Arteriosclerosis, Thrombosis, and Vascular Biology

JOURNAL OF THE AMERICAN HEART ASSOCIATION

Common and Rare *ABCA1* Variants Affecting Plasma HDL Cholesterol

Jian Wang, John R. Burnett, Susan Near, Kue Young, Bernard Zinman, Anthony J. G. Hanley, Philip W. Connelly, Stewart B. Harris and Robert A. Hegele

Arterioscler Thromb Vasc Biol. 2000;20:1983-1989
doi: 10.1161/01.ATV.20.8.1983
Arteriosclerosis, Thrombosis, and Vascular Biology is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1079-5642. Online ISSN: 1524-4636

Common and Rare *ABCA1* Variants Affecting Plasma HDL Cholesterol

Jian Wang, John R. Burnett, Susan Near, Kue Young, Bernard Zinman, Anthony J.G. Hanley, Philip W. Connelly, Stewart B. Harris, Robert A. Hegele

Abstract—Mutations in *ABCA1*, a member of the ATP-binding cassette family, have been shown to underlie Tangier disease (TD) and familial hypoalphalipoproteinemia (FHA), which are genetic disorders that are characterized by depressed concentrations of plasma high density lipoprotein (HDL) cholesterol. An important question is whether common variants within the coding sequence of *ABCA1* can affect plasma HDL cholesterol in the general population. To address this issue, we developed a screening strategy to find common *ABCA1* variants. This strategy involved long-range amplification of genomic DNA by using coding sequences only, followed by deep sequencing into the introns. This method helped us to characterize a new set of amplification primers, which permitted amplification of virtually all of the coding sequence of *ABCA1* and its intron-exon boundaries with a single DNA amplification program. With these new sequencing primers, we found 3 novel *ABCA1* mutations: a frameshift mutation (4570insA, A1484S→X1492), a missense mutation (A986D) in a TD family, and a missense mutation (R170C) in aboriginal subjects with FHA. We also used these sequencing primers to characterize 4 novel common amino acid variants in *ABCA1*, in addition to 5 novel common silent variants. We tested for association of the *ABCA1* I/M823 variant with plasma HDL cholesterol in Canadian Inuit and found that M823/M823 homozygotes had significantly higher plasma HDL cholesterol compared with subjects with the other genotypes. The results provide proof of principle of the effectiveness of this approach to identify both rare and common *ABCA1* genomic variants and also suggest that common amino acid variation in *ABCA1* is a determinant of plasma HDL cholesterol in the general population. **(*Arterioscler Thromb Vasc Biol*. 2000;20:1983-1989.)**

Key Words: DNA ■ genetics ■ complex disease ■ susceptibility ■ aboriginal

Mutations in the gene encoding ATP-binding cassette transporter 1 (*ABCA1*) have been reported in patients with Tangier disease (TD),[1–5] an autosomal recessive disorder characterized by extremely depressed plasma concentrations of HDL cholesterol, abnormal deposition of cholesteryl esters in the reticuloendothelial system, and disordered cellular cholesterol efflux. *ABCA1* has been shown to regulate the apolipoprotein- (apo) mediated lipid removal pathway from cells,[6] and both TD patients and mice with a disrupted *Abca1* gene have defective transport of lipids from the Golgi to the plasma membrane.[7] Furthermore, mutations in *ABCA1* have been found in some patients with familial hypoalphalipoproteinemia (FHA),[8] a relatively more common disorder than TD. Cellular cholesterol efflux appears to be abnormal in some patients with FHA due to mutant *ABCA1*, but there is no obvious tissue deposition of cholesterol esters, and the low plasma HDL cholesterol concentrations demonstrate an apparent autosomal codominant pattern of expression in *ABCA1* mutation carriers.[8] Therefore, *ABCA1* is an attractive candidate gene for the determination of plasma HDL cholesterol concentration in the general population.

To this point, common variants and polymorphisms in *ABCA1* and their potential association with plasma concentrations of HDL cholesterol have not been reported. To identify common genomic variants in *ABCA1* for association studies, we applied a genomic screening strategy under a single set of amplification primers, which would amplify almost all of the *ABCA1* coding sequence by using the same amplification conditions. Using these primers to sequence genomic DNA, we discovered 2 new TD mutations, 1 new FHA mutation, and many common variants in *ABCA1*, 1 of which was associated with variation in plasma HDL cholesterol.

Methods

Subjects

TD Kindred

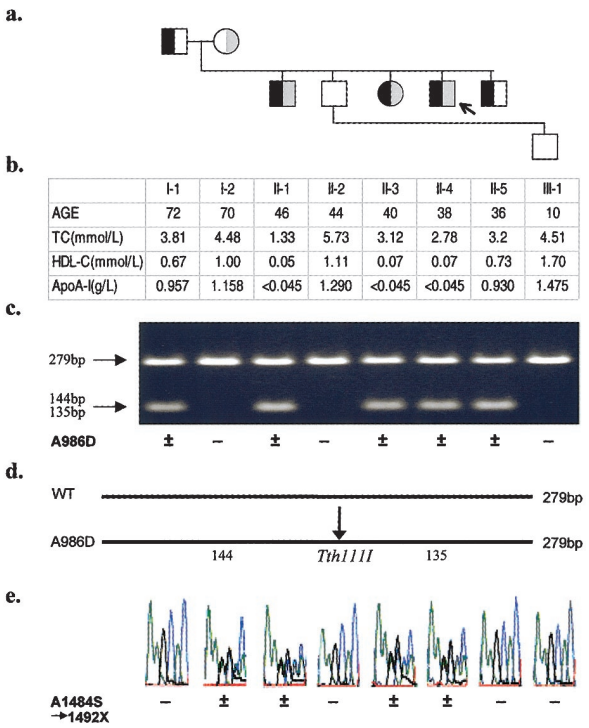
The proband in the TD family (the Figure; subject II-4) was previously reported to have severe aortic stenosis, associated with

Received April 20, 2000; revision accepted May 30, 2000.

