (n = 9,103)	
Characteristic	Women	Men	Women	Men	Women	Men
Age (years)	55 ± 2.9	55 ± 2.9	55 ± 2.7	57 ± 2.8	59 ± 0.2	57 ± 0.2
Sex (%)	50	50	51	49	55	45
Total cholesterol (mmol/l)	5.9 ± 0.2	5.3 ± 0.2	6.5 ± 0.2	6.0 ± 0.1	6.3 ± 0.02	6.0 ± 0.02
ApoB (mg/dl)	94 ± 3.7	82 ± 3.4	66 ± 3.2	63 ± 2.3	86 ± 0.3	86 ± 0.3
HDL cholesterol (mmol/l)	0.8 ± 0.02	0.6 ± 0.02	3.3 ± 0.07	2.9 ± 0.08	1.7 ± 0.01	1.4 ± 0.01
ApoA-I (mg/dl)	98 ± 2.5	84 ± 2.7	214 ± 4.8	196 ± 3.7	151 ± 0.4	130 ± 0.4
Triglycerides (mmol/l)	3.1 ± 0.3	4.6 ± 0.5[a]	1.3 ± 0.1	1.2 ± 0.08	1.7 ± 0.01	2.1 ± 0.03
Body mass index (kg/m^2)	27 ± 0.6	28 ± 0.8	23 ± 0.5	24 ± 0.5	25 ± 0.07	26 ± 0.06

apoA-I, apolipoprotein A-I. Values shown are means ± SEM.
[a] Six extreme outliers (triglycerides > 20 mmol/l) were excluded.

and c.1813A>T (p.R605W)] were identified in or flanking the protein-coding regions of the gene and in the 3′ untranslated region (3′UTR), four of which introduced amino acid substitutions (**Table 2**, **Fig. 1**): c.461C>T (p.A154V) and c.820G>C (p.V274L) substitute similar amino acids, c.775A>G (p.K259E) introduces a shift between positively charged and negatively charged side chains, and c.1813A>T (p.R605W) introduces a shift between positively charged nonpolar and uncharged polar side chains. c.461C>T (p.A154V) is located in the SCAN domain, known to be important for protein-protein interaction, whereas both c.775A>G (p.K259E) and c.820G>C (p.V274L) are in the KRAB A domain, important for transcriptional repression. The amino acid residues affected by these two variants are situated in highly conserved areas of ZNF202 and are themselves completely conserved between species (human, mouse, rat) (**Fig. 2**). c.1813A>T (p.R605W) is located in the seventh zinc finger motif, in a highly conserved area, and p.R605 is also conserved between species (human, mouse, rat). c.*2T>G is located 2 bp downstream of the translation stop site (Table 2, Fig. 1). Of the four nonsynonymous variants identified, c.820G>C (p.V274L), c.1813A>T (p.R605W), and c.775A>G (p.K259E) have not been reported previously.

Four SNPs (g.−685G>A, g.−660A>G, g.−118G>T, and g.+34G>A) and three mutations (g.−447T>C, g.−232C>T, and g.−122C>T) were identified in the promoter region and 5′UTR of exon 1 (Table 2, Fig. 1): g.−685G>A and g.−660A>G are situated in a silencer sequence, g.−447T>C in a putative Yin Yang 1 binding site, g.−122C>T in a putative myeloid zinc finger 1 binding site (12), and g.+34G>A in a putative downstream promoter element. Four of these seven variants have not been reported previously.

Four variants were identified in introns: two SNPs and two new mutations. Judging from their positions in relation to the exon-intron junctions and the nucleotide substitutions introduced, none of these would be predicted to affect splicing.

Individuals in the low and high HDL cholesterol groups

A total of 17 genetic variants were identified in the promoter and in and around the protein-coding region of *ZNF202* (Table 2). Allele frequencies ranged from 1:380 (0.3%) to 147:380 (39%) in the low and high HDL cholesterol groups. Ten SNPs (allele frequencies > 1%) were identified in both HDL cholesterol groups, two of which were new [g.−118 G>T and c.775A>G (p.K259E)].

TABLE 2. Genetic variation in protein-coding and regulatory sequences of *ZNF202* in individuals from the general population with extreme HDL cholesterol levels

Gene Region	Nucleotide Substitution	No. of Alleles: Low HDL Cholesterol (n = 190)	No. of Alleles: High HDL Cholesterol (n = 190)	Allele Frequency in the General Population (n = 9,103)	Amino Acid Residue	Side Chain Substitution	Domain	References
Promoter	g.−685G>A	9	3	—	—	—	Silencer sequence	dbSNP:10736530
Promoter	g.−660A>G	52	51	—	—	—	Silencer sequence	dbSNP:10893081
Promoter	g.−447T>C	1	1	—	—	—	Proximal promoter Yin Yang 1 binding site	New
Promoter	g.−232C>T	1	0	—	—	—	Proximal promoter	New
Promoter	g.−122C>T	1	0	—	—	—	Proximal promoter myeloid zinc finger 1 binding site	New
Promoter	g.−118G>T	13	8	—	—	—	Proximal promoter	New
Exon 1	g.+34G>A	1	4	—	—	—	5′UTR downstream promoter element	dbSNP:2272142
Intron 4	c.IVS4-240A>T	22	32	0.15	—	—	—	dbSNP:2282641
Intron 4	c.IVS4-223T>C	50	55	0.30	—	—	—	dbSNP:2282642
Intron 5	c.IVS5+37G>A	1	0	—	—	—	—	New
Intron 5	c.IVS5+52G>A	0	2	—	—	—	—	New
Exon 6	c.461C>T	52	54	0.30	p.A154V	Nonpolar to nonpolar	SCAN	dbSNP:1144507 and (25)
Exon 8	c.731A>G	4	3	—	p.V244V	—	KRAB A	dbSNP:2282644
Exon 8	c.775A>G	3	3	0.01	p.K259E	Positively charged to negatively charged	KRAB A	New
Exon 8	c.820G>C	1	0	—	p.V274 L	Nonpolar to nonpolar	KRAB A	New
Exon 10	c.1813A>T	0	1	—	p.R605W	Positively charged nonpolar to uncharged polar	Zinc finger motif 7	New
Exon 10	c.*2T>G	78	69	0.39	—	—	3′UTR	dbSNP:3183878

