

High-density genotyping and functional SNP localization in the CETP gene[§]

John F. Thompson,^{1,*} Linda S. Wood,^{*} Eve H. Pickering,[†] Bryan DeChairo,[§] and Craig L. Hyde[†]

Pharmacogenomics,^{*} Statistical Applications,[†] and Molecular Profiling,[§] Pfizer Global Research and Development, Groton, CT 06340

Abstract The cholesteryl ester transfer protein gene (*CETP*) has been the subject of hundreds of genetic analyses that typically focus on a small number of polymorphisms within a single ethnic group. Furthermore, the extent of DNA beyond the transcribed sequence from which single nucleotide polymorphisms (SNPs) may influence *CETP* expression has not been well defined. To better understand the role of natural variation in modulating *CETP* and high density lipoprotein-cholesterol (HDL-C) levels, dense genotyping of *CETP* and regions up to 15 kb on either side of the gene was carried out on >2,000 individuals. A complex, nonlinear set of linkage disequilibrium bins was found, with many bins interspersed along the DNA sequence and spread over large regions of the gene. Bins assigned based on large numbers of individuals matched the small subset of SNPs that had been assigned to bins previously with a small number of individuals. Associations of known functional SNPs with HDL-C were found, but there were suggestions that there are additional functional SNPs not characterized previously. Narrowing of the set of likely functional SNPs was accomplished by comparing associations observed in different ethnic groups. **■** The promoter SNP most highly associated with HDL-C that is likely to be functional, position -4,502, alters a consensus transcription factor binding site.—Thompson, J. F., L. S. Wood, E. H. Pickering, B. DeChairo, and C. L. Hyde. **High-density genotyping and functional SNP localization in the CETP gene.** *J. Lipid Res.* 2007. 48: 434–443.

Supplementary key words genetics • high density lipoprotein • single nucleotide polymorphism • cholesteryl ester transfer protein

The importance of the cholesteryl ester transfer protein gene (*CETP*) in affecting high density lipoprotein-cholesterol (HDL-C) levels in humans was originally detected when individuals lacking active protein were identified based on high HDL-C levels (1). Since then, individuals lacking *CETP* as well as those with varying levels of *CETP* or with variant sequences have been studied extensively (reviewed in Refs. 2, 3). In addition to the complete null mutations, many single nucleotide polymor-

phisms (SNPs) in *CETP* have been found to be reproducibly associated with protein mass/activity and/or HDL-C. When sample sizes are large enough, there is a high degree of consistency across studies and populations. Results with the closely linked phenotypes of *CETP* mass/activity and HDL-C are highly replicated, but associations with other, more complex phenotypes, such as cardiovascular disease, have been less easily replicated. However, when studies are of sufficient size and properly designed, associations with the more complex phenotypes can often be found. For example, a meta-analysis of the TaqIB SNP showed that the allele associated with low *CETP* was also associated with high HDL-C and lower levels of coronary artery disease (4). When studies examine only one or a small number of SNPs, integration of results with other studies can be challenging.

In addition to the numerous amino acid variants that have been detected in *CETP* (5), there is also evidence that promoter SNPs are even more significantly associated with HDL-C than those that change a single amino acid. Common promoter polymorphisms at positions -629 and -971 and a variable repeat sequence have all been reproducibly associated with *CETP* and/or HDL-C levels (6–10). These associations at the 5' end of the gene are stronger than those observed with SNPs causing amino acid changes at the 3' end of the gene. The functionality of the -629 SNP has been linked to changes in an Sp1/Sp3 binding site (7, 11), whereas the SNP at -971 was shown not to affect transcription (8). No results have been reported on the functional role of the variable repeat sequence. Large segments of the promoter have been fused to reporter genes to determine functional regions, but only changes at positions -629 and -38 have been examined as a function of naturally occurring polymorphisms (7, 11).

Linkage disequilibrium (LD) within *CETP* makes the identification of functional SNPs difficult. When only low-density genotypic information was available, genome

¹To whom correspondence should be addressed.
e-mail: john.f.thompson@pfizer.com

§The online version of this article (available at <http://www.jlr.org>) contains supplemental data in the form of one figure and one table.

structure was generally approximated by linear collections of haploblocks. Within the *CETP* gene, initial studies showed two haploblocks, one covering the promoter region and the 5' half of the gene and the second covering the 3' half of the gene and including many nonsynonymous SNPs (6, 10, 12). As higher density SNP information has become available, it became apparent that a linear collection of haploblocks is a poor approximation of genome structure. An improved description consists of multiple LD bins that are interspersed along the linear DNA sequence, with each bin containing a distinct subset of SNPs (13). These bins are described empirically and, for any given region of the genome, the extent and number of LD bins vary across different ethnicities. Typically, there are cutoffs for both minor allele frequency (MAF; 5%) and extent of LD ($R^2 > 0.8$) to define SNPs within each bin (13).

We have previously published association studies with 20 SNPs in and near the *CETP* gene with ~2,500 individuals (5, 10, 11, 14, 15). Although this has provided significant insight into the functional nature of *CETP* SNPs, there are still many unanswered questions about the role of different SNPs, how they are linked to each other, and the detailed genomic structure surrounding *CETP*. To help answer some of these questions, we have genotyped 63 additional SNPs in >2,000 individuals and integrated that data with other sources of information to generate a highly detailed map of *CETP* and its association with HDL-C.

MATERIALS AND METHODS

Samples and genotyping

The Atorvastatin Comparative Cholesterol Efficacy and Safety Study (ACCESS) (16) was designed to determine the safety and efficacy profile of atorvastatin compared with other HMG-CoA reductase inhibitors when used to treat patients with National Cholesterol Education Program LDL-C criteria. Whole blood from participating subjects was obtained with appropriate institutional review and appropriate informed consent documentation that defined the study design and provided an assessment of the risks and benefits associated with study participation. A second European cohort came from a previous Pfizer clinical trial in the cardiovascular area that recruited healthy patients (n = 664). DNA from another African-American cohort (n = 250) was purchased from Genomics Collaborative, Inc. (Cam-

bridge, MA). All laboratory tests were performed at a central laboratory (Medical Research Laboratories, Highland Heights, KY) certified by the National Heart, Lung, and Blood Institute/Centers for Disease Control Part III Program. HDL-C was measured in a fasting sample. No subfraction analysis was done.

Genomic DNA was extracted from whole blood using the PureGene DNA isolation system (Gentra) according to the manufacturer's protocol. Some SNPs discussed here were first reported elsewhere (5, 10, 11, 14, 15) and genotyped as described in those publications. SNPs reported for the first time here were genotyped using either TaqMan or SNPlex technology according to the manufacturer's instructions (Applied Biosystems, Foster City, CA).

Statistical analysis

The goal of the statistical analysis was to test for significant genetic associations between HDL-C levels and *CETP* SNPs across European (n = 3,129) and African (n = 420) subjects. A small number of Asian subjects (n = 36) were also used for comparison of fitted genotype effects, although this population was considered too small for hypothesis testing. The largest cohort studied was from ACCESS and included individuals with European (2,465), African-American (170), and Asian (36) ancestry, after removing subjects who were outliers (beyond 5 sigma) and/or who had missing data for critical phenotypes. Demographic information for the genotyped individuals is listed in **Table 1**.