From the John P. Robarts Research Institute (J.W., S.N., R.A.H.), London, Ontario, Canada; the Department of Core Clinical Pathology and Biochemistry (J.R.B.), Royal Perth Hospital, Perth, Australia; the Northern Health Research Unit (K.Y.), Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada; the Samuel Lunenfeld Research Institute (B.Z., A.J.G.H.), Mount Sinai Hospital, Toronto, Ontario; St. Michael's Hospital (P.W.C.), University of Toronto, Toronto, Ontario, Canada; and the Thames Valley Family Practice Research Unit (S.B.H.), University of Western Ontario, London, Ontario, Canada.

Correspondence to Robert A. Hegele, MD, Blackburn Cardiovascular Genetics Laboratory, Robarts Research Institute, 406-100 Perth Dr, London, Ontario, Canada N6A 5K8. E-mail robert.hegele@rri.on.ca

***Arterioscler Thromb Vasc Biol.* is available at http://www.atvbaha.org**



	I-1	I-2	II-1	II-2	II-3	II-4	II-5	III-1
AGE	72	70	46	44	40	38	36	10
TC(mmol/L)	3.81	4.48	1.33	5.73	3.12	2.78	3.2	4.51
HDL-C(mmol/L)	0.67	1.00	0.05	1.11	0.07	0.07	0.73	1.70
ApoA-I(g/L)	0.957	1.158	<0.045	1.290	<0.045	<0.045	0.930	1.475

Segregation of *ABCA1* mutations in a TD kindred. (a) The TD proband is indicated by an arrow. Solid shading indicates carrier status for *ABCA1* A986D, and half-tone shading indicates carrier status for the 4570insA mutation. (b) Age and selected biochemical data from 8 family members are shown. TC indicates total cholesterol. (c) Segregation of *ABCA1* A986D mutation by restriction enzyme analysis, with genotypes shown beneath gel ("±" means D986/A986 heterozygote and "−" means an A986/A986 homozygote). (d) *Tth111*I restriction map for *ABCA1* A986D, with arrow indicating a polymorphic *Tth111*I site in the A986D allele. WT indicates wild type, or A986 allele. (e) Segregation of 4570insA mutation shown by results of genomic DNA sequencing of *ABCA1* with genotypes shown beneath tracings ("±" means heterozygous for the frameshift mutation and "−" means homozygous for the normal allele).

deposition of cholesteryl ester in the aortic valve, and early atherosclerosis.[9] The kindred has been subsequently extended and a total of 8 members have been studied, including both parents (subjects I-1 and I-2); 2 siblings with TD (II-1 and II-3); a sibling with low HDL cholesterol, but no other findings, thus being clinically compatible with heterozygous FHA (II-5); and a clinically normal brother and his son (II-2 and III-1, respectively). The biochemical data from these 8 family members are shown in the Figure.

FHA Subjects

We selected from our dyslipidemia archival sample repository 6 subjects who conformed to stringent biochemical criteria for heterozygous FHA, which were plasma HDL cholesterol and/or apo AI concentration below the second percentile for age and sex, with plasma triglyceride concentration below the 50th percentile for age and sex. None of these subjects had any physical manifestations compatible with TD. This screening sample with FHA included 3 aboriginal Canadian subjects, 2 taken from our well-characterized sample of Oji-Cree[10] and 1 taken from our well-characterized sample of Inuit.[11]

Control Samples

Samples from normolipidemic subjects taken from 6 ethnic groups (38 Europeans, 44 Oji-Cree, 30 Inuit, 36 Africans, 37 Chinese, and 38 South Asians) were studied to determine allele and genotype frequencies when a new polymorphism was identified. Association analyses with plasma HDL cholesterol concentration were carried out in normolipidemic samples by using amino acid polymorphisms that were shown to have sufficient heterozygosity or information content, thus ensuring sufficient numbers of subjects with each genotype for parametric statistical analysis.

Determining the Intron-Exon Boundaries of *ABCA1*

Genomic DNA was isolated from the whole blood of all study subjects by using an established method (Puregene, Gentra Systems). Long-distance polymerase chain reaction (LPCR) amplification was performed to identify intron-exon boundaries of the *ABCA1* gene. Human *ABCA1* cDNA sequence (Genome Data Base AJ012376) and murine *ABCA1* cDNA sequence (Genome Data Base X75926) were aligned and used to infer the approximate position of each exon for the human *ABCA1* gene. LPCR primers were designed with the intention of annealing within the deduced interior regions of exons, although there could be no a priori guarantee of avoiding annealing to intron-exon boundaries, which would result in a failure to amplify. In such an instance, different primers would be designed and the procedure would be repeated. In addition, primers were chosen to share an optimal annealing temperature of 62°C, which would allow for many reactions to proceed simultaneously within a single amplification device using an identical cycling program, thus increasing the throughput of the entire strategy.

Expand 20-kb Plus enzyme mix (Roche) was used for LPCR according to the manufacturer's instructions. LPCRs were performed in a Perkin-Elmer gene amplification 9600 apparatus. Genomic DNA from a normal subject was used as the amplification template for all of these experiments. The cycle conditions for the LPCR were as follows: denaturing at 92°C for 2 minutes, followed by 10 cycles of a denaturation step at 92°C for 10 seconds, an annealing step at 62°C for 30 seconds, and an extension step at 68°C for 18 minutes. An autoextension feature was used to add 10 seconds to the extension step in each cycle for 20 cycles and a final 7-minute extension step at 68°C. The products of successful amplification reactions from genomic DNA ranged in size from 4 to 15 kb. Each successful LPCR was subsequently found to have amplified between 4 and 6 exons of *ABCA1*.

