

Common Variants in 6 Lipid-Related Genes Discovered by High-Resolution DNA Melting Analysis and Their Association with Plasma Lipids

John F. Carlquist^{1,2*}, Jason T. McKinney³, Benjamin D. Horne^{1,2}, Nicola J. Camp², Lisa Cannon-Albright², Joseph B. Muhlestein^{1,2}, Paul Hopkins², Jessica L. Clarke¹, Chrissa P. Mower¹, James J. Park¹, Zachary P. Nicholas¹, John A. Huntinghouse¹ and Jeffrey L. Anderson^{1,2}

¹Cardiovascular Department, Intermountain Healthcare, USA

²Department of Internal Medicine, University of Utah School of Medicine, USA

³Idaho Technology, Inc., Salt Lake City, UT, USA

Abstract

Background: Total cholesterol was among the earliest identified risk factors for coronary heart disease (CHD). We sought to identify genetic variants in six genes associated with lipid metabolism and estimate their respective contribution to risk for CHD.

Methods: For 6 lipid-associated genes (*LCAT*, *CETP*, *LIPC*, *LPL*, *SCARB1*, and *ApoF*) we scanned exons, 5' and 3' untranslated regions, and donor and acceptor splice sites for variants using Hi-Res Melting® curve analysis (HRMCA) with confirmation by cycle sequencing. Healthy subjects were used for SNP discovery (n=64), haplotype determination/tagging SNP discovery (n=339), and lipid association testing (n=786).

Results: In 17,840 bases of interrogated sequence, 90 variant SNPs were identified; 19 (21.1%) previously unreported. Thirty-four variants (37.8%) were exonic (16 non-synonymous), 28 (31.1%) in intron-exon boundaries, and 28 (31.1%) in the 5' and 3' untranslated regions. Compared to cycle sequencing, HRMCA had sensitivity of 99.4% and specificity of 97.7%. Tagging SNPs (n=38) explained >90% of the variation in the 6 genes and identified linkage disequilibrium (LD) groups. Significant beneficial lipid profiles were observed for *CETP* LD group 2, *LIPC* LD groups 1 and 7, and *SCARB1* LD groups 1, 3 and 4. Risk profiles worsened for *CETP* LD group 3, *LPL* LD group 4.

Conclusions: These findings demonstrate the feasibility, sensitivity, and specificity of HRMCA for SNP discovery. Variants identified in these genes may be used to predict lipid-associated risk and reclassification of clinical CHD risk.

Keywords: Lipids; Genetic variants; Coronary heart disease; High resolution melting curve analysis

Introduction

Coronary heart disease (CHD) is the leading cause of morbidity/mortality in the Western world [1,2]. Among the traditional risk factors for CHD, total cholesterol is one of the earliest and strongest risk factors for the development of the disease [3]. Recently, specific guidelines have addressed the risk associated with individual plasma lipid components, e.g. low density lipoprotein (LDL), high-density lipoprotein (HDL), and triglycerides (<http://www.nhlbi.nih.gov/guidelines/cholesterol/atp3xsum.pdf>). The Intermountain Heart Collaborative Study is a database and biospecimen registry, which has prospectively enrolled over 17,000 consenting subjects undergoing coronary angiography since 1994 [4]. We have reported that lipid-related risk for CHD in subjects enrolled in the registry is explained almost entirely by high LDL and/or low HDL [5]. Given the prevalence of these lipid abnormalities in the registry and the current interest in the human exome as related to lipids [6], we sought to determine whether genetic variants in the protein coding and regulatory regions of some principal genes associated with lipid metabolism contribute to differences in lipid levels and to risk for CHD among enrolled subjects.

The selection of genes examined in this study was based on known lipid metabolic pathways [7] and included those that encode principal enzymes, transfer proteins and cell receptors that mediate the formation and degradation of LDL and HDL. The chosen genes were: *LCAT* (encodes the lecithin-cholesterol acyl transferase that converts cholesterol to cholesterol esters and facilitates the maturation of the HDL particle), *CETP* (the product of which is cholesteryl ester transfer

protein which redistributes triglycerides and cholesteryl esters between lipoproteins), *LIPC* (the gene for hepatic lipase), *LPL* (encodes lipoprotein lipase which transfers phospholipids and free cholesterol to HDL), *SCARB1* (the gene that encodes the hepatic lipid receptor), and *ApoF* (the product of which is the lipid transfer inhibitor protein which inhibits the action of the *CETP* protein). Three of these (*CETP*, *LPL*, *LIPC*) have recently been reported in genome-wide association studies (GWAS) to be implicated as risk factors for CHD [8]. Polymorphisms or over-expression of the remaining three genes have been found to influence lipids and/or CHD risk in other studies [9,10].