It was determined via standard inspection of qq-plots that a log transformation was appropriate for the HDL-C response. Unfortunately, the variance in log(HDL-C) varied significantly by cohort as well as by gender, although to a lesser degree. However, within the ACCESS cohort, the African females exhibited much higher variance than either the African males or the Europeans of either gender. Hence, variance was allowed to vary by three factors: cohort, ethnicity, and gender.

The model used had log(HDL-C) as the response and genotype (coded as a three-level factor) as the main effect to be tested for. Explanatory covariates, all of which were significant against log(HDL-C), were as follows: age, gender, ethnicity, cohort, and alcohol consumption (coded as a three-level factor on weekly alcohol consumption: no drinks, 1 to <10 drinks, and 10 or more drinks, based on the distribution of values).

A generalized least-squares model was used, allowing for the heterogeneous variance components described above. Genotype significance for individual ethnicities was evaluated by likelihood comparisons of the full model with one with the genotypes of the targeted ethnicity assigned to a single factor, using a Chi-square test. The final model used for overall significance was

$$\log(\text{HDL} - \text{C}) \sim \text{age} + \text{gender} + \text{ethnicity} + \text{cohort} + \text{alcohol} + \text{genotype}$$

TABLE 1. Population demographics

Study	Ethnicity	Gender	Number	Age	HDL-C	Triglycerides	Body Mass Index
ACCESS	European	Male	1,498	60.8	44.0	184	28.5
	European	Female	967	62.9	54.0	198	28.4
	African American	Male	91	58.0	46.5	155	30.4
	African American	Female	79	59.6	57.4	158	31.7
Phase II	European	Male	412	49.3	42.0	192	28.4
	European	Female	253	54.4	53.6	173	27.8
GCI	African American	Male	120	58.5	47.7	158	30.1
	African American	Female	130	63.4	62.8	136	31.8

HDL-C, high density lipoprotein-cholesterol. Demographic and lipid values are provided for individuals who were genotyped from the three trials described in Materials and Methods. Values for age (years), HDL-C (mg/ml), triglycerides (mg/ml), and body mass index (m/kg²) are all means.

This model tested the hypothesis for overall genotype effect, and in addition, it was compared against an identical model with an ethnicity \times genotype interaction term added, and the significance of this interaction term was evaluated using a Chi-square test comparing the deviances of the two models.

The hypothesis tests were validated with a permutation test. Specifically, because many of the SNPs tested were in high LD with each other (and hence were far from independent), multiple-testing adjustment was performed by comparing the rank-ordered hypothesis test results against 5,000 permutations in which sample IDs were permuted within each ethnicity and then remerged to the full set of genotype values, thus preserving both the underlying LD structure and the explanatory covariates as they related to HDL-C while still simulating the null hypothesis for genotype effects. The adjusted P value for each SNP is the maximum (worst) of the false discovery rate so calculated and the point estimate of its P value from the individual permutation test.

In generating effects plots, the Asian population was merged with the African and European populations, then the entire data set was fit to a null model (i.e., one without a genotype term), and the residuals were plotted by ethnicity and genotype using box and whisker plots. Genotypes are denoted by A, B, or C, with A denoting the wild-type homozygotes (as defined by empirical observation of the pooled population), B denoting the heterozygotes, and C denoting the homozygotes in the minor allele, still based on empirical observation over all subjects. These definitions remained fixed across ethnicities, even if the MAF crossed the 50% barrier going from one ethnicity to another.

RESULTS

Genomic structure and LD blocks

To gain maximal information about the genetic structure surrounding *CETP*, databases and the literature were searched for all SNPs and other polymorphisms. The region we examined included 15 kb upstream of the *CETP* gene, the 22-kb gene, and 13 kb downstream of the gene. Segments of this region have been resequenced in 10 to 200 individuals (12, 13, 15, 17–21), with the greatest focus on the promoter and exons. Many laboratories carried out the sequencing in multiple ethnic groups. Within the heavily sequenced regions, all common SNPs have been identified, but some introns and regions outside of the gene have not been as well characterized for variation.

Only two sets of genotype data published to date span the entire 50 kb region of interest, HapMap (http://www.hapmap.org/cgi-perl/gbrowse/hapmap20_B35/) and those SNPs published by Hinds et al. (13) (<http://genome.perlegen.com/browser/index.html>). Within this 50 kb region, there are 103 HapMap SNPs with nonzero MAF, including a subset of 87 with a frequency of $>5\%$ in at least one ethnic group. Similarly, there are 33 SNPs reported by Hinds et al. (13) (all nonzero MAF) in this region. Thirteen of these 33 SNPs are not genotyped in the HapMap set. Using the individual genotype data from the data sets for which they are available, an initial set of LD bins was determined as described (13). The resulting bins were compared across data sets to the extent possible to determine which bins could be combined. Because all of these data sets have a limited number of individuals and none includes complete sequence information over the

entire region, many of the LD bins are poorly defined. To better understand the nature of these bins and to collapse them into a smaller set, SNPs were chosen from across the region for analysis in a large, multi-ethnic population. Genotype data from the original databases and the literature were generated using a variety of techniques. Some SNPs are not amenable to genotyping by particular technologies, and not all could be assayed by the SNPlex technology used here, preventing us from obtaining a complete set of genotypes. To the extent possible, some SNPs unique to each data set were included so that comparison across studies could be accomplished.

All SNPs that we genotyped were in Hardy-Weinberg equilibrium with $P > 0.05$, with the exception of one SNP each in individuals of European and African ancestry. Both of these SNPs were $P > 0.03$ and thus within the expected range of normal variation for a study with this number of SNPs tested. Of the 103 published HapMap SNPs, we generated more in-depth data for 40 of them, including 38 with MAF $> 5\%$ in at least one population. Of the 33 Hinds et al. (13) SNPs, we generated data for 29 of them. By generating genotypes for thousands of individuals that includes SNPs from each of these sets of individuals, we are able to establish LD bins representing SNPs spanning all of these data sets. A summary of the allele frequency for each ethnicity for polymorphisms with MAF $> 5\%$ for which we have generated data (reported previously or here for the first time) is shown in **Fig. 1**. For completeness, uncommon SNPs and an additional 101 SNPs for which others have published individual genotype data are also included in supplementary Table I. Some SNPs have multiple dbSNP identifiers, and we have used the one chosen by National Center for Biotechnology Information and listed alternative numbers. Even though the group size for some populations is relatively small (20 to 50 individuals), the minor allele frequencies are consistent across studies within the same ethnic group (see supplementary Table I).