LPCR amplification products were electrophoresed in 0.6% agarose gels and purified by using the QIAEX II gel extraction kit (QIAGEN). Each pair of LPCR primers was used to define the 3′ boundary of the first exon by using the 5′-most primer and the 5′ boundary of the last exon by using the 3′-most primer of each LPCR. For the exons contained within each LPCR product, internal primers were then designed to sequence in forward and reverse directions, beginning within the deduced central regions of the exons. When successful, these tracings would begin with a known exon sequence but would either end before the intron-exon boundary, in which case a new primer would be designed and the sequencing strategy would continue in a "walking" fashion, or continue through the intron-exon junction and into a previously unknown intron sequence. The position of an intron-exon boundary within the LPCR product was obvious after alignment with the published cDNA sequences by using Sequence Navigator (Applied Biosystems). When additional intron sequence information was required, a sequencing primer could be designed by using the newly derived intron sequence information, and the sequencing could continue, in theory, throughout the whole intron.

Once the intron-exon boundaries were determined (Table I; please see http://www.ahajournals.org), the *ABCA1* gene–coding region in study samples could be sequenced with different PCR primers, which were designed to amplify the genomic region containing each individual exon, the intron-exon boundaries, and a long intronic stretch based on the genomic sequence obtained from the LPCR mapping strategy described above. In addition, the primers were chosen to share an optimum annealing temperature of 56°C (Table II; please see http://www.ahajournals.org), which allowed for optimal use of a single amplification apparatus.

ABCA1 Genomic Mutations and Polymorphisms

Samples from 7 subjects were analyzed extensively: TD family subject II-3 and the 6 subjects with FHA, as described above. Direct

genomic DNA sequencing of the individual exons was carried out by using the primers in Table II (please see http://www.ahajournals.org). All amplification reactions were carried out in a 50-μL mixture containing 32 pmol of each primer; 0.2 mmol/L each of dATP, dCTP, dGTP, and dTTP; 1.5 mmol/L $MgCl_2$; 50 mmol/L KCl; 20 mmol/L Tris HCl (pH 8.4); and 2.5 U *Taq* Platinum DNA polymerase (GIBCO BRL, Life Technologies). Each amplification cycle consisted of a denaturation step at 94°C, an annealing step at 56°C, and an extension step at 72°C for 30 seconds each, followed by a final extension step for 10 minutes at 72°C and cooling to 4°C. The amplification products were then run on 1.5% agarose gels and purified with the QIAEX II gel extraction kit (QIAGEN). Purified amplification products were directly sequenced by the ABI 377 Prism automated DNA sequencer. DNA sequences were aligned both with published *ABCA1* exon sequences[3,4] and with each other by using ABI Sequence Navigator software.

Once a sequence variant was identified in a sample, its presence was confirmed by fresh reamplification and resequencing of the culprit exon from the patient sample on another day. When the DNA change, which was usually a single nucleotide polymorphism (SNP), affected a restriction endonuclease recognition site, restriction fragment length polymorphism analysis would be used as the basis of a rapid method to genotype the specific SNP for the purposes of determining allele and genotype frequencies in various ethnic groups and for studies of association with plasma HDL cholesterol. In contrast, when the DNA change or SNP did not affect a restriction endonuclease recognition site, genotypes would be inferred from electrophoretogram traces from direct genomic DNA sequencing to determine allele and genotype frequencies in whites only.

Statistical Analysis

The significance of deviations of the observed genotype frequencies from those predicted by the Hardy-Weinberg equation were evaluated with χ^2 tests. SAS (version 6.12) was used for statistical comparisons involving HDL cholesterol.[12] To test whether a common variation in *ABCA1* was associated with variation in plasma HDL cholesterol, we genotyped the *ABCA1* I/M823 variant in a previously reported normolipidemic Inuit sample.[11] We assumed a recessive model for M823, in which the genotype variable was set at 0 for subjects with I823/I823 and M823/I823 genotypes and 1 for subjects with the M823/M823 genotype. Covariates in the ANOVA were age, sex, body mass index (BMI), and current cigarette smoking.[11] The proportion of variation in a plasma HDL cholesterol concentration from each covariate was estimated from partial regression coefficients.

Results

Intron-Exon Boundaries of *ABCA1* and Primers for Genomic Sequencing

The strategy outlined in the Methods section allowed for determination of >96% of the intron-exon boundaries within a period of 6 weeks (Table I; please see http://www.ahajournals.org). Owing to insufficient human cDNA sequence information, we could not determine the 2 boundaries of exon 1 and the 5′ boundary of exon 2. For as-yet-undetermined reasons, the 5′ boundary of exon 5 and the 3′ boundaries of exon 4 were not accessible through this strategy. The availability of heretofore unpublished intron sequences allowed for the development of amplification primers for individual exons (Table II; please see http://www.ahajournals.org). These primers could amplify at a constant annealing temperature, allowing for all amplification of all exons from a single genomic source simultaneously within a single amplification apparatus.

Rare Mutations in *ABCA1* in TD and FHA

The sequencing strategy outlined above resulted in the identification of 2 novel *ABCA1* mutations in the TD family. These were a frameshift mutation in exon 33, namely 4570insA, A1484S→X1492, and a missense mutation in exon 21, namely A986D (the Figure and Table I). Although the frameshift mutation within codon 1484 did not alter a restriction site, it was easily genotyped because of the characteristic appearance of the electrophoretogram tracings of genomic DNA sequences of exon 33 of *ABCA1* in heterozygotes for the frameshift (the Figure). Digestion of an amplified fragment containing exon 21 with restriction endonuclease *Tth111*I revealed a 279-bp fragment for alleles containing A986 and smaller 144- and 135-bp fragments for alleles containing D986 (the Figure).

The sequencing strategy also resulted in the identification of a rare *ABCA1* variant, namely R170C, in 1 of the subjects with FHA who was part of the screening experiment. This subject was Oji-Cree, and subsequent screening of 80 Oji-Cree samples by using mismatch amplification primers and digestion with *Sfu*I revealed 2 more Oji-Cree heterozygotes for *ABCA1* C170/R170. Both of these subjects were found to have plasma HDL cholesterol concentrations below the 5th percentile for age and sex. However, the mutation was completely absent from 230 subjects from other ethnic groups, suggesting that it is probably a population-specific variant that determines FHA in Oji-Cree. Its presence in such a small number of Oji-Cree suggests that it arose relatively recently in the history of these people.