KRAB, Krüppel-associated box; SCAN, SCR-ZBP, CT-fins1, AW-1, Number 18; UTR, untranslated region; *ZNF202*, Zinc Finger Protein 202. Nomenclature for the description of sequence variation in the promoter region and exon 1 is according to Langmann et al. (12), where nucleotide 1 denotes the transcription start site. Nomenclature for the description of sequence variation in and around the protein-coding region and for changes affecting protein level is according to den Dunnen and Antonarakis (25), where nucleotide 1 denotes A in the start codon ATG (methionine = amino acid 1) in exon 5, deduced from cDNA sequence NM_003455 and protein sequence NP_003446. c.*2T>G = T to G substitution in the second nucleotide after the translation stop site. SNPs are characterized by more than four allele counts of a total of 380 (rare allele frequency > 1%). Mutations are characterized by four or fewer allele counts of a total of 380 (rare allele frequency ≤ 1%).

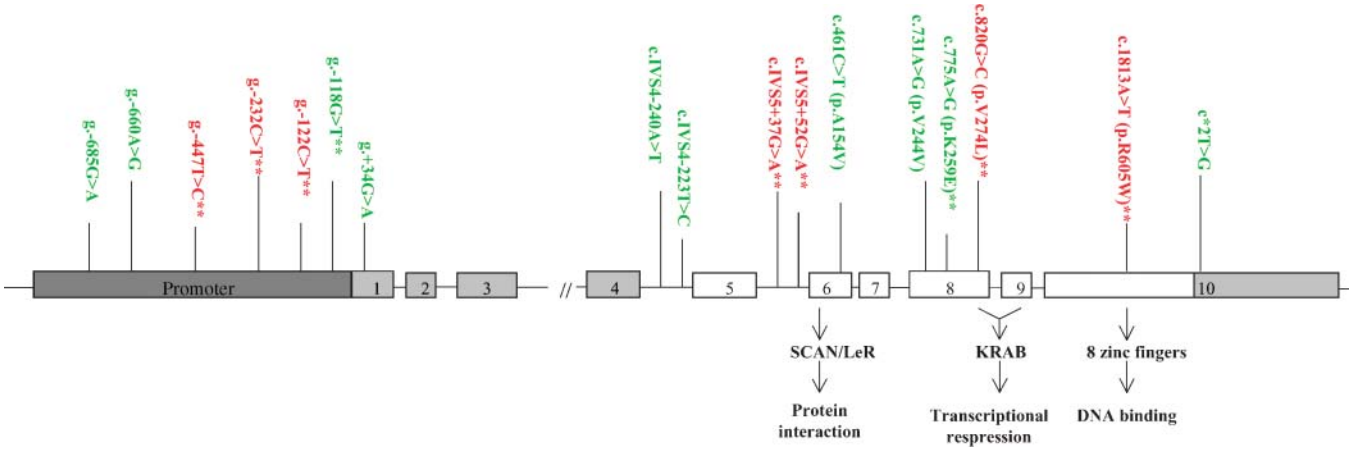

Fig. 1. Schematic model of Zinc Finger Protein 202 (*ZNF202*) according to Wagner et al. (10) and corresponding protein domains. Single-nucleotide polymorphisms (SNPs; green) and mutations (red) identified here are superimposed. ZNF202 protein is characterized by a SRE-ZBP, CT-finS1, AW-1, Number 18 (SCAN) oligomerization domain, a Krüppel-associated box (KRAB) repression domain, and eight zinc finger (Cys_2His_2) DNA binding motifs. Two asterisks indicate new variants identified in this study.

None of these 10 SNPs differed in frequency between individuals with the lowest 1% and the highest 1% HDL cholesterol levels. Of the seven mutations (allele frequencies $\leqslant$ 1%), four were identified exclusively in four individuals in the low HDL group, two in three individuals in the high HDL group, and one was present in both HDL groups (Table 2). Together, the frequencies of mutations in *ZNF202* in the low and high HDL groups were similar (4–5%).

Haplotype analysis of the five SNPs located in and around the coding region estimated that six haplotypes accounted for >98% of all haplotypes in the low HDL cholesterol group and for almost 100% in the high HDL cholesterol group (**Table 3**). However, none of these haplotypes differed in frequency between the low and high HDL groups, suggesting that they did not affect HDL cholesterol levels.