An LD chart with individual R^2 values for all of the SNPs we genotyped in 2,458 individuals with European ancestry with MAF $> 5\%$ is shown in supplementary Fig. I. This includes 56 SNPs across 50 kb. Using Haploview, seven LD blocks are identified that span 32 kb and include 50 SNPs (**Fig. 2**). Over the same genomic region, 59 SNPs genotyped in HapMap have MAF $> 5\%$ in the 90 person CEU population. Thirty-one of these SNPs are identical to those we genotyped. With the HapMap SNPs, six LD blocks are identified with Haploview that cover 25 kb and include 43 SNPs. Although the blocks in the CEU and ACCESS populations align well for the most part, there are differences in both the number of blocks and their boundaries, primarily in the 5' region of the gene. Because of the much larger ACCESS population, many more SNPs are incorporated into the LD blocks, including three that were genotyped by HapMap but not placed in LD blocks. Thus, the large number of individuals genotyped allows many more SNPs to be placed in LD blocks, but the relevance of these blocks to the detailed genomic structure is also much more apparent, as perhaps best visualized by the "check-

dbSNP	Trials	Euro Bin	Asian Bin	Af Am Bin	pt Euro	adj Euro	pt Asian	adj Asian	pt African	adj Afr	Exon/ Intron	Position	Build 35	Euro N=2415	Asian N=36	Af Am N=168
rs173537	All	1	1	1								-14957	55538409	0.25	0.21	0.62
rs247611	AC/Pf	1	1	2								-14625	55538741	0.25	0.22	0.34
rs247614	AC/Pf	1	1	3								-11878	55541488	0.26	0.21	0.62
rs9989419	All	4	4	4	<.00001	<.001						-10726	55542640	0.61	0.73	0.41
rs7203286	All	5	5	5	<.00001	<.001						-9103	55544263	0.45	0.49	0.55
rs7205692	All	6	6	7	<.05	<.001						-8951	55544415	0.16	0.15	0.34
rs173539	AC/Pf	8	8	9	<.00001	<.001	<.05	<.05				-7821	55545545	0.31	0.32	0.33
rs247618	AC/Pf	10	10	10		<.05						-5062	55548304	0.76	0.82	0.86
rs183130	All	8	8	8	<.00001	<.001	<.05	<.05	<.01	<.05		-4502	55548864	0.31	0.29	0.25
rs6499861	All	6	6	11	<.01	<.001			<.05	<.05		-4370	55548996	0.16	0.17	0.21
rs6499862	All	6	6	6	<.01	<.05						-4341	55549025	0.16	0.17	0.28
rs6499863	AC/Pf	6	6	6	<.001	<.001						-3848	55549518	0.17	0.18	0.26
rs12920974	AC/Pf	12	12	12	<.0001	<.001						-2840	55550526	0.28	0.31	0.20
rs12149545	AC/Pf	8	8	13	<.00001	<.001	<.05					-2704	55550662	0.30	0.28	0.07
rs3764261	AC/Pf	8	8	8	<.00001	<.001	<.05	<.05				-2541	55550825	0.31	0.29	0.28
VNTR	AC	8	8	8	<.00001	<.001	<.01	<.05		<.05		-1990	55551376	0.31	0.26	0.26
rs17368435	All	8	8	14	<.00001	<.001	<.05	<.05				-1337	55552029	0.31	0.29	0.15
rs12708968	AC/Pf	15	15	15								-1046	55552320	0.10	0.19	0.06
rs4783961	AC/Pf	16	16	16	<.0001	<.001						-971	55552395	0.49	0.40	0.43
rs4783962	AC/Pf	10	10	10								-827	55552539	0.76	0.79	0.86
rs1800775	AC/GC	17	17	17	<.00001	<.001						-629	55552737	0.48	0.52	0.57
rs708272	All	18	18	19	<.00001	<.001			<.05	<.05	1/2		55553789	0.42	0.47	0.26
rs4369653	All	20	20	20							2/3		55555052	0.31	0.25	0.24
rs11508026	All	18	18	21	<.00001	<.001					2/3		55556829	0.42	0.43	0.11
rs820299	All	22	20	22							2/3		55557785	0.61	0.71	0.53
rs12720922	All	23	23	23	<.00001	<.001					2/3		55558386	0.18	0.25	0.34
rs12597002	All	20	20	24							2/3		55559905	0.30	0.22	0.16
rs9926440	All	25	25	25	<.00001	<.001					2/3		55560164	0.71	0.65	0.35
rs9939224	All	23	23	23	<.00001	<.001					2/3		55560233	0.80	0.74	0.65
rs11076174	All	26	26	26	<.0001	<.001	<.05		<.05		2/3		55560647	0.08	0.18	0.08
rs12720861	All			64							5/6		55561960	0.0021	0	0.12
rs7205804	All	18	18	27	<.00001	<.001	<.05	<.05			5/6		55562390	0.43	0.44	0.15
rs1532624	AC	18	18	27	<.00001	<.001		<.05			7/8		55562980	0.43	0.43	0.15
rs11076175	AC	23		28	<.00001	<.001					8/9		55563879	0.18	0.26	0.19
rs5883	All	29		29	<.05	<.05			<.05	<.05	9	F270F	55564854	0.06	0.03	0.11
rs11076176	All	30	30	30	<.00001	<.001					9/10		55564947	0.17	0.26	0.42
rs158477	All	31		31						<.05	9/10		55565111	0.46	0.54	0.54
rs158480	All	32		32	<.05	<.001					9/10		55565728	0.12	0.27	0.37
rs289715	All	32	32	33	<.05	<.001					9/10		55566009	0.89	0.79	0.81
rs289716	AC	34	34	34	<.0001	<.001					10/11		55566877	0.31	0.40	0.51
rs289719	All	34	34	35	<.0001	<.001					10/11		55567442	0.69	0.64	0.58
rs4784744	All	36	36	36	<.05	<.05					10/11		55568686	0.35	0.34	0.15
rs291043	All	36	80	36	<.05	<.001					11/12		55570200	0.34	0.32	0.14
rs5880	All	37		37	<.00001	<.001	<.05	<.05			12	A373P	55572592	0.05	0.03	0.01
12 bp insert	All	32	32	38	<.01	<.001	<.05	<.05			12/13		55572838	0.12	0.21	0.24
rs1800774	All	39	39	40		<.01					12/13		55573046	0.36	0.31	0.26
rs5882	All	34	34	41	<.00001	<.001					14	I405V	55573593	0.68	0.62	0.41
rs289741	All	34	41	41	<.0001	<.001	<.05	<.05	<.05	<.05	15/16		55574975	0.69	0.67	0.40
rs1801706	All	42		42	<.05	<.001					16		55575163	0.18	0.13	0.18
rs289742	AC	32		44	<.01	<.05						9	55575263	0.88	0.77	0.65
rs289746	All	45	45	45								2452	55577706	0.27	0.16	0.25
rs17369163	All	46		46							2574	55577828	0.12	0.03	0.02	
rs12934552	All	47		47							3680	55578934	0.13	0	0.09	
rs289747	All	48	48	48	<.05	<.001					6185	55581439	0.39	0.22	0.39	
rs17290922	All	49		49		<.05					6564	55581818	0.11	0	0.07	
rs1566439	All	50		50		<.05					6909	55582163	0.39	0.60	0.26	
rs7197864	All	49		49		<.05					7131	55582385	0.11	0	0.07	
rs289748	All	51	48	51	<.0001	<.001			<.05	<.05	7310	55582564	0.52	0.22	0.61	
rs7205459	All	49		49	<.05	<.05					7418	55582672	0.11	0	0.07	
rs289749	All	52	45	52							8541	55583795	0.30	0.17	0.26	
rs17369468	All	53	53	53							10012	55585266	0.29	0.39		
rs289735	All	52	52	52							11809	55587063	0.30	0.24	0.25	