Common Amino Acid Variants in *ABCA1*

The sequencing strategy outlined above resulted in the identification of 4 common amino acid variants in *ABCA1*, namely R/K159 in exon 6, V/I765 in exon 16, I/M823 in exon 17, and R/K1527 in exon 34 (Table 1) in the subjects who had been selected for the screening experiments. Restriction analysis by mismatch priming and digestion with restriction endonucleases *Bsm*I and *Bst*NI for I/M823 and R/K1527, respectively, was carried out to determine the genotype and allele frequencies of these markers in the normolipidemic ethnic samples. Restriction analysis with the use of the naturally occurring *Dpn*II recognition site affected by V/I765 was carried out to determine the genotype and allele frequencies of this marker in the normolipidemic ethnic samples. Genotypes for the R/K159 variant were determined from the characteristic appearance of electrophoretogram tracings of genomic DNA sequences of exon 6 for calculation of the genotype and allele frequencies in whites only. None of the genotype frequencies for any of the variants deviated from the expectations of the Hardy-Weinberg equation in any study sample.

The *ABCA1* R/K159, V/I765, I/M823, and R/K1527 variants were each markedly polymorphic in whites (Table 1). The *ABCA1* V/I765, I/M823, and R/K1527 variants were each also markedly polymorphic in other ethnic groups, with a wide range of allele frequencies.

Common Silent Variants in *ABCA1*

The sequencing strategy outlined above also resulted in the identification of 5 common silent variants in *ABCA1*, namely 876C/T and 888G/A in exon 8, 1980C/A in exon 14, 2820G/A in exon 19, and 4221G/A in exon 30 (Table 1) in white subjects. Although none of these variants altered a

TABLE 1. Characterization of *ABCA1* Genomic Variants Affecting the Amino Acid Sequence

Exon	Nucleotide	Amino Acid	Allele Frequencies	
6	596G/A	R/K159	596A	0.464 in whites
	C628T	R170C	C170	Absent from 230 samples
8	876C/T	P252	876T	0.250 in whites
	888G/A	G256	888A	0.143 in whites
14	1980C/A	I620	1980A	0.321 in whites
16	2413G/A	V/I765	2413A	0.028 in Africans
				0.053 in South Asians
				0.145 in whites
				0.200 in Inuit
				0.250 in Oji-Cree
				0.514 in Chinese
17	2589A/G	I/M823	2589G	0.079 in whites
				0.145 in South Asians
				0.294 in Inuit
				0.419 in Africans
				0.690 in Oji-Cree
				0.737 in Chinese
19	2820G/A	L900	2820A	0.036 in whites
21	3077C/A	A986D	3077A	Absent from 230 samples
30	4221G/A	T1367	4221A	0.179 in whites
33	4570 ins A	A1484S to X1492	4570 ins A	Absent from 230 samples
34	4700G/A	R/K1527	4700A	0.211 in whites
				0.303 in South Asians
				0.378 in Oji-Cree
				0.447 in Chinese
				0.561 in Inuit
				0.724 in Africans

naturally occurring restriction endonuclease recognition site, genotypes could be easily determined from the characteristic appearance of electrophoretogram tracings of genomic DNA sequences of the individual exons. This comparison allowed for calculation of the genotype and allele frequencies in whites only. The *ABCA1* 876C/T, 888G/A, 1980C/A, 2820G/A, and 4221G/A SNPs were each markedly polymorphic in whites (Table 1). None of the genotype frequencies for any of the variants deviated from the expectations of the Hardy-Weinberg equation in any study sample.

In addition, we observed several disparities in the *ABCA1* genomic DNA sequence from the screening sample and sequences that had been previously published. Allele frequencies for these disparate genotypes were determined from the characteristic appearance of electrophoretogram tracings of genomic DNA sequences of the individual exons in 28 whites; for each disparity, the DNA sequence was monomorphic in the control sample (Table 2).

Association of *ABCA1* I/M823 With Plasma HDL Cholesterol Variation

ANOVA revealed that age, sex, BMI (in kg/m^2), smoking, and *ABCA1* I/M823 genotype (assuming a recessive model for M823) were each a significant source of variation for plasma HDL cholesterol in Inuit (Tables 3 and 4). Simple mean HDL cholesterol concentration was significantly higher in M823/M823 homozygotes than in subjects with the I823/I823 and M823/I823 genotypes (Table 4). Multivariate regression analysis indicated that age, sex, BMI, smoking, and *ABCA1* I/M823 genotype (assuming a recessive model for M823) accounted for 29.4%, 42.0%, 13.5%, 8.2%, and 7.0%, respectively, of attributable variation in the plasma HDL cholesterol concentration ($P<0.0001$, <0.0001, $=0.0034$, $=0.021$, and $=0.031$, respectively). Post hoc analyses revealed no associations with other lipoprotein variables, such as plasma triglycerides and apo AI (data not shown).

Discussion

We report herein (1) a strategy to identify intron-exon boundaries of a gene, without any need for cloning; (2) the application of this method to identify intron-exon boundaries of *ABCA1*; (3) the use of the acquired intron information to derive a set of amplification primers that amplified >96% of the *ABCA1* coding sequence at a single annealing temperature, allowing for more effective use of amplification hard-

TABLE 2. Characterization of Monomorphic Sequence Disparities With Published *ABCA1* Sequences