Individuals in the general population

With the exception of a synonymous SNP [c.731A>G (p.V244V)], we genotyped the total general population

```
PGELQSWVRGQRPESGEEAVTLVEGLQKQPRRPRRWVTVHVHGQEVLSEE 148 Homo sapiens
PGELQSWVRGQRPESGEEAVTLVEGLQKQPRRPRRWVTVHVQGQEVLSEE 148 Mus musculus
PGELQSWVRGQRPESGEEAVTLVEGLQKQPRRPRRWVTVHVQGQEVLSEE 150 Rattus norvegicus
****************************************** ********
   p.A154V(c.461C>T)
TVHLGAEPESPNELQDPVQSSTPEQSPEETTQSPDLGAPAEQRPHQEEEL 198 Homo sapiens
TLHLGLEPESSSEQQDPTQTLTTEQPHEEALRSPDLGTQEQETLQHDEEH 198 Mus musculus
TLHLGVEPESSNEQQEPTQTLTAEQPHEEALRSTDLGTLEQESLQHEGER 200 Rattus norvegicus
* *** ****  * * * *  * **  **   * ***           *

    p.K259E(c.775A>G)  p.V274L(c.820G>C)
DQWSDLDPTQKEFYGEYVLEEDCGIVVSLSFPIPRPDEISQVREEEPWVP 298 Homo sapiens
DQWSDLDPTQKEFYGEYVLEEDCGIVVSLSFPIPRLDDTSQIREEEPQVP 298 Mus musculus
DQWSDLDPTQKEFYGEYVLEEDCGIVVSLSFPIPRLDDPSQIREEEPQVP 300 Rattus norvegicus
************************************ *  ** ***** **

HRRYLAHRKTHAAEELYLCSECGRCFTHSAAFAKHLRGHASVRPCRCNEC 598 Homo sapiens
QKQYLTHRKTHVSEEHHLCNECGRSFSHSAAFAKHLKGHASVRNCRCDEC 590 Mus musculus
QKQYLAHRKTHVSEENHLCSECGRSFNHSAAFAKHLKGHASVRNCRCDEC 592 Rattus norvegicus
   ** *****  **  ** **** * ********* ****** *** **
  p.R605W(c.1813A>T)
GKSFSRRDHLVRHQRTHTGEKPFTCPTCGKSFSRGYHLIRHQRTHSEKTS- 648 Homo sapiens
GKTFSRRDHLVRHQRTHTGEKPFTCATCGKSFSRGYHLIRHQRVHTGKTKT 641 Mus musculus
GKSFSRRDHLVRHQRTHTGEKPFTCATCGKSFSRGYHLIRHQRIHTGKTKT 643 Rattus norvegicus
** ********************** ***************** *  **
```

Fig. 2. Partial alignment of human ZNF202 and orthologous murine and rat protein sequences. For alignment, the ClustalW program was used. Parts of the SCAN domain (top, in italics), the KRAB domain (middle), and zinc finger motifs (bottom) are presented and underlined. The four nonsynonymous variants identified in this study are highlighted. SNPs are shown in green, and mutations are shown in red. Asterisks indicate that amino acid residues are identical in all sequences aligned.

TABLE 3. Estimated haplotype frequencies of SNPs in and around the protein-coding region of *ZNF202* in individuals from the general population with the lowest 1% and highest 1% HDL cholesterol levels

Haplotype	Frequency (95% Confidence Interval) Low HDL Cholesterol (n = 190)	High HDL Cholesterol (n = 190)	*P* (Empirical) Haplotype-Specific
ATCAG	39.9% (33.3–47.1)	36.1% (29.8–43.4)	0.53
ACTAT	24.8% (19.1–31.4)	27.8% (22.0–34.7)	0.56
ATCAT	19.8% (14.9–26.3)	18.3% (13.5–24.6)	0.79
TTCAT	11.3% (7.3–16.4)	16.1% (11.7–22.3)	0.18
ACTGT	1.6% (0.3–4.8)	1.6% (0.3–4.8)	1.00
ACTAG	1.2% (0.05–4.0)	0	0.50

SNP, single nucleotide polymorphism. Haplotypes of SNPs are shown in the following order (left to right): c.IVS4-240A>T, c.IVS4-223T>C, c.461C>T (p.A154V), c.775A>G (p.K259E), and c.*2T>G (substitution in the second nucleotide after the translation stop site in the 3′UTR). Haplotypes are ranked according to frequency in the low HDL cholesterol group. The ATCAT haplotype (boldface) harbors wild-type alleles at all loci and differs from the most common haplotype only at the c.*2T>G position underlined.

sample (n = 9,259) for all SNPs located in and around the protein-coding region [c.IVS4-240A>T, c.IVS4-223T>C, c.461C>T (p.A154V), c.775A>G (p.K259E), and c.*2T>G] (Table 2). Allele frequencies of these five SNPs ranged from 1% to 39% in the total general population sample. Genotype frequencies did not differ from those predicted by Hardy-Weinberg equilibrium.