The current generations of SNP arrays used for GWAS have average marker densities of 4k-5k bp. Thus, following the independent validation of marker(s) identified by a GWAS or by whole-genome sequencing, it remains necessary to examine the identified region(s) to identify the specific variant(s) associated with the disease. High

***Corresponding author:** Dr. John F. Carlquist, Intermountain Healthcare Cardiovascular Research Laboratory, LDS Hospital, 8th Avenue & C Street, Salt Lake City, UT 84132, USA, Tel. 1-801-4081028; Fax 1-801-408-5820; E-mail: john.carlquist@imail.org

Received May 02, 2011; Accepted July 06, 2011; Published July 10, 2011

Citation: Carlquist JF, McKinney JT, Horne BD, Camp NJ, Cannon-Albright L, et al. (2011) Common Variants in 6 Lipid-Related Genes Discovered by High-Resolution DNA Melting Analysis and Their Association with Plasma Lipids. J Clin Exp Cardiol 2:138. doi:10.4172/2155-9880.1000138

Copyright: © 2011 Carlquist JF, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

resolution melting curve analysis (HRMCA) is a rapid and inexpensive method to scan a gene for sequence variants. It is based on the principle that heteroduplex double-stranded DNA molecules are formed when a genomic region carrying a variant is amplified by polymerase chain reaction. The heteroduplex molecule melts at a temperature that is relatively less than that of a double-stranded molecule formed when no variants are present. We have previously used melting curve analysis to identify rare/novel variants in the bone morphogenic protein receptor-2 (BMP2) gene [11] and selected this method for identifying variants in the above genes. Additionally, we assessed the degree of linkage disequilibrium (LD) among the variants within each gene, identified LD groups, and assigned a tagging SNP(s) for each LD group. This report presents the results of the SNP discovery phase by HRMCA and the description of LD among SNPs in each gene. For each LD group, the tagging SNP(s) approximates the variation within that LD group. We applied this approach to identify variants in the above genes and test them primarily for association with lipid measures in healthy volunteers. A follow-up, companion study will use this approach to seek associations among lipid gene LD groups with the clinical endpoint of CHD.

Materials and Methods

We sought to validate HRMCA as a sensitive and specific approach to identify sequence variants in the coding, splicing, and regulatory regions of selected genes. We selected genes (described above) specifically involved in lipid metabolic pathways. Secondly, we attempted to associate variation in the selected genes with variation in plasma lipids. To this end, a probabilistic adjustment of the size of the SNP discovery cohort was used to allow identification of those variants occurring at a sufficiently large frequency to potentially explain variations in lipid measures on a population basis. Finally, we tested for the presence of such associations.

Subjects for SNP discovery

Participants were healthy volunteers attending a community health fair. Enrollment was approved by the IRB, and written informed consent was obtained prior to enrollment. Peripheral blood was collected in ethylenediaminetetraacetic acid (EDTA) by venipuncture. A SNP discovery cohort of 50 subjects representative of the local population, principally of Euro-American heritage, was scanned for variants. Additionally, 14 subjects qualifying as NIH-defined underrepresented minorities were similarly scanned. The determination of LD groups was accomplished in a haplotype discovery set of 339 subjects of Euro-American descent. The method of principal components analysis (PCA) [12] was used for haplotype identification, which is particularly useful for characterizing intragenic LD structure because it accounts for mutations, which may contribute importantly to genetic variation at high resolution, in addition to the large (intergenic) changes due to recombination.

Subjects for assessment of lipid/genetic associations

A probabilistic adjustment of the size of the SNP discovery cohort was used to allow identification of those variants occurring at a sufficiently large frequency to potentially explain variations in lipid measures on a population basis. A population-based, randomly generated control sample was assembled by invitation (letters and follow-up telephone calls) to a random sample of the general population from the greater Salt Lake City metropolitan area based on the Utah Population Database (UPDB). Invitees agreeing to participate completed a health-related questionnaire, had vital signs

taken, and donated a blood sample for study-related testing. The study was approved by the LDS Hospital IRB and the UPDB. Those without clinically evident CHD were assigned to the population-based control set (n= 786).

DNA extraction

DNA was extracted from blood samples using a Gentra Autopure LS automated DNA extractor. Intactness of DNA is routinely gauged by gel electrophoresis. The quantity and purity of eluted DNA was determined by UV absorption at 260 and 280 nm; DNA was adjusted to 200 µg/mL and stored at -70°C.