Exon	Nucleotide Disparity	Source	Amino Acid	Allele Frequencies (in Whites)
7	705C/T	Reference 1	A195	705T:1.000
11	gtaa/gtgaa	Reference 4	3 nt to 3′ end exon 11	gtgaa:1.000
33	4604C/T	Reference 1	T1495I	4604T:1.000
35	4883C/T	Reference 1	P1588L	4883T:1.000
43	5861G/A	Reference 1	R1914K	5861A:1.000
44	gtactagt/gtactgtg	Reference 4	6 nt to 3′ end exon 44	gtactgtg:1.000
45	atcactag/atcactgtag	Reference 4	5 nt to 5′ end exon 45	atcactgtag:1.000
45	gtgagtgt/gtgagtat	Reference 4	7 nt to 3′ end exon 45	gtgagtat:1.000
48	6443C/T	Reference 1	P2108L	6443T:1.000

ware; (4) the use of the amplification primers for genomic DNA sequencing, which resulted in (5) the identification of 3 novel rare *ABCA1* mutations in TD and FHA, of which 1 appears to be an Oji-Cree–specific variant; and (6) the identification of 4 novel common amino acid variants and 5 novel common silent variants in *ABCA1*. We have noted a wide range of allele frequencies of these common markers across ethnic groups. Finally, we found an association of the common *ABCA1* I/M823 variant with plasma HDL cholesterol in Canadian Inuit: M823/M823 homozygotes had significantly higher plasma HDL cholesterol compared with subjects with the other genotypes. This finding suggests that common variation in *ABCA1* may be a source of variation of plasma HDL cholesterol in the general population.

The aforementioned method to **t**ransform a cDNA sequence into **g**enomic **a**mplification primers (designated TCGA) allowed us to rapidly identify 12 novel *ABCA1* genomic sequence variants, with no need for cloning. The large quantity of intron sequence provided by TCGA was sufficient to allow the design of genomic amplification primers that fulfilled 3 criteria: (1) the amplified fragment contained the entire exon sequence; (2) the amplified fragment contained both intron-exon boundaries; and (3) the GC content of the primers could be optimized to allow for annealing at a specific temperature. The optimal annealing temperature could be kept constant for all genomic DNA amplification reactions developed this way, thereby permitting many reactions to proceed simultaneously in a single amplification apparatus under a single set of conditions.

In the postgenomic era, expressed sequence tags (ESTs), defined as complete or partial cDNA sequences, can mark genes for further study and analysis. However, the EST database will not necessarily be directly helpful to investigators who seek to identify human mutations from genomic DNA. For this application, reliable primers and conditions are required for amplification from human genomic DNA samples. Because there is no information regarding the intron-exon boundaries for the vast majority of ESTs mapped to date, a robust, general method such as TCGA is required, which might assist the large-scale, rapid conversion of EST data into intron sequence data. The intron sequences can be used, in turn, to create primers for the amplification and sequencing of coding regions from a genomic DNA template. TCGA will not necessarily fully characterize the intron size and complete intron sequence: only enough sequence is obtained to allow for clean amplification of coding regions and margins from a genomic template. However, the "walking" TCGA strategy of sequencing through the intron can be serially used until the next contiguous intron-exon boundary is crossed and the next exon is entered.

There are several reasons to acquire intron sequence for the study of candidate genes in human diseases. First, some diseases can result from mutations directly at or close to the intron-exon boundaries. Thus, sufficient flanking sequence within the intron is required to develop amplification primers to first identify and then to screen for a mutation close to these boundaries. Second, a small subset of some diseases can be due to mutations well inside introns. Again, sufficient flanking sequence within the intron is required to develop amplification primers to identify and then screen for an intron mutation in such patients. Finally, knowledge of intron sequence can permit development of clean and effective amplification strategies that fully capture coding regions, where most mutations would be expected to reside. Furthermore, TCGA allows for the use of genomic DNA from easily

TABLE 3. Sources of Variation of Plasma HDL Cholesterol Concentration in Inuit From ANOVA

Source	F Value	Probability>F
Age	22.9	<0.0001
BMI	32.0	<0.0001
Sex	10.9	0.0011
Current smoking	3.63	0.021
ABCA1 I/M823 genotype	4.93	0.028

TABLE 4. *ABCA1* I/M823 Association With Plasma HDL Concentration in Inuit

	I823/I823 Plus M823/I823	M823/M823
Number	229	16
Age, y	36.6±14.9	37.6±17.5
Females, %	54.7	50.0
BMI, kg/m^2	26.8±4.5	24.5±3.2
HDL cholesterol, mmol/L	1.42±0.39	1.67±0.53*

*P<0.05.

accessible whole-blood samples as the template for amplification and sequencing, with no need for cloning or further modifications. Approaches to find disease mutations in the absence of knowledge of the intron-exon boundaries, such as sequencing of the cDNA derived from reverse transcriptase–treated mRNA from tissues in diseased patients, are more cumbersome by comparison. In addition, such methods depend on mRNA being expressed in the diseased tissue, in contrast to the independence of mRNA expression required to detect DNA changes at the genomic level.

We used TCGA to identify rare *ABCA1* mutations in TD and FHA subjects. The 2 TD family members who were simple heterozygotes for *ABCA1* A986D, namely I-1 and II-5, had depressed though not undetectable plasma concentrations of HDL cholesterol and apo AI (the Figure). In contrast, the plasma HDL cholesterol and apo AI concentrations in the simple heterozygote for 4570insA, namely I-2, were within the normal range (the Figure). This finding suggests that the A986D mutation, which occurs between the Walker domains A and B of the first ABC in *ABCA1*,[1] had an autosomal codominant influence on HDL cholesterol and apo AI. In contrast, the 4570insA mutation, which creates a premature truncation that results in the deletion of most of the second transmembrane domain and the second ABC of *ABCA1*,[1] had a recessive influence on the TD phenotype. The Oji-Cree FHA R170C mutation, which occurs in the 5′ intracytoplasmic domain of *ABCA1*,[1] appeared to have an autosomal codominant influence on the biochemical phenotype. The mechanism(s) through which *ABCA1* mutations can lead to the TD and FHA phenotypes are not understood at present. Presumably, once reliable in vitro tests of function are developed, these 3 rare *ABCA1* mutations can be studied to determine their impact on cellular cholesterol trafficking, which might then contribute to an understanding of the phenotype in intact humans.