Overall associations (regardless of variation at the other four sites) for each of the five SNPs with HDL cholesterol and apoA-I levels are presented separately by gender in **Fig. 3.** In men (n = 4,064), homozygosity for c.*2G in the 3′UTR was associated with an apparent decrease in HDL cholesterol of ~0.05 mmol/l (ANOVA: $P = 0.05$; posthoc test: TT versus GG, $P = 0.04$, TG versus GG, $P = 0.02$) and a corresponding decrease in apoA-I (ANOVA: $P = 0.006$; posthoc test: TT versus GG, $P = 0.01$, TG versus GG, $P = 0.002$). If *P* values were corrected for multiple comparisons using the Bonferroni method, only the genotype association with apoA-I levels remained significant (corrected *P* value with five comparisons: $P = 0.05/5 = 0.01$). None of the remaining SNPs [c.461C>T (p.A154V), c.775A>G (p.K259E), c.IVS4-223T>C, and c.IVS4-240A>T] were associated with variation in HDL cholesterol or apoA-I levels.

To further explore these results, we tested the isolated single-site effect of the c.*2T>G SNP on HDL cholesterol and apoA-I levels in men by comparing five SNP genotypes differing only at this site (**Fig. 4**). Using this approach, we found an association between c.*2T>G genotype and HDL cholesterol in one of five tests and an association between c.*2T>G genotype and apoA-I levels in three of five tests. After Bonferroni correction, only one association between genotype and apoA-I levels remained significant ($P = 0.01$).

Pairwise LD was tested for the five SNPs located in and around the coding region and genotyped in 9,103 individuals from the general population (**Table 4**). Strong pairwise LD was present for all SNP pairs. However, because allele frequencies varied widely between SNPs (from 1% to 39%), these LDs should be interpreted with caution.

DISCUSSION

With the aim to identify genetic variation in *ZNF202* affecting HDL cholesterol levels in the general population, we used a systematic approach in which we screened 700 bp of the promoter and all protein-coding exons of the *ZNF202* gene in 190 individuals with extreme HDL cholesterol levels selected from a large sample of the general population (n = 9,259). We subsequently genotyped the entire general population sample for the five SNPs located in and around the protein coding region and determined the association with HDL cholesterol and apoA-I levels.

Novel observations in this study include the following: *1*) we identified nine new genetic variants (two SNPs and seven mutations), of which two were located in predicted transcription factor binding sites in the proximal promoter, and three were nonsynonymous variants; *2*) common SNPs in and around the coding region and haplotypes harboring these SNPs did not segregate differently in individuals with low and high HDL cholesterol levels; *3*) these SNPs also did not have a major effect on HDL cholesterol and apoA-I levels in the total general population, as single sites or as combined genotypes differing only at the relevant SNP; and *4*) mutations in *ZNF202* were equally frequent (5%) in individuals with low or high HDL cholesterol levels.

This is the first systematic study to determine the genetic variation in the promoter and in and around the protein-coding region of *ZNF202* in a substantial number of individuals with low or high HDL cholesterol levels and to investigate the role of ZNF202 in HDL metabolism in the general population. Because *ZNF202* has been identified in a low hypoalphalipoproteinemia locus (8, 19), and because in vitro studies suggest that ZNF202 is a transcriptional repressor of several key genes in HDL cholesterol homeostasis (10), genetic variation in *ZNF202* might affect HDL cholesterol levels in plasma and perhaps play a role in the development of atherosclerosis in the general population. However, the data summarized above suggest that genetic variation in and around the protein-coding region of *ZNF202* is not a major determinant of HDL cholesterol or apoA-I levels in the general population.

We detected all previously reported SNPs in the promoter and coding region of the gene as well as two new SNPs, one in the proximal promoter (g.−118G>T) and a nonsynonymous SNP in exon 8 [c.775A>G (p.K259E)]. The c.775A>G (p.K259E) SNP was the least frequent of the six SNPs identified in and around the coding region, suggesting that we did not overlook any important SNPs. Three SNPs were identified in the protein-coding region, one synonymous [c.731A>G (p.V244V)] and two nonsynonymous [c.461C>T (p.A154V) and c.775A>G (p.K259E)], and one in the 3′UTR (c.*2T>G). The c.461C>T (p.A154V) SNP is located in the SCAN