Polymerase chain reaction and melting curve analysis

The overall objective was to scan all exons to capture coding variants and 20 bp 5' and 50 bp 3' to capture donor and acceptor splice variants. Target regions for scanning were amplified by the polymerase chain reaction (PCR) in the presence of the double strand DNA-binding dye LCGreen® Plus+, and presumptive variants were detected by Hi-Res Melting®. Targets for determination of melting curves were prepared using either a PTC 200®, DYAD® thermal cycler (both from MJ Research) or an ABI 9700® thermal cycler (Applied Biosystems). Amplification reactions were individually optimized. PCR products were transferred to a LightScanner® (Idaho Technology, Inc.) and subjected to a slow (0.1° C/sec) thermal denaturation (melt) during which fluorescence of each sample is continuously monitored. Melting data are analyzed by the LightScanner software, which automatically groups sample profiles according to melting curve shape. Samples containing heteroduplex molecules melt at relatively lower temperatures and display premature decreases in fluorescence. Fluorescence and temperature normalized melting profiles are displayed as subtractive difference plots where heteroduplex variant profiles are easily distinguished from the most common, presumably wild type, melt profile group (Figure 1). Suspected variants were confirmed by fluorescence cycle sequencing at the University of Utah Core Sequencing Facility using Big Dye® terminator chemistry v3.1.

SNPs identified at an allele frequency ≥ 0.05 within the discovery cohort were carried forward for haplotype determination. For high throughput genotyping of discovered SNPs, two methods were used: 1) 5'-3' nuclease (Taqman®) allelic discrimination assays and 2) Hi-

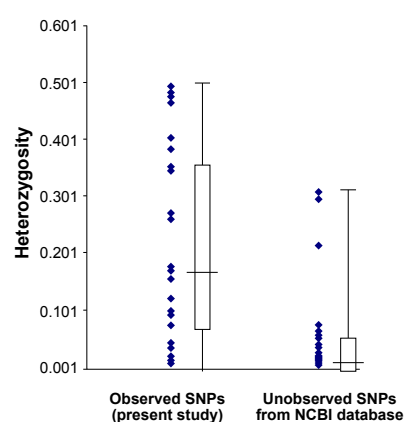


Figure 1. Distribution of SNP heterozygosity (ratio of the number of observed heterozygous genotypes/total genotypes) for the SNPs discovered in the present study versus SNPs listed in NCBI dbSNP. Shown are range (whiskers) median (horizontal line) and inter-quartile range (box).

GENE	Exons	5' UTR	3'UTR	# Bases	# Assays
<i>CETP</i>	16	700	2000	5505	42
<i>SCARB1</i>	12	1055	200	3762	17
<i>LCAT</i>	6	100	200	2042	18
<i>APOF</i>	2	38	665	1815	8
<i>LIPC</i>	9	1288	200	3625	18
<i>LPL</i>	4(10)*	461	200	1091	12

*Exons 3-9 sequenced by Nickerson et al. [15]

Table 1: SNP Discovery Summary.

GENE	5'UTR	3'UTR	Intron	Synon	Non-syn	Novel/ Total
<i>CETP</i>	2	7	18	4	3	3/34
<i>SCARB1</i>		3	3	4	5	5/15
<i>LCAT</i>			1	1	1	0/3
<i>APOF</i>			3	3	2	5/8
<i>LIPC</i>	11	1		6	4	1/22
<i>LPL</i>	3	1	3		1	3/8
Total	16	12	28	17	18	17/90

Table 2: Number and Location of Discovered SNP's.

Res Melting[®] on the LightScanner using LunaProbes[™]. For Taqman[®] assays, primers and probes were designed using the ABI Prism[®] Primer Express[™] and purchased from Applied Biosystems. The assays were performed on an ABI 7000 Sequence Detection System. LunaProbes[™] are unlabeled oligonucleotides with a 3' block to prevent extension during PCR. Genotype is determined by monitoring the temperature at which the probe: target hybrid melts (Figure 2). A single LunaProbe can be used to genotype multiple alleles and/or combinations of alleles [13].

LD and tagging SNP determinations

PCA is an established method for determining the contributions of an individual variable to an observed factor. PCA calculates a factor loading for each variable, or component, within each factor. When squared, this factor loading represents a multivariate r^2 . For this study, LD groups were determined from a PCA including all SNPs genotyped in the haplotype discovery set (n=339) that had a minor allele frequency >0.01. The optimal number of tSNPs for an LD-group was the number of factors that explain >90% of the group's variance [12].

Tagging SNP association with lipid measurements

To evaluate the effect of genetic variants on lipid values, tagging SNPs were tested for associations with standard lipid measurements in a cohort of healthy volunteers without a history of CHD or lipid-lowering medication use (n=786). Genetic associations with HDL were tested using the linear test of trend in analysis of variance.