To this point, common variants and SNPs in *ABCA1* and their potential associations with plasma HDL cholesterol have not been reported in the literature. The R/K159 variant occurs within the 5′ intracytoplasmic domain of *ABCA1*,[1] the V/I765 variant occurs within the first group of transmembrane domains of *ABCA1*,[1] the I/M823 variant occurs upstream of the Walker A motif of the first ABC of *ABCA1*,[1] and the R/K1527 variant occurs within the second group of transmembrane domains of *ABCA1*.[1] Interestingly, the allele frequencies vary widely across populations (Table 1), a finding whose significance is uncertain at this time.

That common *ABCA1* variants might affect HDL in subgroups of the general population was demonstrated by the finding in normal Inuit that the *ABCA1* I/M823 genotype was a significant source of variation in plasma HDL cholesterol concentration (Tables 3 and 4). It should be noted that the sample sizes of the various ethnic groups were relatively small for this screening experiment and that replication in larger samples would be important. Inuit with the M823/M823 genotype had significantly higher mean plasma HDL cholesterol than did Inuit with the other 2 genotypes (1.42 ± 0.39 versus 1.67 ± 0.53 mmol/L). The I/M823 marker accounted for 7% of the variation in HDL cholesterol, an effect size that was comparable to that of cigarette smoking, although it was smaller than those due to age, sex, and BMI. More association studies with these common *ABCA1* variants are now required to determine whether such modest associations with HDL cholesterol are more general. The mechanism(s) through which this common *ABCA1* variant can affect plasma HDL cholesterol is not clear at present and cannot be resolved by the current study.

In addition to the rare mutations and common variants shown in Table 1, we observed several disparities between the *ABCA1* genomic DNA sequence obtained from the screening sample of whites and *ABCA1* sequences that had been previously published (Table 2). For each disparity, the DNA sequence was monomorphic in our control sample of 28 whites (Table 2), suggesting that small amendments might be required in some of the deposited reference sequences for *ABCA1*.[1,4]

In summary, our results provide proof of principle that TCGA is an effective way to develop primers that are part of an amplification strategy to find mutations in human genomic DNA. TCGA permitted rapid assessment of >96% of the *ABCA1* coding sequence and resulted in the identification of 12 novel *ABCA1* sequence variants. Three of the novel variants were extremely rare and were seen in subjects with either TD or FHA. One of the common amino acid variants, namely *ABCA1* I/M823, was significantly associated with variation in plasma HDL cholesterol in Inuit. The results suggest that TCGA can identify both rare and common *ABCA1* genomic variants and that common variation in *ABCA1* can be a determinant of plasma HDL cholesterol.

Acknowledgments

This work was supported by grants from the National Institutes of Health (1-R21-DK44597-01), the Ontario Ministry of Health (No. 04307), the MRC (Canada, No. MT13430), the Canadian Genetic Diseases Network, and the Blackburn Group. Dr Hegele is a Career Investigator (No. CI-2979) of the Heart and Stroke Foundation of Ontario. Dr Harris is supported by the Ontario Ministry of Health Career Scientist Program. Dr Young is a Senior Scientist of the MRC (Canada). The authors wish to thank Carol Anderson and Pearl Campbell for their expert assistance. We acknowledge the cooperation and assistance from the chief and council of the community of Sandy Lake, the Sandy Lake community surveyors, the Sandy Lake nurses and the staff of the University of Toronto Sioux Lookout program, and the members of the Health Canada research team.

References

1. Bodzioch M, Orso E, Klucken J, Langmann T, Bottcher A, Diederich W, Drobnik W, Barlage S, Buchler C, Porsch-Ozcurumez M, Kaminski WE, Hahmann HW, Oette K, Rothe G, Aslanidis C, Lackner KJ, Schmitz G. The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat Genet.* 1999;22:347–351.
2. Brooks-Wilson A, Marcil M, Clee SM, Zhang LH, Roomp K, van Dam M, Yu L, Brewer C, Collins JA, Molhuizen HO, Loubser O, Ouelette BF, Fichter K, Ashbourne-Excoffon KJ, Sensen CW, Scherer S, Mott S, Denis M, Martindale D, Frohlich J, Morgan K, Koop B, Pimstone S, Kastelein JJ, Hayden MR. Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat Genet*. 1999;22:336–345.
3. Rust S, Rosier M, Funke H, Real J, Amoura Z, Piette JC, Deleuze JF, Brewer HB, Duverger N, Denefle P, Assmann G. Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nat Genet*. 1999;22:352–355.
4. Remaley AT, Rust S, Rosier M, Knapper C, Naudin L, Broccardo C, Peterson KM, Koch C, Arnould I, Prades C, Duverger N, Funke H, Assmann G, Dinger M, Dean M, Chimini G, Santamarina-Fojo S, Fredrickson DS, Denefle P, Brewer HB Jr. Human ATP-binding cassette transporter 1 (ABC1): genomic organization and identification of the

genetic defect in the original Tangier disease kindred. *Proc Natl Acad Sci U S A*. 1999;96:12685–12690.
5. Brousseau ME, Schaefer EJ, Dupuis J, Eustace B, Van Eerdewegh P, Goldkamp AL, Thurston LM, FitzGerald MG, Yasek-McKenna D, O'Neill G, Eberhart GP, Weiffenbach B, Ordovas JM, Freeman MW, Brown RH Jr, Gu JZ. Novel mutations in the gene encoding ATP-binding cassette 1 in four Tangier disease kindreds. *J Lipid Res.* 2000;41:433–441.
6. Lawn RM, Wade DP, Garvin MR, Wang X, Schwartz K, Porter JG, Seilhamer JJ, Vaughan AM, Oram JF. The Tangier disease gene product *ABC1* controls the cellular apolipoprotein-mediated lipid removal pathway. *J Clin Invest*. 1999;104:R25–R31.
7. Orso E, Broccardo C, Kaminski WE, Bottcher A, Liebisch G, Drobnik W, Gotz A, Chambenoit O, Diederich W, Langmann T, Spruss T, Luciani MF, Rothe G, Lackner KJ, Chimini G, Schmitz G. Transport of lipids from Golgi to plasma membrane is defective in Tangier disease patients and *Abc1*-deficient mice. *Nat Genet.* 2000;24:192–196.
8. Marcil M, Brooks-Wilson A, Clee SM, Roomp K, Zhang LH, Yu L, Collins JA, van Dam M, Molhuizen HO, Loubster O, Ouellette BF, Sensen CW, Fichter K, Mott S, Denis M, Boucher B, Pimstone S, Genest J Jr, Kastelein JJ, Hayden MR. Mutations in the ABC1 gene in familial HDL deficiency with defective cholesterol efflux. *Lancet.* 1999;354: 1341–1346.
9. Burnett JR, Law AJ, Yeong ML, Crooke MJ, Sharma AK. Severe aortic stenosis and atherosclerosis in a young man with Tangier disease. *Am J Cardiol*. 1994;73:923–925.
10. Hegele RA, Harris SB, Hanley AJ, Sun F, Connelly PW, Zinman B. Angiotensinogen gene variation associated with variation in blood pressure in aboriginal Canadians. *Hypertension*. 1997;29: 1073–1077.
11. Hegele RA, Young TK, Connelly PW. Are Canadian Inuit at increased genetic risk for coronary heart disease? *J Mol Med*. 1997;75:364–370.
12. *SAS/STAT Guide for Personal Computers*. Cary, NC: SAS Institute; 1987.