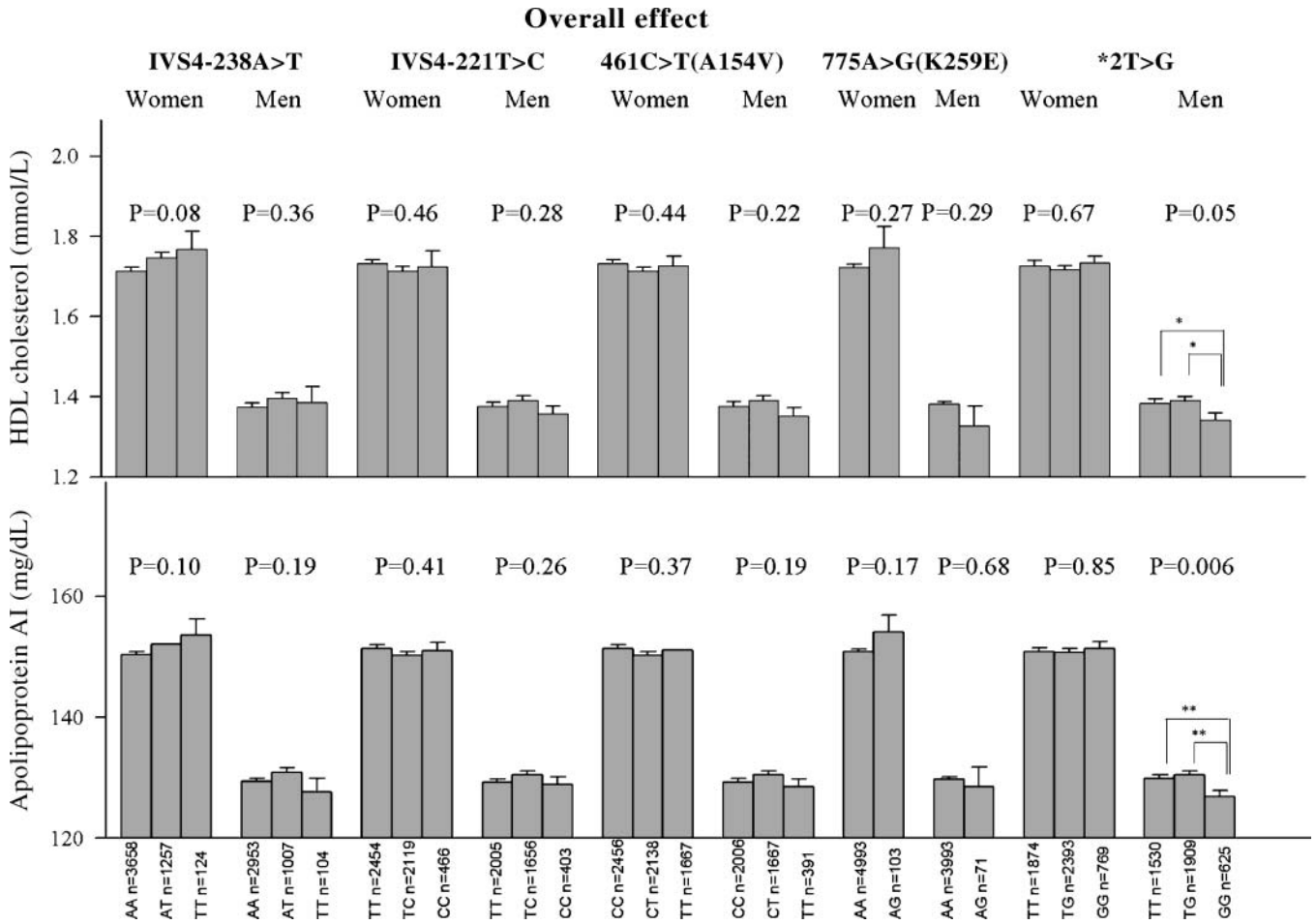

Fig. 3. Mean plasma HDL cholesterol and apolipoprotein A-I (apoA-I) levels as a function of five SNPs [c.IVS4-240A>T, c.IVS4-223T>C, c.461C>T(p.A154V), c.775A>G(p.K259E), and c.*2T>G] in *ZNF202* regardless of variation at the other four sites (overall association) in women (n = 5,039) and men (n = 4,064) in the general population. Values shown are means ± SEM. *P* values are by ANOVA or Student's *t*-test. After Bonferroni correction for multiple comparisons, only $P \leq 0.01$ remained significant (corrected value with five comparisons: $P = 0.05/5 = 0.01$). * $P < 0.05$, ** $P \leq 0.01$ by posthoc test.

domain of ZNF202. The SCAN domain is important for protein-protein interaction and thus regulates ZNF202 activity (11, 20, 21). However, A154 was not conserved between species (human, mouse, rat), and alanine and valine are very similar amino acids. c.461C>T (p.A154V) was equally distributed between the two HDL cholesterol groups, suggesting that this SNP did not affect HDL cholesterol levels. This was confirmed when genotyping c.461C>T (p.A154V) in 9,103 individuals from the general population; no association between c.461C>T (p.A154V) genotype and HDL cholesterol levels was found. c.731A>G (p.V244V) and c.775A>G (p.K259E) are both located in the KRAB A domain, an important repressor domain. The KRAB A domain is highly conserved, and selected conserved amino acids have been shown to be essential for repression (22). K259 is highly conserved between species (human, mouse, rat), and the positive charge of the K259 residue is also conserved between paralogous human proteins with the same protein architecture, SCAN-KRAB-Cys_2His_2, as ZNF202. K259 is located close to residues proven to be essential for the repressor activity of the KRAB A domain (22); therefore, it is possible that a SNP in this position could affect interaction with other protein residues, leading to decreased repressor activity. However, c.775A>G (p.K259E) was identified in both high and low HDL cholesterol groups with equal frequencies and showed no association with HDL cholesterol levels in the general population. c.*2T>G is a very common SNP located in the 3′UTR, 2 bp downstream of the translation stop site. In men in the general population, c.*2G homozygosity appeared to be associated with marginally lower levels of HDL cholesterol and apoA-I compared with c.*2T heterozygotes and homozygotes. However, this association was not robust and may be attributable to a type I error: only the association with apoA-I levels remained significant after correction for multiple comparisons. Thus, iteration in a large, independent population to confirm or reject the association between c.*2G homozygosity and low HDL cholesterol and apoA-I levels would be desirable.