Results

Scanning exons and adjacent splice regions

The details of scanning the coding and the 5' and 3' intronic sequences (to capture donor and acceptor splice variants, respectively) are summarized in Table 1. In all, 18,029 bases of sequence were interrogated, albeit with some redundancy (approximately 1-2% due to overlapping primer sets required to scan larger exons); 115 target sequences were scanned for 64 subjects or for a total of 7360 total scans. Scanning with sequence verification identified 90 SNPs for the six genes under study. Of the 90 observed SNPs, 19 (21.1%) were previously unreported variants. Thirty four (37.8%) of the variants were identified in exons; 18 (53%) exonic variants were not predicted to result in

amino acid substitutions (synonymous substitutions), and 16 (47.0%) were predicted to result in the alteration of the protein sequence (non-synonymous substitutions). Twenty eight (31.1%) of the identified SNPs were in intron-exon boundaries containing splice sequences, and 28 (31.1%) were in the 5' and/or 3' untranslated regions. The summary of discovered SNPs is presented in Table 2.

Interestingly, for the six genes studied, the National Center for Biotechnology Information (NCBI) Single Nucleotide Polymorphism database lists 81 SNPs in the 49 exons (including 20 bp 5' and 50 bp 3') that were interrogated. Of those 81, we found 30 (37.0%) in the discovery set of 50 subjects of Northern European descent. Three SNPs were found exclusively in the minority sample (n=14).

The SNPs discovered by HRMCA were typically more common polymorphisms than those not identified as judged by their relative heterozygosity (the ratio of the number of heterozygous genotypes to the total number of genotypes) as found in the NCBI database. Heterozygosity estimates were available for 63 of the 81 NCBI SNPs; the median heterozygosity for the SNPs discovered in the present study (discovered SNPs with calculated heterozygosity: n=27) was 0.16 (range: 0.003 to 0.492) versus 0.02 (range: 0.002 to 0.309) for those 56 SNPs not identified in our sample set (n=36 with calculated heterozygosity). This suggests that the unidentified SNPs were generally of a low frequency and likely not represented in our SNP discovery sample. In support of the notion that HRMCA is a reasonably sensitive technique to discover rare/novel SNPs, 19 of 90 variants (21.1%) observed in this study were previously unreported or novel. The distributions of the heterozygosity values for SNPs found in this study versus those that were not identified are given in Figure 1.

Comparison of melting-curve analysis to sequencing

During the initial SNP discovery phase, 169 of the scanned targets (n=7360, 2.3% of all scanned segments) were found to produce a melting profile indicative of a polymorphism. All presumptive gene variations observed by scanning were resequenced for SNP validation. Of the 169 presumptive positives, 159 were verified by resequencing. Additionally, as a measure of sensitivity, 420 targets (5.7% of total scanned) without evidence of a SNP by melting curve analysis were also resequenced. Of the 420 presumptive negatives, 419 were determined to be true negatives. It is of interest to note that the false negative sample harbored a variant that was correctly identified in multiple other samples on the same run, introducing the possibility of a sample mix up or pipetting error. As determined by this study, melting curve analysis had an overall sensitivity of 99.4% and a specificity of 97.7%.

LD and tagging SNP determination

The most highly variant gene was *CETP*, with 34 total SNPs/5505 bases), and the least variation was found in *LCAT* (3 SNPs/2042 bases). The highest proportion of novel SNP discoveries were for *SCARB1*, 5/15 (33%). All SNPs with a minor allele frequency >0.01 in the haplotyping cohort (n=339) were assigned to an LD group (or discontinuous haplotype) using PCA. Sixty seven of the 90 discovered SNPs (74.4%) had a minor allele frequency >0.01 and were included in the LD analysis. The remaining 23 SNPs were of very low frequency and were not considered to be reliably assigned to one or more groups. Overall, 29 LD groups were determined across the 6 genes (Table 3).

The number of LD groups/gene ranged from 2 to 9, and the number of SNPs present within the individual LD groups varied from 1 to 9. Ten of the 29 LD groups (34.5%) were represented by a single SNP.