Fig. 1. Cholesteryl ester transfer protein (CETP) single nucleotide polymorphisms (SNPs). The dbSNP nomenclature for each SNP examined by us with minor allele frequency (MAF) > 5% in at least one population is listed in the first column. A more complete list of SNPs and additional data for all SNPs are provided in supplementary Table I. Trials [AC = ACCESS (for Atorvastatin Comparative Cholesterol Efficacy and Safety Study); PF = Pfizer Phase II cardiovascular trial; GC = Genomics Collaborative Incorporated (GCI) African-American cohort] in which individuals were genotyped are listed in the second column. Bin numbers for each ethnic group are listed in columns 3–5, with shaded blocks representing tagging SNPs identified using HapMap samples for individuals of European (column 3) and African (column 5) ancestry. *P* values for association with high density lipoprotein-cholesterol (HDL-C) generated using a generalized least-squares method and adjusting for covariates but not for multiple testing are provided in columns 6, 8, and 10, with the smallest *P* values listed as <.00001 even though some are <10⁻¹⁰. False-discovery rates generated from a permutation analysis that better corrects for multiple testing, linkage disequilibrium (LD) structure, and violation of modeling assumptions are provided in columns 7, 9, and 11. Because only 5,000 permutations were done, the smallest *P* values are listed as <.001, but these could be substantially smaller if more permutations were attempted. Of note, the permutation test also revealed the conservative nature of the underlying model, as many uncorrected point estimates for *P* values were more significant in the permutation test than in the original model (particularly among Caucasians); this is the reason why many of the adjusted false-discovery rates are actually lower than the unadjusted *P* values. In columns 6–11, <x means the *P* value is >x/10 but not >x, unless x is one of the minimum listed values cited above, or unless x = 0.05, in which case the value is >.01 but not >.05. All SNPs were tested for association. In column 12, the exon/intron positions are provided. In column 13, the position relative to the start and end of transcription and SNPs that are located within the coding sequence is listed. The nucleotide position on chromosome 16 in Build 35 is listed in column 14. Columns 15–17 provide allele frequencies in the ACCESS trial.

dbSNP	CETP Position	ACCESS	HapMap
rs173537	-14957	1	1
rs247611	-14625	1	1
rs247614	-11878	1	
rs247615	-11101		0
rs9989419	-10726	0	0
rs12448528	-10309		0
rs7203286	-9103	0	
rs7205692	-8951	2	
rs173539	-7821	2	2
rs247616	-6275	2	2
rs183130	-4502	2	
rs6499861	-4370	2	
rs6499862	-4341	2	
rs6499863	-3848	2	2
rs12920974	-2840	2	
rs12149545	-2704	2	
rs12708967	-2653		2
rs3764261	-2541	2	2
rs12447839	-1929		2
rs12447924	-1672		2
rs17368435	-1337	2	2
rs12708968	-1046	2	
rs4783961	-971	2	2
rs4783962	-827	2	
rs1800775	-629	2	0
rs711752	347		3
rs708272	424	2	
rs1864163	1369		3
rs4369653	1687	3	
rs9929488	2708		4
rs7203984	3394		4
rs11508026	3464	3	4
rs820299	4420	3	4
rs17237974	5021	3	
rs12597002	6540	3	4
rs9926440	6799	4	4
rs9939224	6868	4	4
rs11076174	7282	4	
rs7205804	9025	4	4
rs1532624	9615	4	4
rs12708974	9686		4
rs11076175	10514		4
rs7499892	10726		4
rs9930761	11328		0
rs5883	11489	5	0
rs11076176	11582	5	
rs289714	11587		0
rs158477	11746	5	
rs17238100	12363	6	
rs289715	12644	6	6
rs289716	13512	6	
rs289717	13524		6
rs289719	14077	6	6
rs2033254	14121		6
rs4784744	15321	6	6
rs12720898	15379		6
rs12708980	16515		6
rs17231848	16835	6	
rs4784745	19011		6
rs5880	19227		6
rs1800774	19681	6	6
rs5882	20228	6	6
rs9923854	21138		6
rs289741	21610	6	6
rs1801706	21798	6	6
rs289742	+9		6
rs289744	+349		0
rs12720917	+1639		0
rs289745	+1779		0
rs289746	+2452	0	
rs17369163	+2574	7	
rs12934552	+3680	7	0
rs289747	+6185	7	7
rs17290922	+6564	7	
rs1566439	+6909	7	7
rs7197864	+7131	7	
rs289748	+7310	7	7
rs7205459	+7418	7	
rs289749	+8541	0	
rs289750	+9011		0
rs17369468	+10012	0	0
rs8058353	+10762		0
rs289734	+11602		0
rs289735	+11809	0	0

Fig. 2. The dbSNP identifiers (column 1) and positions relative to the CETP gene (column 2) are shown for all SNPs with MAF > 5% that we genotyped in ACCESS or that are in HapMap for individuals of European ancestry. LD blocks as determined by Haploview are listed as 1 through 7, skipping 5 in HapMap, and are shown for the SNPs genotyped by us (column 3) or by HapMap (column 4). If 0 is listed, the SNP is not present in a block. If no number is listed, the SNP was not genotyped. Contiguous blocks are shaded light gray, and discordant areas are shaded dark gray.

erboard” pattern of LD in block 7, 3 to 8 kb downstream of the CETP gene. In addition, most of the weak LD interactions that appear in the HapMap population disappear with the much larger ACCESS population.

LD bins

Initial determination of LD bins was carried out to select SNPs for genotyping, but these bins were redefined after the complete set of genotypes was obtained. Numbering of LD bins is arbitrary. We have used the same numbers across ethnicities where possible, but it is clear that the boundaries for these bins are not the same across ethnic groups. As noted previously by Hinds et al. (13), the LD bins are highly noncolinear, with significant interdigitation of SNPs in different bins.

When we define LD bins using our data, the SNPs grouped together are very consistent within an ethnic group compared with those generated by Hinds et al. (13) or using the same definitions with HapMap data, even though both sets had far fewer individuals. The only discrepancies found with either data set are with SNPs that are very close to either the MAF or R^2 cutoff in one population or the other. Otherwise, there is perfect agreement for LD bin composition.

Using the cutoffs of MAF > 5% and $R^2 > 0.8$ for LD bin determination and all available data, 102 SNPs can be placed in 49 bins for individuals of European ancestry. The most populated bin contains eight SNPs, and the largest span covered by a single bin is 10,731 bp. For HapMap samples, 40 tagging SNPs representing 59 SNPs are identified. Nearly identical bins are generated with only two HapMap tagging SNPs falling into the same bin as defined here. When the additional 43 SNPs not found in HapMap are added, this required only an additional nine bins, confirming that there is a point of diminishing returns for genotyping but that genotyping at a frequency of greater than one SNP per kilobase still has the potential to generate useful data.

For individuals of African ancestry, there are 97 SNPs in 66 bins. The most populated bin contains seven SNPs, and the largest spans 6,621 bp. The larger number of bins and shorter extent of DNA sequence covered by each bin in those of African-American versus European ancestry are similar to what has been observed previously (13). Our bins match well with the African Americans characterized by Hinds et al. (13) but poorly with the Nigerian HapMap data. This highlights the difficulties of comparing across populations and ensuring that ancestry is matched appropriately. We have not attempted a comparison of our Asian ancestry bins with others because of the small number of individuals genotyped.