The HapMap data for Caucasians uses nine SNPs [in introns and protein-coding exons, including dbSNP 1144507 and c.461C>T (p.A154V); Table 2] in 60 indi-

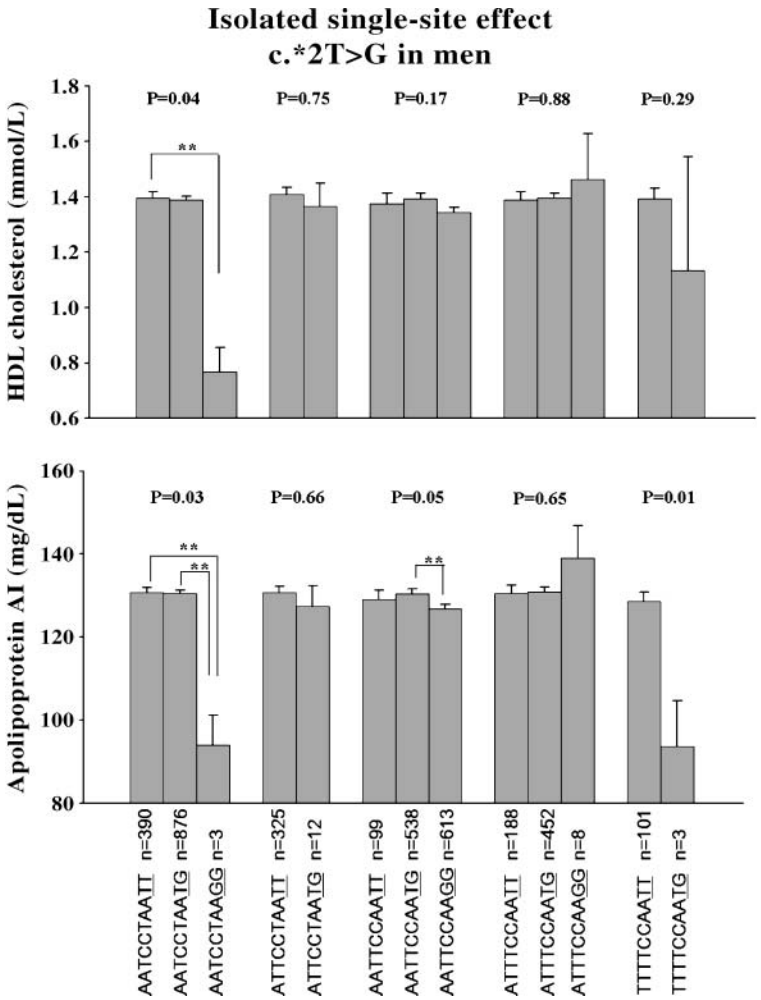

Fig. 4. Plasma HDL cholesterol and apoA-I levels in men in the Copenhagen City Heart Study as a function of five SNP genotypes differing only at c.*2T>G (isolated single-site association). Values shown are means ± SEM. *P* values are by one-way ANOVA and Student's *t*-test. After Bonferroni correction for multiple comparisons, only $P \leq 0.01$ remained significant (corrected value with five comparisons: $P = 0.05/5 = 0.01$). * $P < 0.05$, ** $P \leq 0.01$ by posthoc test.

viduals (120 alleles) spanning both non-protein-coding and protein-coding exons of *ZNF202* to define four haplotypes with frequencies of >1% and three with frequencies of >5% (www.hapmap.org). In our sample of individuals with low and high HDL cholesterol, we used five SNPs in 190 individuals (380 alleles) spanning the protein-coding exons only to define five haplotypes with frequencies of >1% and four with frequencies of >5%. Adding one (g.+34 G>A in exon 1; Table 2) of only two SNPs reported in non-protein-coding exons from the National Center for Biotechnology Information SNP database (http://www.ncbi.nlm.nih.gov/entrez/) to our haplotype data did not add additional haplotypes in our sample. Finally, had we used HapMap SNPs spanning only the protein-coding exons to determine haplotypes in our data (assuming that the same haplotypes are found in the two groups of Caucasians), we would have detected only two common haplotypes.

Four new mutations were identified in and around the protein-coding region, two of which were nonsynonymous. The affected amino acid residues were located in functionally important areas of ZNF202 in the KRAB A

TABLE 4. Pairwise linkage disequilibrium between *ZNF202* SNPs in and around the protein-coding region of *ZNF202* in the general population (n = 9,103)

	c.IVS4-240A>T	c.IVS4-223T>C	c.461C>T(p.A154V)	c.775A>G(p.K259E)	c.*2T>G
c.IVS4-240A>T		−0.9	−1.0	−1.0	−1.0
c.IVS4-223T>C	—		+1.0	−0.9	−1.0
c.461C>T(p.A154V)	—	—		+1.0	−1.0
c.775A>G(p.K259E)	—	—	—		−1.0
c.*2T>G	—	—	—	—	

Disequilibrium statistics are reported as D′, ranging from −1.0 to +1.0 ($D' = D/D_{max}$ when $D > 0$, $D' = D/D_{min}$ when $D < 0$). $D = h - pq$, where h is the frequency of the rare estimated haplotype for a pair of sites and p and q are the frequencies, assuming no linkage, of the alleles in that haplotype. D_{max} is $p(1 - q)$ and D_{min} is pq, assuming that $p < q < 0.5$. Plus signs indicate that rare alleles at each locus segregate together; minus signs indicate that the rare allele at one locus segregates with the common allele at the other locus. For all pairwise linkage disequilibria, $P < 0.001$.

domain [c.820G>C (p.V274L)] and in the zinc finger domain [c.1813A>T (p.R605W)]. V274 is conserved between species but does not introduce a change in amino acid charge. R605 is conserved between species but not between paralogous proteins. However, a change from a positively charged nonpolar amino acid (arginine) to an uncharged polar amino acid (tryptophan) might alter the conformation of the zinc finger motif and perhaps change the DNA binding specificity.