ApoF:	LD group 1: ss142463324*
	LD group 2: rs34934555*
LIPC:	LD group 1: rs1077835, rs1077834, rs1800588*, rs2070895
	LD group 2: rs6082*, rs3829462, rs3829461*
	LD group 3: rs6083, rs6084*, rs6074*
	LD group 4: rs6078*
	LD group 5: rs36041167*
	LD group 6: rs35588604*
	LD group 7: rs6082*, ss142463322*
	LD group 8: rs690*
	LD group 9: rs35511894*
LPL:	LD group 1: rs248*, rs316*, rs4922115
	LD group 2: ss142463328*, rs1800590*, rs1801177
	LD group 3: ss142463330, ss142463332*
	LD group 4: rs11570897*, rs11570891*
SCARB1:	LD group 1: rs5889*, rs5888*
	LD group 2: rs61932577*, rs5888*
	LD group 3: rs5891*, rs5888*
	LD group 4: rs4238001*
	There were 5 SNPs in SCARB1 with low MAF that were excluded as not helpful: ss142463311, ss142463314, ss142463316, ss142463318, ss142463320
LCAT:	LD group 1: rs4986970*
	LD group 2: ss142463326*
	LD group 3: rs5923*
CETP:	LD group 1: rs289715*, rs289718, rs289719*, rs1800774*, rs5882, rs289741, rs289742, rs289743, rs289744
	LD group 2: rs1800775, rs708272*, rs1532625, rs11076175*, rs158477, rs291044*, rs1800774*
	LD group 3: rs1800776*, rs11076175*, rs11076176, rs289714, rs289715, rs158477, rs291044*, rs1800774*
	LD group 4: rs5880*, rs1800777
	LD group 5: rs158477, rs289715*, rs12720889, rs1801706*, rs289742
	LD group 6: rs1800776*, rs289745
	LD group 7: rs5883*, rs12720917

*tagging SNP

Table 3: Linkage groups and tagging SNPs for the variants discovered by scanning in the 6 lipid-related genes.

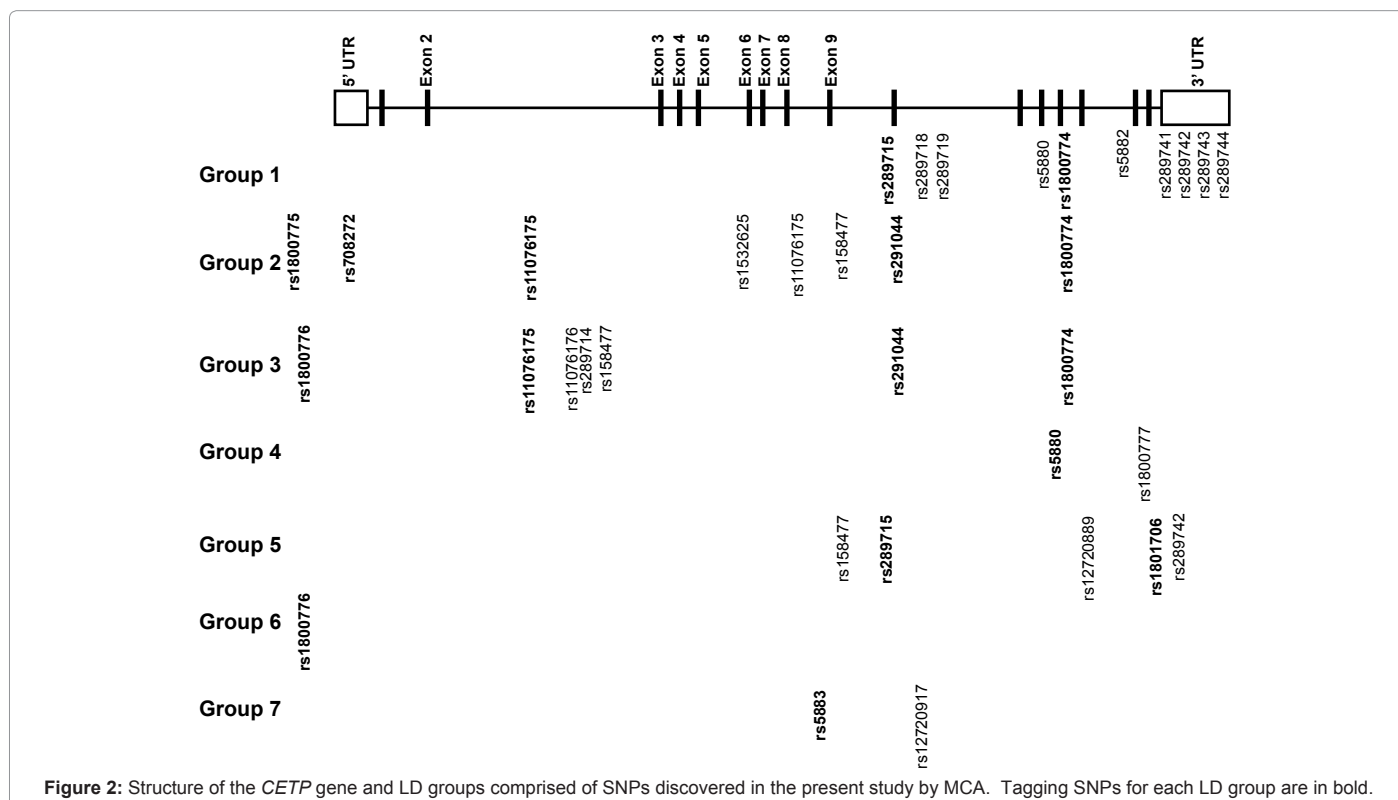


Figure 2: Structure of the CETP gene and LD groups comprised of SNPs discovered in the present study by MCA. Tagging SNPs for each LD group are in bold.