Associations with HDL-C

In addition to characterizing the genomic structure surrounding *CETP*, we also wanted to determine how the association with HDL-C was superimposed on that structure. Among individuals of European ancestry, the strongest associations are clearly in the promoter but span a very

broad region. We have extended coverage to within 3 kb of the neighboring upstream gene, HERPUD1, 15 kb from the CETP transcriptional start. Our finding that a SNP >10 kb away (rs9989419) from the start site is associated with HDL-C suggests that distal interactions may play a role in regulating CETP levels. However, much of this association appears to arise from LD with nearby SNPs. All of the SNPs most highly associated with HDL-C among individuals of European ancestry are in LD bin 8. LD bins 6 and 10 are interspersed in this part of the promoter region but are orders of magnitude less significantly associated with HDL-C based on point estimates (Fig. 1). Several singleton SNPs in the promoter are also associated with HDL-C, but not as strongly as those in bin 8. One of the most highly associated SNPs, rs183130 in bin 8, has a consistent effect across ethnic groups, as shown in Fig. 3. The sequence conservation for this region among primates and the sequence surrounding two other functional SNPs is shown in Fig. 4. In each region, the chimpanzee sequence is identical to the human sequence, whereas other species have up to several changes. For each of the putative transcription factor binding sites, some of these nonconserved positions would be predicted to affect protein binding.

At the other end of the gene, the associations with HDL-C are not as strong as observed in the promoter region. Several nonsynonymous SNPs in this 3' region have been shown to be functional, with effects on CETP activity or stability. There may be functional SNPs in addition to the nonsynonymous SNPs. rs289748, which is >7,000 bp from

the end of transcription, is associated with HDL-C in individuals of both European ($P < 0.001$) and African ($P < 0.02$) ancestry. This SNP is not tightly linked with other SNPs tested in either group. The functional source of this association is unknown.

All associations were tested for gender effects as well. Two SNPs in individuals of European ancestry and one SNP in individuals of African-American ancestry yielded gender-genotype interactions with P values between 0.03 and 0.05, not significant after correction for multiple testing. Interactions between SNPs were also tested. Because of the number of SNPs and comparisons involved, this was done only for individuals of European ancestry in the ACCESS trial. Four pairs of SNPs show strong, nonadditive interactions, even after correction for multiple testing. The most significant interaction, between rs12920974 and rs4783961, has an uncorrected P value of 1.7×10^{-8} , remaining significant even after correction for 2,616 tests. Both of these SNPs are in the promoter region. The HDL-C means for each combination of genotypes are shown in Fig. 5. The other SNP pairs that remain significant after multiple testing correction all include rs7203286 (in the distal promoter region) in combination with rs820299 (intron 2), rs158477 (intron 9), or rs4783961 (promoter).

DISCUSSION

Low HDL-C is known to be a major risk factor for cardiovascular disease (reviewed in Ref. 22). HDL-C is af-

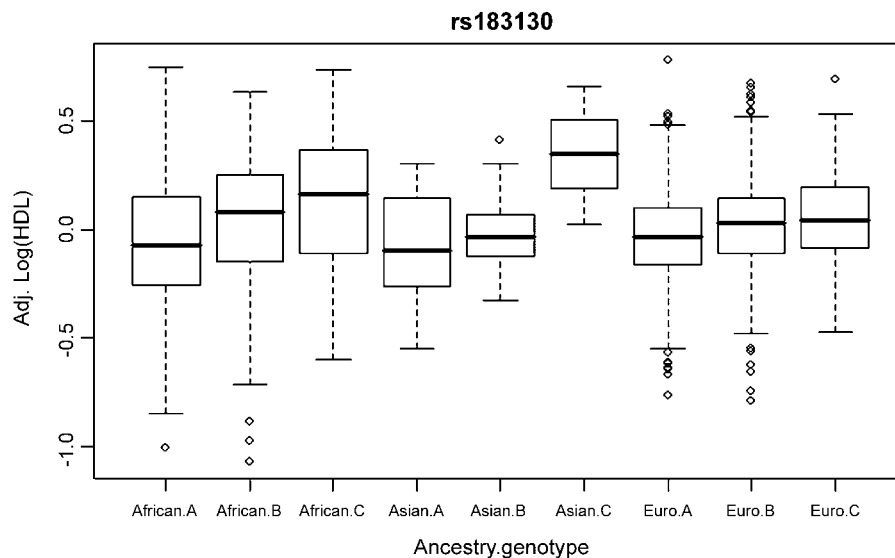


Fig. 3. The plot depicts the ordered genotype of rs183130 for each ancestry on the horizontal axis [labeled as ancestry genotype, where the genotype is named A, B, or C for homozygotes in the most common allele (among pooled ancestries), heterozygotes, or homozygotes in the less common allele, respectively]. The vertical axis represents the values of log(HDL) after adjusting for nongenotypic covariates in the model. The boxes in the plot are bounded above and below by the 75th and 25th percentiles, respectively (the quartiles), and the error bars/whiskers extend an additional 1.5 times the interquartile range from either boundary of the box. Points outside this extension are plotted as outliers. The horizontal line inside the box denotes the median. The deviation from mean HDL-C for each of the rs183130 genotypes is shown for the three ethnicities examined.

		GGRRAHYCCC
-4502	Human	AGACAAATCTTCTGTGCAGGAGAAATCCAAATGAATTCTGTAGATTCTCCA
	Chimpanzee	AGACAAATCTTCTGTGCAGGAGAAATCCAAATGAATTCTGTAGATTCTCCA
	Baboon	AGACAAATCTTCCATGCGGAAGAACCACAAATGAATTGTGTAGATTCTCCA
	Marmoset	AGACAGATCGTCCATGCAGAGAATCCCAATGAATTCTGTGAATTCTCTA
	Rhesus	AGACAAATCTTCCATGCAGAGAATCCCAATGAATTCTGTAGATTCTCCA
		***** ** *
		YCCGCCCM
-629	Human	CAATGATCTCAGAGGCTGTATACCCGCCAGAGTTATTTTATGCATATCAA
	Chimpanzee	CAATGATCTCAGAGGCTGTATACCCGCCAGAGTTATTTTATGCATATCAA
	Baboon	CAATGATCTCAGAGGCTGTATACCCGCCAGAGTTATTTTATGCATATCAA
	Marmoset	CAATGATCTCAGGGGCTGTATACCCGCCAGAGTTATTTTATGCATATCAA
	Dusky Titi	CAATGATCTCAGAGGCTGTATACCCGCCAGAGTTATTTTATGCATATCAA
	Rhesus	CAATGATCTCAGAGGCTGGGTACCCATCCAGAGTTATTTTATGCATATCAA
		***** ** *
		KGGCGGGR
-38	Human	GCCTCATGTTCCGTGGGGGCTGGGCGGACATACATATACGGGCTCCAGGCT
	Chimpanzee	GCCTCATGTTCCGTGGGGGCTGGGCGGACATACATATACGGGCTCCAGGCT
	Baboon	GCCTCATGTTCCATGGGGGCTGGGAGGACATATATACAGGCTCCAGGCT
	Marmoset	GCCTCATGTTCCGTGGGGGCTGGGCGGACATACATATACAGGCTCCAGGCT
	Dusky Titi	GCCGATGTTCCGTGGGGGCTGGGCGGACACTTATACGGGCTCCAGGCT
	Rhesus	GCCTCATGTTCCATGGGGGCTGGGAGGACATATATACAGGCTCCAGGCT
		*** ** *