The systematic approach by which we screened individuals with extreme HDL cholesterol levels has previously proven sensitive in the detection of both mutations with strong phenotypic effects and SNPs with more modest effects on HDL cholesterol levels, illustrated by differential segregation of the functional SNPs in groups with extreme phenotypes (15, 23). Whether one can detect such a frequency difference between extreme phenotype groups depends on the size of the study, the frequency of the SNP, the order of magnitude of the phenotype effect of the SNP (in this case, on HDL cholesterol or apoA-I), and whether this effect is equally strong in both genders. In this study, we used this approach to determine whether genetic variation in *ZNF202* affected HDL cholesterol, because the majority of ZNF202 target genes play a role in HDL metabolism. However, because ZNF202 is a transcriptional repressor of the actual functional gene, genetic variation in *ZNF202* may have a less distinct effect on the intermediate phenotype than genetic variation in the structural gene itself.

Limitations

HDL cholesterol levels are affected by many factors, both environmental and genetic, among them body mass index (BMI), triglycerides, and genetic variation in other genes (*cholesteryl ester transfer protein*, *hepatic lipase*, *lipoprotein lipase*, *endothelial lipase*, *APOAI*, *APOE*, and others). This is illustrated in our study by the significant increase in mean BMI in the low HDL group compared with the high HDL group (27 kg/m^2 in women and 25 kg/m^2 in men compared with 23 and 24 kg/m^2, respectively) and by triglyceride concentrations in the low HDL group that were 2.4- to 3.8-fold higher than those in high HDL individuals. Nevertheless, we knew from experience in previous studies (15) that by selecting individuals in the low HDL group with normal BMI and triglycerides only we would have failed to detect any of the nine carriers with mutations associated with low HDL cholesterol, so we chose only to correct for age and gender in the extreme HDL groups. Had we only included individuals with BMI < 25 kg/m^2 and/or triglycerides < 2.2 mmol/l in the low HDL group, we would have failed to detect five of six mutations identified in the low HDL group.

The high HDL cohort had extremely elevated HDL cholesterol concentrations (2.9–3.3 mmol/l), so we cannot rule out that a subset of these subjects may harbor functional variants in other genes, such as *cholesteryl ester transfer protein*, *hepatic lipase*, *lipoprotein lipase*, and *endothelial lipase*, that are also associated with increased HDL cholesterol levels.

Finally, we screened a relatively small portion of the promoter (700 bp), all protein-coding exons (exons 5–10), and the corresponding exon/intron boundaries of *ZNF202*, raising the possibility that functional variants affecting gene expression, or variants in introns or non-protein-coding exons affecting gene regulation, may have been missed. In support of intronic variants being important for gene regulation, a functional variant in an intron was recently identified in the USF1 gene (24).

In conclusion, we show that genetic variation in *ZNF202* is common in the general population. However, SNPs in and around the protein-coding region of *ZNF202* do not make a major contribution to HDL cholesterol levels in the general population.

The authors thank Mette Refstrup for expert technical assistance. The authors also thank the subjects who participated in the study. This work was supported by the Danish Heart Foundation, the Danish Medical Research Council, Ingeborg and Leo Dannin's Grant, and the Research Fund at Rigshospitalet, Copenhagen University Hospital.