The largest LD groups were identified for the *CETP* gene; for *CETP*, 4 distinct LD groups containing 9, 8, 7 and 5 SNPs were identified, with 5 SNPs assigned to more than one LD group: rs158477 was linked to groups 2, 3 and 5; rs291044, rs11076175 and rs1800774 were linked to groups 2 and 3; and rs1800776 was linked to groups 2 and 6. The largest *CETP* LD group (group 1, n=9 SNPs) did not share a SNP with any other group. The relative locations of linkage groups for the highly variant and structurally complex *CETP* gene are presented in Figure 2. The least degree of LD was observed for the *APOF* and *LCAT* genes for which all LD groups (n=5) consisted of a single SNP.

Lipid associations with tagging SNPs

A preliminary assessment of the association of plasma HDL measurements and tagging SNPs was completed for healthy population controls who were without evidence of coronary disease and not taking lipid-lowering medication (n=786). Seven SNPs representing six of the LD groups in 4 of the studied genes showed significant associations with plasma HDL. Elevated HDL, a presumed protective phenotype, was observed for *CETP* group 2 (rs1800775 and rs708272); *LIPC* group 1 (rs1800588) and group 7 (rs6082); *SCARB* group 1 (rs5889), all p<0.05. Significantly reduced HDL was observed for *CETP* group 3 (rs11076175, p<0.0001) and *LPL* group 4 (rs11570897), p<0.05. Of these, only *CETP* rs11076175 remained significant using the multiple comparison-adjusted (Bonferroni), alpha=0.0017. For that SNP, GG homozygotes had a significant, 15% reduction in HDL (44.4±12.5 mg/dL versus 52.3±14.9 mg/dL, p<0.0001) and a significant 18% increase in triglycerides (137±102mg/dL versus 116±68mg/dL for AA homozygotes, p<0.05). The decrease in HDL associated with *LPL* rs11570897 CT genotype was also associated with a marked (25%) increase in triglycerides (120±72mg/dL versus 150±37mg/dL), but this was not significant.

Discussion

Technological developments and increasingly sophisticated analytical methods continue to further our understanding of the human genome. A current approach to understanding the genetic contribution to disease is GWAS with deep sequencing of disease-associated regions. Another approach is to study the variants in protein coding regions of the genome (exome) using exon capture and high-throughput sequencing. These applications are providing genomic information at a resolution previously unimaginable. However, there is limited availability to such technologies, and other methods may provide similar, useful information.

We applied the method of HRMCA to scan for variants in the coding regions (exons), splicing, and regulatory regions of six genes that are involved in lipid metabolism. This approach combines both physical and statistical methods to assess the impact of all variants in these regions on plasma lipids. A discovery set of 50 Caucasian samples and 14 samples from locally underrepresented ethnicities were scanned for the presence of variants by HRMCA. The size of the discovery set was chosen to give adequate power to detect polymorphisms with a minor allele frequency of ≥5% (C.I. 1-9%). The allele frequencies and LD of SNPs identified in the discovery set were estimated by genotyping each HRMCA observed SNP in a cohort of 339 individuals drawn from the same ancestry as the discovery set. SNPs with a minor allele frequency >1% were assigned to an LD group, and a tagging SNP(s) for each LD group was determined by PCA [12] and tested for association with plasma lipids.

In the present study, HRMCA showed good sensitivity (99.4%)

as well as good specificity (97.7%), but failed to identify 56 SNPs listed in the NCBI SNP database. It does not appear that lack of sensitivity explains the failure to identify these 56 SNPs. A more plausible explanation is the rarity and/or absence of these variants in the population we examined. Of the database SNPs that we failed to identify, only 3 of 37 (8.1%) had a calculated heterozygosity greater than 0.1. In contrast, 16 of 25 (64%) of the HRMCA observed SNPs had a calculated heterozygosity greater than 0.1. As previously mentioned, our discovery sample was drawn from the Utah population, which is comparatively homogeneous in ancestral composition. The NCBI dbSNP contains SNPs from 11 broad population classes with over 700 sample descriptions [14]. It is to be expected that a relatively homogeneous population would not possess many variants and/or would not possess variant alleles at a frequency likely to be observed in a sample of 50 individuals. Studies suggest that genetic variation differs by ancestry [15]. The fact that in our study, three SNPs were identified in the minority samples (n=14) that were not observed in the Caucasian discovery sample supports this hypothesis. Additionally, the rarity of the unobserved SNPs (0.02 median heterozygosity) suggests that these variants would have little impact on plasma lipids or risk for CHD on a population basis, and almost certainly would not have met the criteria (>1% minor allele frequency) for assignment to any specific LD group in this study.