Table I. Intron-exon boundaries of the *ABC-1* gene

exon	5' boundary	3' boundary
2		TGAATgtaagtaactgtggatgttg
3	ctactcctcgatttcaacagGCCAT	TCCATgtaagtatcagatcaggttt
4	ttttcttgcacccttacagTGTGG	
5		ACAAGgtaagctgatgcctccagct
6	tttgcatgaaatgcttccagGTATT	TCCTGgtgagtagacttgctcactg
7	agttttctttctcattacagAGAAC	AGGAGgtaagttgtgtctttccagt
8	gtccccatccctgacgccagCTGTT	TACAAgtgagtgtccatgcagaccc
9	gtctctcctcttcttcatagCTCCT	CTGAGgtaagctgcccccagcccaa
10	ggctccccacctgactccagGTGAA	TCCGGgtgagtgtccctcccattat
11	ctttctcactggaatggcagATGCT	TGGAGgtgaatctgttgctgggatc
12	gaaggctgttcttctatcagTGTGT	GATGGgtaagtggaatcccatcaca
13	tacctgtggtttggttttagGTACT	GACATgtaagttacctgcaagccac
14	tccttccttggggcctccagCTTTC	TGAAGgtaaggcagcctcactcgct
15	atgtgtctcttccttcccagTTAGG	TCGCTgtgagtacctctggcctttc
16	actttgcctgtttctcacagAGCCT	TCCAGgtacactgctttgggcatct
17	tacctgtgctttgctttcagGCCAG	AGAAAgtaagtgctgttgacctcct
18	tgtcttctctcctgtcacagTCTGC	ACCATgtaagaagagggtgtggttc
19	ttctaaatttttgtcctcagGTCAA	GACATgtgagtaccagcagcacgtt
20	acctgcctccctgtccccagGCTGA	GTCAGgtgcggcccagagctacctt
21	cttgctctgtttttttgtagGTGGA	ACAAGgtgcctgatgtgtatttatt
22	cttgtgtttgtgcctggcagGCCGC	AAAAGgtgagctgcagtcttggtgc
23	acctccttctttgccatcagGAGGA	CATCGgtaaggactctggggtttct
24	ttgcattctggtccctgcagATGTC	AAGAAgtaagttaagtggctgactg
25	tcatctctctcatttttccagATATT	CTCAGgtaactgccttgagggagaa
26	ttctttccttgtgccttcagATGGT	CCCAGgtctgttagggcaagatcaa
27	tcagtgttggcttcttgcagAATCC	CTCAGgtgagacgtgctgttttcgc
28	cctccactgcctgggtaaagATTGT	GTCAGgtatgtttgtcttctacatc
29	tgcctctcatgtacccacagCAATG	ATCCCgtgagtgccactttagccat
30	ttgctgtcttccctttgcagAGACA	CACAAgtgagtcactttcagggggt
31	tggttttcatgtgatgacagAGAAA	AAAAGgtgacttttactaaacttg
32	ttgtttctttgtctctgcagCTTAA	TTTAGgtaagttgctgtctttctgg
33	atgtcaggctgttcttttagGTATG	CCAAGgtaaaatatctatcgtaaga
34	caggtttctgttgtttacagGACAG	TCAAGgtaaaccgctgtctttgttc
35	cctggtatgttttctttcagGTGTG	GCTCTgtaagtgtggctgtgtctgt
36	cacacgctttttctctccagGATGA	ATATGgtaaggacacaggcctgcgt
37	gtctttgttttcacttttagTGCAA	TATGGgtaagtcacctctgagtgag
38	atcttttgtgtctcaacagGTGGT	ACAATgtgagtcatgcagagagaac
39	ttttttttttttttttttagAAGCT	GTTTGgtgagtgaagcagtggctgt
40	ctctgcattgtgtctactagGGGAG	CCCAGgtgagctttttcttagaacc
41	ttcttgtttttttcccccaagACCTG	CGAAGgtgagagagtacaggttaca
42	ttacattttgtttgttgcagATATA	GTGAGgtaaagacactttgtctata
43	acttaatctttccttttcagTGCTT	AATAGgtgagaaaagaagtggcttg
44	agttttcttgtcctttgcagTATCT	GCAAGgtactgtgggcacctgaaag
45	ggtgctttttatcactgtagGTTGG	TTCTGgtgagtataactgtggatgg
46	ttattttattctaattacagGATGA	CATAGgtccgtagtaaagtcttggg
47	ctctttctgttttaccgcagTATGG	AATAGgtaataaagataatttcttt
48	gtatttttttcctgctttcagGTTTG	ACCAAgtaagctttgagtgtcaaaa
49	ccatcttaatcttttaacagGTATT	