REFERENCES

1. Stampfer, M. J., F. M. Sacks, S. Salvini, W. C. Willett, and C. H. Hennekens. 1991. A prospective study of cholesterol, apolipoproteins, and the risk of myocardial infarction. *N. Engl. J. Med.* **325:** 373–381.
2. Genest, J., Jr., M. Marcil, M. Denis, and L. Yu. 1999. High density lipoproteins in health and in disease. *J. Investig. Med.* **47:** 31–42.
3. Tall, A. R., J. L. Breslow, and E. M. Rubin. 2001. Genetic disorders affecting plasma high-density lipoproteins. *In* The Metabolic and Molecular Bases of Inherited Disease. C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, editors. McGraw-Hill, New York. 2915–2936.
4. Friedlander, Y., J. D. Kark, and Y. Stein. 1986. Biological and environmental sources of variation in plasma lipids and lipoproteins: the Jerusalem Lipid Research Clinic. *Hum. Hered.* **36:** 143–153.
5. Hunt, S. C., S. J. Hasstedt, H. Kuida, B. M. Stults, P. N. Hopkins, and R. R. Williams. 1989. Genetic heritability and common environmental components of resting and stressed blood pressures, lipids, and body mass index in Utah pedigrees and twins. *Am. J. Epidemiol.* **129:** 625–638.
6. Perusse, L., J. P. Despres, A. Tremblay, C. Leblanc, J. Talbot, C. Allard, and C. Bouchard. 1989. Genetic and environmental determinants of serum lipids and lipoproteins in French Canadian families. *Arteriosclerosis.* **9:** 308–318.
7. Prenger, V. L., T. H. Beaty, and P. O. Kwiterovich. 1992. Genetic determination of high-density lipoprotein-cholesterol and apolipoprotein A-1 plasma levels in a family study of cardiac catheterization patients. *Am. J. Hum. Genet.* **51:** 1047–1057.
8. Kort, E. N., D. G. Ballinger, W. Ding, S. C. Hunt, B. R. Bowen, V. Abkevich, K. Bulka, B. Campbell, C. Capener, A. Gutin, et al. 2000. Evidence of linkage of familial hypoalphalipoproteinemia to a novel locus on chromosome 11q23. *Am. J. Hum. Genet.* **66:** 1845–1856.
9. Collins, T., J. R. Stone, and A. J. Williams. 2001. All in the family: the BTB/POZ, KRAB, and SCAN domains. *Mol. Cell. Biol.* **21:** 3609–3615.
10. Wagner, S., M. A. Hess, P. Ormonde-Hanson, J. Malandro, H. Hu, M. Chen, R. Kehrer, M. Frodsham, C. Schumacher, M. Beluch, et al. 2000. A broad role for the zinc finger protein ZNF202 in human lipid metabolism. *J. Biol. Chem.* **275:** 15685–15690.
11. Porsch-Ozcurumez, M., T. Langmann, S. Heimerl, H. Borsukova, W. E. Kaminski, W. Drobnik, C. Honer, C. Schumacher, and G. Schmitz. 2001. The zinc finger protein 202 (ZNF202) is a transcriptional repressor of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene expression and a modulator of cellular lipid efflux. *J. Biol. Chem.* **276:** 12427–12433.

12. Langmann, T., C. Schumacher, S. G. Morham, C. Honer, S. Heimerl, C. Moehle, and G. Schmitz. 2003. ZNF202 is inversely regulated with its target genes ABCA1 and apoE during macrophage differentiation and foam cell formation. *J. Lipid Res.* **44:** 968–977.
13. Appleyard, M., A. Tybjærg-Hansen, G. Jensen, P. Schnohr, and J. Nyeboe. 1989. The Copenhagen City Heart Study, Østerbrounder-søgelsen. A book of tables with data from the first examination (1976–78) and a five year follow-up (1981–83). The Copenhagen City Heart Study Group. *Scand. J. Soc. Med. Suppl.* **41:** 1–160.
14. Schnohr, P., G. Jensen, P. Lange, H. Scharling, and A. Appleyard. 2001. The Copenhagen City Heart Study, Østerbroundersøgelsen. Tables with data from the third examination 1991–1994. *Eur. Heart J.* **(Suppl.) 3:** 1–83.
15. Frikke-Schmidt, R., B. G. Nordestgaard, G. B. Jensen, and A. Tybjærg-Hansen. 2004. Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J. Clin. Invest.* **114:** 1343–1353.
16. Pericak-Vance, M. A. 1998. Linkage disequilibrium and allelic association. *In* Approaches to Gene Mapping in Complex Human Diseases. J. L. Haines and M. A. Pericak-Vance, editors. John Wiley & Sons, New York. 323–333.
17. Thompson, E. A., S. Deeb, D. Walker, and A. G. Motulsky. 1988. The detection of linkage disequilibrium between closely linked markers—RFLPs at the AI-CIII apolipoprotein genes. *Am. J. Hum. Genet.* **42:** 113–124.
18. Excoffier, L., and M. Slatkin. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* **12:** 921–927.
19. Monaco, C., C. M. Helmer, E. Caprini, I. Vorechovsky, G. Russo, C. M. Croce, G. Barbanti-Brodano, and M. Negrini. 1998. Molecular cloning and characterization of ZNF202: a new gene at 11q23.3 encoding testis-specific zinc finger proteins. *Genomics.* **52:** 358–362.
20. Schumacher, C., H. Wang, C. Honer, W. Ding, J. Koehn, Q. Lawrence, C. M. Coulis, L. L. Wang, D. Ballinger, B. R. Bowen, et al. 2000. The SCAN domain mediates selective oligomerization. *J. Biol. Chem.* **275:** 17173–17179.
21. Babb, R., and B. R. Bowen. 2002. SDP1 is a PPARgamma2 coactivator that binds through its SCAN domain. *Biochem. J.* **370:** 719–727.
22. Witzgall, R., E. O'Leary, A. Leaf, D. Onaldi, and J. V. Bonventre. 1994. The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc. Natl. Acad. Sci. USA.* **91:** 4514–4518.
23. Cohen, J. C., R. S. Kiss, A. Pertsemlidis, Y. L. Marcel, R. McPherson, and H. H. Hobbs. 2004. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science.* **305:** 869–872.
24. Naukkarinen, J., M. Gentile, A. Soro-Paavonen, J. Saarela, H. A. Koistinen, P. Pajukanta, M. R. Taskinen, and L. Peltonen. 2005. USF1 and dyslipidemias: covering evidence for functional intronic variant. *Hum. Mol. Genet.* **14:** 2595–2605.
25. den Dunnen, J. T., and S. E. Antonarakis. 2001. Nomenclature for the description of human sequence variations. *Hum. Genet.* **109:** 121–124.