The degree of variation in the six genes investigated in this study showed marked diversity. The *CETP* gene showed the most variation with 34 variants; contrasted with *LCAT* with only 3 variants discovered. Moreover, the LD structure of the *CETP* gene revealed a great deal more complexity than the other 5 genes. Twenty four of the 34 variants comprised 3 of 7 LD groups. The remaining LD groups included 2-5 SNPs. Four of the SNPs (rs1107617, rs158477, rs1800774, and rs291044) were included in both LD groups 2 and 3, suggesting a complex gene structure. SNPs showing LD with more than one haplotype were most common in the *CETP* gene but were also observed for *LIPC* where rs6082 linked to both groups 2 and 7 and in the *SCARB1* gene where rs5888 linked to groups 1, 2 and 3. Eight of the 25 LD groups contained a single SNP. The phylogenetic significance of a SNP showing no appreciable LD to other SNPs is not presently known. Although a high rate of recombination in the region would provide one explanation and is consistent with the presence of SNPs within several LD groups, further study is necessary to fully explain this observation.

Thirteen variants in five of the six genes studied showed associations with lipid values. Despite significant associations, the results were not subjected to a formal correction for multiple comparisons and should be interpreted with caution. Four of these variants were in the *CETP* gene and all were in *CETP* LD group 2. Although rs291044 and rs11076175 were also in LD group 3, it appears that the functional variant(s) are most likely linked to group 2 because no lipid-associated SNP was found to be exclusively a member of group 3. The fact that four of the eight SNPs in *CETP* group 2 were found to be associated with lipid values provides an "internal validation" and increases the degree of confidence in the observed lipid associations. Further, our results are consistent with a recent meta-analysis of 92 studies that showed rs1800775 and rs708272 (both in LD group 2) to be associated with modestly higher HDL levels [16]. Although we did not examine cardiovascular risk, in the meta-analysis the modestly increased HDL associated with these two SNPs translated into a similarly sized reduction in risk for CHD. As a follow-up to the present study, we plan to test the candidate SNPs reported here for association with CHD in a separate set of angiographically defined cases and controls.

The method used to investigate the contribution of genetic variation to plasma lipid composition proved to be a useful and robust approach. However, there are limitations to its applicability. Sample stratification between cases and controls can generate artifactual associations in any SNP-disease association study and is also true with this approach. Additionally, the various cohorts (SNP discovery, LD determination, disease cases and controls) must all share similar ancestry. Although a high degree of racial homogeneity in this study elevates confidence in the study findings and mitigates spurious associations, it also limits the application across groups of different ancestry and necessitates that similar studies be performed for individual racial groups.

A second limitation to this approach is that rare variants may be missed. According to the recently proposed rare variant-common disease hypothesis [17], one or a few rare variants with high penetrance may underlie common complex diseases. Under this model, rare variants may be undetected by the method reported here if the discovery set is inadequately large or is adequately large but consists primarily of healthy individuals lacking the high penetrance, albeit rare variant. Conversely, the traditional common disease-common variant hypothesis would predict that a predisposing variant should be detectable by this method. In the latter case, the difficulty comes in the analytical phase, where adequate power is needed to discern allele frequency differences between cases and controls.

Another limitation is the failure to address the additional genetic complexity imparted by the wealth of variation within the introns. Although in our approach we captured the conserved canonical donor-acceptor splice sites, other conserved regions within the intron which were not studied, such as the branch point, conserved adenosine residue, and the polypyrimidine tract, may also be relevant for splicing [18]. Under the current discovery scheme, variations in these latter splice elements would be undetected. Further, it is known that intronic regulatory regions exist for some genes [19]. Variation in such an intronic regulatory region that modifies gene expression would potentially have phenotypic effects and would not be detected by the scanning of exons. However, similar limitations are also associated with the Exome Project, aiming to develop and validate a cost-effective, high-throughput sequencing application for sequencing all of the protein coding regions of the human genome. This method is proposed to be the next level of resolution in the identification of genetic variation underlying Mendelian traits [20]. Although the Exome Project will achieve previously unattainable sensitivity for mutation detection, its availability is limited for the immediate future. The method of exon scanning by HRMCA offers an accessible alternative that was shown to provide excellent detection of heterozygous variants.