Table II. *ABC-1* amplification primers for individual exons

exon	primer sequences (5' to 3')	product size
3	F: AAATGGAGCCTCAAAATCGCTTCA	302
	R: TCCAGCCATTCAAAATTCTCCAGA	
6	F: GACCCAGCTTCCAATCTTCATAA	288
	R: TTCCGAAAGCATTAGTGCTTGAAG	
7	F: GTTAGGAGTCGGTTTCTTGTTTG	292
	R: TCTTGCCAGACTCAGGAATGA	
8	F: AAGGCTTTGAGATGACATCAGACA	517
	R: GCAAATGGGCATGTAGACATCTAA	
9	F: CAGTATATGACCATGCATATGAGC	302
	R: AATTCCAGAAGAGCCGTGAGTATA	
10	F: TCAGGGAGGAGCCAAACGCTCATTGTCTGT	245
	R: CTCACCTCTCCAGTATCAAGCAGGCACATG	
11	F: TGCTCCTCAAGATTTAGTTGGTG	325
	R: GAACCTTATTGTAACGTCTCAGAG	
12	F: ATTTGTCTCAGCCACTGCTTGTT	333
	R: TAATATAATAGGTGCTCTGGACCT	
13	F: GAATGGTTGATTACCTGTGGTTTG	243
	R: AACTGGTTAAAAACAGTGGCTTGC	
14	F: ACCCTGGCTCACGTGTGTCT	342
	R: CCTTAGCCCGTGTTGAGCTA	
15	F: TCTGTAGATCATACAAGTGAGTGC	381
	R: ATCTTCTGTTCTCAACTTGCTGCT	
16	F: CTTCTGCACCTTATGATTGATCCA	386
	R: AGAAAGGACATCAGCTAGAAGTCA	
17	F: ATGATGCTGAGCTTGGCTCATAC	199
	R: GCAGCAGCACTAGGTTAAAGAAAG	
18	F: TTCTGTGGGTTCATTTCTGTCTTC	374
	R: ACGTGGAAAGACACTCATGATTAG	
19	F: TCAACCCTGACTTCTGTACTTCTA	212
	R: CATCCAGAAAAGGCCTATTCTTAAC	
20	F: TGGTTGCTCACTGTTGTTCTAC	450
	R: AGCCAGGGACAAGTTTCTGTTA	
21	F: TTGGTAACAGAAACTTGTCCCTG	279
	R: CTCTCTCAGTCCATTTACTCAGA	
22	F: CACTGTGACAAATCAGAACTGAGA	358
	R: AGAAATCATTCACAGCCAGCAAGT	
23	F: GTGCATGTGTCCCAGGGAAAG	324
	R: AGCAGCAAACCTTGAGTCAGC	
24	F: TGAGTTGATTGAGCCATCTTTGCA	341
	R: TGACTAACTCCATGATAATCCTGC	
25	F: CTACTCTCGTAATGTTGGTGC	174
	R: GAACAATACTCGTGCACTGAG	
26	F: GCATTCTCATTTGTAGAGTGTAGG	245
	R: ACAGTAGGACACTGTTTGATCTTG	
27	F: GAGAAACTTCTAGCTTCAGACACA	368
	R: ATCCATATCTTGACCAGGATGCTA	
28	F: TTCTAACACTTGCCGTTTCCTG	242
	R: TCGTAAACATCTTTGGTCTG	
29	F: TTAAGAGCCTGTCACAGAGAAATG	309
	R: TCACAATGTGGCATGCAGTTGATA	
30	F: CCATCTCTGGTAATCCTACTCTTG	316
	R: GATATCTCACTCATTCCTGCTTCC	

31	F: ACATGGTCACTGTTGTCCTTGA	298
	R: TAGACAGAATCAGGCCATAATCTG	
32	F: ATTAGGAATAGGCTCAGTGCT	167
	R: CAAGGCTTTCTTCAATCCAAG	
33	F: TGACTCAGTGAACCTTGTACACAC	263
	R: TACTCCTATATCCCAGCAGCTACA	
34	F: ACAGGTTCCAGATTGACTGTTG	216
	R: GAAAGTACTCCAGGAAACATAAGG	
35	F: GCCTTGGTATGTGACTAGAATTTC	458
	R: CCAGCCAATTCGTAAACTTATCAC	
36	F: CTCAGATATCTCTGTGTGTGTGTA	332
	R: TGACAGACATCAGAAAGATACAGCA	
37	F: GGTGAAAACTTGTGTGATGCATTG	232
	R: CAAATGCCTTATCCACTGTGCA	
38	F: CTACTTTTGCTGTATGGGTAAGTC	366
	R: CACTGACAAGTGGAATAAGATCAC	
39	F: CTGGTTTGTCCCTAGAGAATC	313
	R: TGCAGAGTGCCATCTCCATTA	
40	F: ACTTGGAAAGTGATCACTCTCTTG	515
	R: GACAAAGATTCAGTCTCTAACGTA	
41	F: GTTAGAGACTGAATCTTTGTCCTG	223
	R: CTGAAGATGAGCTATTGTAACCTG	
42	F: TGTGAGTCACTGAGGCTTATGAA	169
	R: GATTGGGTAGAGATAGTCTGAACT	
43	F: GAAGGTTTGAGGTAGTTACGTGTT	226
	R: GTCTTTGCAGCAAAATACAAGCCA	
44	F: AACCAGAGTATCTCTACCTGCTTA	315
	R: ATTTTGTGCTGCTGCATTCATGAG	
45	F: TACATGTATGTGTAGGACAGCATG	254
	R: TTTCCACTCAGGCCAGAACAA	
46	F: GCTGTTTCAAAGATGCTTCTGCA	237
	R: TAAAACATCCCACAGTGAGGAAC	
47	F: GTGATGTTCTCATGGTTACAGAGA	290
	R: CCCTCATTCTTTCCACTATACTG	
48	F: TAACTGGTTCCTTCTCTCAGAAAC	534
	R: GAATCCACACCCTGAGAAGTA	
49	F: TCTGTATGCTCCTACTTGACCTA	320
	R: CACTTACATGGTGCAAAGGAAAG	