Fig. 4. The sequence for 25 bp on either side of the three SNPs in transcription factor consensus sites is shown for humans and up to five other primates. Dusky titi is not shown for rs183130 because the sequence conservation was too low in that region. Each SNP is shown in reverse color, and putative transcription factor binding sites are shown above each sequence. The chimpanzee sequence was obtained from the University of California, Santa Cruz website (<http://www.genome.ucsc.edu/cgi-bin/hgGateway?org=Chimp&db=panTro2>), and all other sequences were obtained from the Lawrence Berkeley Laboratory website (http://pga.lbl.gov/cgi-bin/get_gene?id=131).

ected by a variety of environmental factors such as alcohol intake, estrogen administration (23), and exercise (24) as well as by a host of genetic factors. The effect of *CETP* variation on HDL-C is robust (2) but can be obscured when small samples or particular SNPs are examined. For example, of the SNPs we examined, there are five within the *CETP* gene and one in the proximal promoter with minor allele frequencies of >10% that are not associated with HDL-C ($P > 0.05$) among the >2,400 individuals of European ancestry. If one were to look exclusively at these SNPs, one would mistakenly conclude that *CETP* is not associated with HDL-C levels.

All common *CETP* variants have, at most, modest effects on either *CETP* mass or activity. If the most significantly associated SNP, rs183130, is examined, the mean HDL-C for the common versus rare homozygote varies only from 46.3 to 49.7 mg/dl among individuals of European an-

cestry, a difference of <10%. In contrast, both gender and alcohol consumption have larger effects on HDL-C, with European females in the ACCESS trial having significantly higher HDL-C (54 mg/dl) than European males (44 mg/dl). Similarly, European males who consume >10 drinks per week have higher HDL-C (50.7 mg/dl) than those who consume none (41.9 mg/dl). Only rare, non-functional *CETP* variants have a large impact on HDL-C, and these also have a protective effect with respect to disease in large, prospective studies (25). Thus, the impact of common *CETP* SNPs can be readily observed on a population basis, but these are of little value when examining small numbers of individuals.

It is possible to overinterpret the effect of *CETP* on HDL-C and attribute any large change in HDL-C to *CETP* variants of modest functional significance, even when known environmental effects such as exercise are present

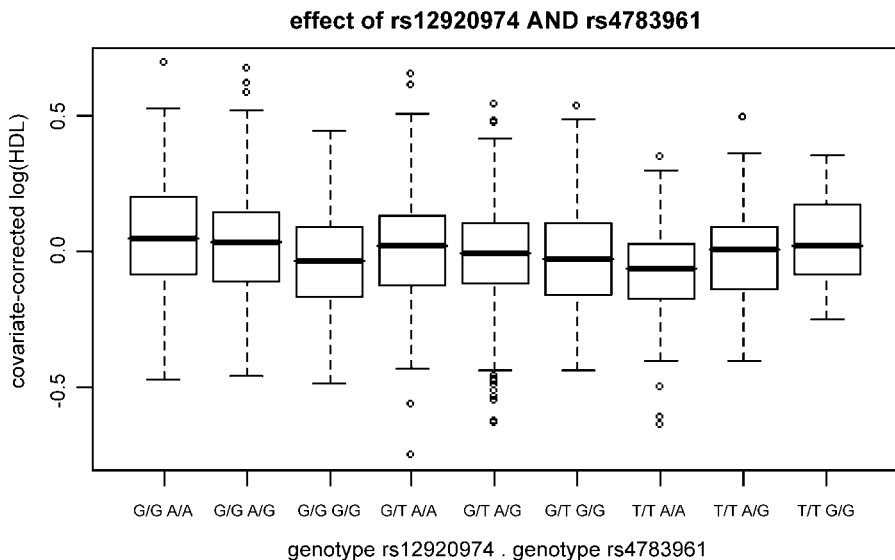


Fig. 5. Mean HDL-C values for each combination of genotypes for rs12920974 and rs4783961. Definitions for the box and whiskers are as described for Fig. 3.

(26). Even the high HDL-C induced by extreme exercise, such as running marathons (27), is not always protective for cardiovascular disease. Risk of sudden cardiac death is generally attributable to a variety of defects in cardiac structural and channel proteins, with >50% of such deaths attributed to hypertrophic cardiomyopathy (28), independent of *CETP* genotype.

When data for only a limited number of SNPs were available, it was most convenient to describe the genomic structure of *CETP* and other genes as a series of large haploblocks. Initial work was consistent in showing that a large number of SNPs in the promoter and 5' region of the gene were in LD with each other, whereas another set of SNPs in the 3' region constituted another haploblock (6, 10, 12). This simplified view of the gene was useful as a rough approximation but is not accurate when comparing SNPs supposedly in the same haploblock but really having little linkage. With much more data now available, it is clear that the more detailed approach of using LD bins or tagging SNPs from nonlinear parts of the genome is necessary for an accurate view. Although there are many approaches and definitions that can be used for defining LD bins and tagging SNPs, similar results are obtained across populations with the same ancestry. The bins defined by Hinds et al. (13) with only 24 individuals were nearly identical to those defined by us with >2,000 individuals.

Horne et al. (21) identified a set of tagging SNPs for *CETP* that overlap the SNPs examined here. Despite using a very different approach, many SNPs represented by their tagging SNPs fall into separate bins, as defined here. The overlap is not perfect, but the same overall picture of the genomic structure is generated. However, comparison of our data with this set of tagging SNPs also highlights the need to characterize extensive regions of DNA sequence for a given gene. Their most distal SNP examined was only 631 bp from the start of transcription; thus, the LD bins with the most significant associations with HDL-C were not tagged.

The promoter SNP at -629 has been shown to affect Sp1 binding in vitro, reporter activity in cells, and *CETP* mass levels in humans (7, 10, 11). In contrast, the TaqIB SNP (rs708272) is among the most extensively studied, but there has been no indication that it exerts any functional effect. Although TaqIB is not in the same LD bin as the -629 SNP, it is in reasonably high LD ($R^2 = 0.72$) in individuals of European ancestry. Furthermore, TaqIB is in the same LD bin as other SNPs that span a large region of the gene, including SNPs in intron 2 ($R^2 = 0.85$), intron 5 ($R^2 = 0.81$), and intron 7 ($R^2 = 0.91$) in individuals of both European and Asian ancestry. Any of these SNPs could potentially have some functional effect, or it could arise from some other uncharacterized SNP in this 9-kb region. The TaqIB SNP is also associated with HDL-C in African Americans, but it is a singleton in terms of LD bins. Unlike individuals of European and Asian ancestry, in whom the most strongly linked SNPs are 3' to TaqIB, the SNPs most strongly linked to TaqIB among individuals of African ancestry are in the promoter with rs183130

($R^2 = 0.68$) and the variable number of tandem repeats ($R^2 = 0.64$) being in tightest LD. This suggests that the functional SNP(s) assessed when TaqIB was examined in Africans is different from the functional SNP(s) assessed in Asians and Europeans.