Conclusions

In the present study, we have used HRMCA to discover common variants in genes encoding enzymes, transport proteins and cell receptors that are key components of lipid metabolism. Using this method, along with LD group analysis and tagging SNP identification, we were able to capture the total variation in 90 SNPs for six genes with 38 tagging SNPs and identify associations with lipid measurements in healthy volunteers. Melting curve/tagging SNP association studies provide a level of resolution intermediate to GWAS and high-throughput sequencing. Whereas deep sequencing will be the tool of choice for identification of causative variations, it is presently cost-prohibitive for a study of this nature and is not readily available. The method described here can significantly reduce the genomic region of interest identified by a GWAS and facilitate the discovery of functional

genetic variations. The actual relationship of the lipid-gene variants reported here in relationship to CHD will require further investigation.

Funding Sources

Supported in part by grant NIH-NHLBI R01 HL071878 (JLA).

References

1. Rosamond W, Flegal K, Friday G, Furie K, Go A, et al. (2007) American Heart Association Statistics Committee and Stroke Statistics Subcommittee Heart disease and stroke statistics—2007 update: a report from the American Heart Association Statistics Committee and the Stroke Statistics Subcommittee. *Circulation* 115: 69-171.
2. Yusuf S, Reddy S, Öunpuu S, Anand S (2001) Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization. *Circulation* 104: 2746-2753.
3. Kannel WB, Castelli WP, Gordon T, McNamara PM (1971) Serum cholesterol, lipoproteins, and the risk of coronary heart disease. The Framingham study. *Ann Int Med* 74: 1-12.
4. Khor LL, Muhlestein JB, Carlquist JF, Horne BD, Bair TL, et al. (2004) Intermountain Heart Collaborative Study Group. Sex- and age-related differences in the prognostic value of C-reactive protein in patients with angiographic coronary artery disease. *Am J Med* 117: 657-664.
5. Horne BD, Muhlestein JB, Carlquist JF, Bair TL, Madsen TE, et al. (2000) Statin therapy, lipid levels, C-reactive protein and the survival of patients with angiographically severe coronary artery disease. *J Am Coll Cardiol* 36: 1774-1780.
6. Horswell SD, Ringham HE, Shoulders CC (2009) New technologies for delineating and characterizing the lipid exome: prospects for understanding familial combined hyperlipidemia. *J Lipid Res* 50: S370-S375.
7. Brewer HB (1999) Hypertriglyceridemia: changes in the plasma lipoproteins associated with an increased risk of cardiovascular disease. *Am J Cardiol* 83: 3F-12F.
8. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, et al. (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 40: 161-169.
9. Boekholdt SM, Sovereign OW, Tanck MW, Hovingh GK, Kuivenhoven JA, et al. (2006) Common variants of multiple genes that control reverse cholesterol transport together explain only a minor part of the variation of HDL cholesterol levels. *Clin Genet* 69: 263-270.
10. Lagor WR, Brown RJ, Toh SA, Millar JS, Fuki IV, et al. (2009) Overexpression of apolipoprotein F reduces HDL cholesterol levels in vivo. *Arterioscler Thromb Vasc Biol* 29: 40-46.
11. Elliott CG, Glissmeyer EW, Havlena GT, Carlquist JF, McKinney JT, et al. (2006) Relationship of BMPR2 mutations to vasoreactivity in pulmonary arterial hypertension. *Circulation* 113: 2509-2515.
12. Horne BD, Camp NJ (2004) Principal component analysis for selection of optimal SNP-sets that capture intragenic genetic variation. *Genet Epidemiol* 26: 11-21.
13. Zhou L, Myers AN, Vandersteen JG, Wang L, Wittwer CT (2004) Closed-tube genotyping with unlabeled oligonucleotide probes and a saturating DNA dye. *Clin Chem* 50: 1328-1335.
14. Tang H, Quetermous T, Rodriguez B, Kardia SLR, Zhu X, et al. (2005) Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *Am J Hum Genet* 76: 268-275.
15. Thompson A, Di Angelantonio E, Sarwar N, Erqou S, Saleheen D, et al. (2008) Association of cholesteryl ester transfer protein genotypes with CETP mass and activity, lipid levels, and coronary risk. *JAMA* 299: 2777-2788.
16. Schork NJ, Murray SS, Frazer KA, Topol EJ (2009) Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 19: 212-219.
17. Baralle D, Baralle M (2005) Splicing in action: assessing disease causing sequence changes. *J Med Genet* 42: 737-748.
18. Shamsheer MK, Chuzhanova NA, Friedman B, Scopes DA, Alhaq A, et al. (2000) Identification of an intronic regulatory element in the human protein C (PROC) gene. *Hum Genet* 107: 458-465.

19. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, et al. (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 42: 30-35.
20. Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, et al. (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet* 19: 216-217.