Even though the promoter -629 SNP has been shown to have functional effects, it is clear that other SNPs in the promoter are also independently associated with HDL-C. The most highly associated SNPs are those in LD bin 8. However, among individuals of European ancestry, there are seven SNPs spread over 6,500 bp in bin 8, six of which we genotyped. Each could be examined on its own for functionality, but that is a challenging and not always fruitful endeavor. The availability of results from multiple ethnic groups makes it possible to decrease the number of potentially functional SNPs by taking advantage of the different LD structures. All six of the bin 8 SNPs we examined in Europeans are highly associated with HDL-C. Among Asians, all five of the SNPs in the homologous bin are also associated with HDL-C. Among Africans, the seven SNPs from bin 8 in Europeans are split into four separate bins, and only one of them is associated with HDL-C, rs183130 at position -4502. Although it is possible that there are distinct functional SNPs in the different ethnic groups, the consistent results (Fig. 3) with this SNP suggest that rs183130 may be functional. Additional experiments will be necessary to confirm this.

When 11 promoter SNPs that are associated with HDL-C among individuals of European ancestry are scanned for transcription factor binding sites (29), the only one that results in a change is rs183130. When a G is present on the bottom strand (GGGATTCTCC), an 8:10 match to the consensus site for nuclear factor κ B (GGGGYNNCCY) described by Ghosh, May, and Kopp (30) and a 10:10 match to the consensus site (GGGRDITYCC) described by Liu et al. (31) are found. The alteration from G to A creates a mismatch in both consensus sequences. The Liu et al. (31) consensus is particularly interesting in that it also appears to bind members of the Sp1 family of proteins that have been shown to be important in regulating *CETP* at the proximal promoter SNPs at -629 and -38 (7, 11).

Large meta-analyses of some *CETP* SNPs have been published, and these show a consistent but variable association with HDL-C, *CETP* mass/activity, and other phenotypes (2, 3). Results in studies with large numbers of individuals with European ancestry are in agreement with our results (supplementary Table I). Promoter SNPs rs12149545, rs4783961, and rs1800775 are significantly associated with HDL-C (6, 8, 32, 33). Similarly, results with large numbers of individuals of Asian ancestry found rs3764261 and the VNTR highly associated with HDL-C (9), as we found. All of these studies have focused on the promoter sequence within 3,300 bp of the transcriptional start site. Most functional characterization of the promoter has been restricted to the proximal 3,000 bp (34), with only limited analysis beyond that region (35).

In addition to the univariate analyses, nonadditive interactions between SNPs may also be important, as seen with other lipid-related genes (36). Our observation that the

association of some promoter SNPs (rs4783961 at -971) is nonadditive with other SNPs mirrors previous findings (33) in which that SNP's function in vitro was dependent on other nearby SNPs. The fragment tested in vitro was only 1,707 bp long, so it did not include the SNP we found most significant, rs12920974 at -2,940, but the concordance of results clearly shows the complexity of genetic modulation of transcription.

The data provided here yield a means of comparing results across many studies by determining which SNPs are likely to yield similar information and which will not. The extensive genotyping also allows other investigators to compare their populations with an independent set to test whether differences in allele frequencies found in a case-control study might be attributable to problems with the control population rather than a true association. As stated above, the SNPs tested here are all in Hardy-Weinberg equilibrium, unlike some control populations described elsewhere. By examining the *CETP* gene in detail and across populations, we are able to predict which SNPs are likely to be functional. This approach is generalizable to other genes in which robust associations are found. ■■

REFERENCES

1. Brown, M. L., A. Inazu, C. B. Hesler, L. B. Agellon, C. Mann, M. E. Whitlock, Y. L. Marcel, R. W. Milne, J. Koizumi, H. Mabuchi, et al. 1989. Molecular basis of lipid transfer deficiency in a family with increased high-density lipoproteins. *Nature*. **342**: 448-451.
2. Boekholdt, S. M., and J. F. Thompson. 2003. Natural genetic variation as a tool in understanding the role of CETP in lipid levels and disease. *J. Lipid Res.* **44**: 1080-1093.
3. Boekholdt, S. M., J.-A. Kuivenhoven, G. K. Hovingh, J. W. Jukema, J. J. P. Kastelein, and A. van Tol. 2004. CETP gene variation: relation to lipid parameters and cardiovascular risk. *Curr. Opin. Lipidol.* **15**: 393-398.
4. Boekholdt, S. M., F. M. Sacks, J. W. Jukema, J. Shepherd, D. J. Freeman, A. D. McMahon, F. Cambien, V. Nicaud, G. J. de Grooth, P. J. Talmud, et al. 2005. Cholesteryl ester transfer protein TaqIB variant, high-density lipoprotein cholesterol levels, cardiovascular risk, and efficacy of pravastatin treatment—individual patient meta-analysis of 13,677 subjects. *Circulation*. **111**: 278-287.
5. Lloyd, D. B., M. E. Lira, L. S. Wood, L. K. Durham, T. B. Freeman, G. Preston, X. Qiu, E. Sugarman, P. Bonnette, A. Lanzetti, et al. 2005. Cholesteryl ester transfer protein variants have differential stability but uniform inhibition by torcetrapib. *J. Biol. Chem.* **280**: 14918-14922.
6. Corbex, M., O. Poirier, F. Fumeron, D. Betouille, A. Evans, J. B. Ruidavets, D. Arveiler, G. Luc, L. Tiret, and F. Cambien. 2000. Extensive association analysis between the CETP gene and coronary heart disease phenotypes reveals several putative functional polymorphisms and gene-environment interaction. *Genet. Epidemiol.* **19**: 64-80.
7. Datchet, C., O. Poirier, F. Cambien, M. J. Chapman, and M. Rouis. 2000. New functional promoter polymorphism, CETP/-629, in cholesteryl ester transfer protein (CETP) gene related to CETP mass and high density lipoprotein cholesterol levels: role of Sp1/Sp3 in transcriptional regulation. *Arterioscler. Thromb. Vasc. Biol.* **20**: 507-515.
8. Le Goff, W., M. Guerin, V. Nicaud, C. Datchet, G. Luc, D. Arveiler, J.-B. Ruidavets, A. Evans, F. Kee, C. Morrison, et al. 2002. A novel cholesteryl ester transfer protein promoter polymorphism (-971G/A) associated with plasma high-density lipoprotein cholesterol levels: interaction with the TaqIB and -629C/A polymorphisms. *Atherosclerosis*. **161**: 269-279.
9. Lu, H., A. Inazu, Y. Moriyama, T. Higashikata, M. Kawashiri, W. Yu,

- Z. Huang, T. Okamuri, and H. Mabuchi. 2003. Haplotype analyses of cholesteryl ester transfer protein gene promoter: a clue to an unsolved mystery of TaqIB polymorphism. *J. Mol. Med.* **81**: 246-255.
10. Thompson, J. F., L. K. Durham, M. E. Lira, C. Shear, and P. M. Milos. 2005. CETP polymorphisms associated with HDL cholesterol may differ from those associated with cardiovascular disease. *Atherosclerosis*. **181**: 45-53.
11. Thompson, J. F., D. B. Lloyd, M. E. Lira, and P. M. Milos. 2004. CETP promoter SNPs in Sp1 binding sites affect transcription and are associated with HDL cholesterol. *Clin. Genet.* **66**: 223-228.
12. Thompson, J. F., M. E. Lira, L. K. Durham, R. W. Clark, M. J. Bamberger, and P. M. Milos. 2003. Polymorphisms in the CETP gene and association with CETP mass and HDL levels. *Atherosclerosis*. **167**: 195-204.
13. Hinds, D. A., L. L. Stuve, G. B. Nilsen, E. Halperin, E. Eskin, D. G. Ballinger, K. A. Frazer, and D. R. Cox. 2005. Whole-genome patterns of common DNA variation in three human populations. *Science*. **307**: 1072-1079.
14. Lira, M. E., D. B. Lloyd, S. Hallowell, P. M. Milos, and J. F. Thompson. 2004. Highly polymorphic repeat region in the CETP promoter induces unusual DNA structure. *Biochim. Biophys. Acta*. **1684**: 38-45.
15. Lloyd, D. B., J. M. Reynolds, M. T. Cronan, S. P. Williams, M. E. Lira, L. S. Wood, D. R. Knight, and J. F. Thompson. 2005. Novel variants in human and monkey CETP. *Biochim. Biophys. Acta*. **1737**: 69-75.
16. Ballantyne, C. M., T. C. Andrews, J. A. Hsia, J. H. Kramer, and C. Shear. 2001. Correlation of non-high-density lipoprotein cholesterol with apolipoprotein B: effect of 5 hydroxymethylglutaryl coenzyme A reductase inhibitors on non-high-density lipoprotein cholesterol levels. *Am. J. Cardiol.* **88**: 265-269.
17. Cargill, M., D. Altshuler, J. Ireland, P. Sklar, K. Ardlie, N. Patil, N. Shaw, C. R. Lane, E. P. Lim, N. Kalyanaraman, et al. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **22**: 231-238.
18. Morabia, A., E. Cayanis, M. C. Costanza, B. M. Ross, M. S. Flaherty, G. B. Alvin, K. Das, and T. C. Gilliam. 2003. Association of extreme blood lipid profile phenotypic variation with 11 reverse cholesterol transport genes and 10 non-genetic cardiovascular disease risk factors. *Hum. Mol. Genet.* **12**: 2733-2743.
19. Chasman, D. I., D. Posada, L. Subrahmanyam, N. R. Cook, V. P. Stanton, and P. M. Ridker. 2004. Pharmacogenetic study of statin therapy and cholesterol reduction. *J. Am. Med. Assoc.* **291**: 2821-2827.
20. Stansson, P., M. Zabeau, G. Meersseman, G. Remes, Y. Gansemans, N. Storm, R. Hartmer, C. Honisch, C. P. Rodi, S. Bocker, et al. 2004. High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Res.* **14**: 126-133.
21. Horne, B. D., J. F. Carlquist, L. A. Cannon-Albright, J. B. Muhlestein, J. T. McKinney, M. J. Kolek, J. L. Clarke, J. L. Anderson, and N. J. Camp. 2006. High-resolution characterization of linkage disequilibrium structure and selection of tagging single nucleotide polymorphisms: application to the cholesteryl ester transfer protein gene. *Ann. Hum. Genet.* **70**: 524-534.
22. Linsel-Nitschke, P., and A. R. Tall. 2005. HDL as a target in the treatment of atherosclerotic cardiovascular disease. *Nat. Rev. Drug Discov.* **4**: 193-205.
23. Lamou-Fava, S. 2002. High-density lipoproteins: effects of alcohol, estrogen, and phytoestrogens. *Nutr. Rev.* **60**: 1-7.
24. Kelley, G. A., K. S. Kelley, and Z. V. Tran. 2004. Aerobic exercise and lipids and lipoproteins in women: a meta-analysis of randomized controlled trials. *J. Womens Health.* **13**: 1148-1164.
25. Curb, J. D., R. D. Abbott, B. L. Rodriguez, K. Masaki, R. Chen, D. S. Sharp, and A. R. Tall. 2004. A prospective study of HDL-C and cholesteryl ester transfer protein gene mutations and the risk of coronary heart disease in the elderly. *J. Lipid Res.* **45**: 948-953.
26. Sirtori, C. R., L. Calabresi, D. Baldassarre, G. Franceschini, A. B. Cefalu, and M. Averna. 2006. CETP levels rather than polymorphisms as markers of coronary risk: healthy athlete with high HDL-C and coronary disease—effectiveness of probucol. *Atherosclerosis*. **186**: 225-227.
27. Hartung, G. H., J. P. Foreyt, R. E. Mitchell, I. Vlasek, and A. M. Gotto, Jr. 1980. Relation of diet to high-density-lipoprotein cholesterol in middle-aged marathon runners, joggers, and inactive men. *N. Engl. J. Med.* **302**: 357-361.
28. Firoozi, S., S. Sharma, and W. J. McKenna. 2003. Risk of competitive sport in young athletes with heart disease. *Heart*. **89**: 710-714.

29. Prestridge, D. S., and G. Stormo. 1993. Signal Scan 3.0. New database and program features. *Comput. Appl. Biosci.* **9**: 113–115.
30. Ghosh, S., M. J. May, and E. B. Kopp. 1998. NF- κ B and Rel proteins: evolutionarily conserved mediators of immune responses. *Annu. Rev. Immunol.* **16**: 225–260.
31. Liu, A., P. W. Hoffman, W. Lu, and G. Bai. 2004. NF- κ B site interacts with Sp factors and up-regulates the NR1 promoter during neuronal differentiation. *J. Biol. Chem.* **279**: 17449–17458.
32. Klerkx, A. H. E. M., M. W. T. Tanck, J. J. P. Kastelein, H. O. F. Molhuizen, J. W. Jukema, A. H. Zwinderman, and J. A. Kuivenhoven. 2003. Haplotype analysis of the CETP gene: not Taq1B, but the closely linked $-629C \rightarrow A$ polymorphism and a novel promoter variant are independently associated with CETP concentration. *Hum. Mol. Genet.* **12**: 111–123.
33. Frisdal, E., A. H. E. M. Klerkx, W. Le Goff, M. W. T. Tanck, J.-P. Lagarde, J. W. Jukema, J. J. P. Kastelein, M. J. Chapman, and M. Guerin. 2005. Functional interaction between $-629C/A$, $-971G/A$, and $-1337C/T$ polymorphisms in the CETP gene is a major determinant of promoter activity and plasma CETP concentration in the REGRESS study. *Hum. Mol. Genet.* **14**: 2607–2618.
34. LeGoff, W., M. Guerin, L. Petit, M. J. Chapman, and J. Thillet. 2003. Regulation of human CETP expression: role of Sp1 and Sp3 transcription factors at promoter sites -690 , -629 , and -37 . *J. Lipid Res.* **44**: 1322–1331.
35. Williams, S., L. Hayes, L. Elsenboss, A. Williams, C. Andre, R. Abramson, J. F. Thompson, and P. M. Milos. 1997. Sequencing of the cholesteryl ester transfer protein 5' regulatory region using artificial transposons. *Gene.* **197**: 101–107.
36. Hamon, S. C., J. H. Stengard, A. G. Clark, V. Salomea, E. Boerwinkle, and C. F. Sing. 2004. Evidence for non-additive influence of single nucleotide polymorphisms within the apolipoprotein E gene. *Ann. Hum. Genet.* **68**: 521–535.