APOA1 gene polymorphisms in the South Asian immigrant population in the United States

Rebecca S. Henkhaus, Sunita Dodani, Ann M. Manzardo, Merlin G. Butler
University of Kansas Medical Center, Departments of Psychiatry and Behavioral Sciences and Pediatrics; University of Kansas School of Medicine, Department of Internal Medicine, 3901 Rainbow Blvd., Kansas City, Kansas, 66160, USA.

BACKGROUND: Coronary artery disease (CAD) is a leading cause of death in the United States. South Asian immigrants (SAIs) from the Indian subcontinent living in the US are disproportionately at higher risk of CAD than other immigrant populations. Unique genetic factors may predispose SAIs to increased risk of developing CAD when adopting a Western lifestyle including a higher-fat diet, more sedentary behavior and additional gene-environment interactions. SAIs are known to have low levels of the protective high density lipoprotein (HDL) and an altered function for Apo-lipoprotein A-1 (ApoA1), the main protein component of HDL cholesterol. One gene that may be genetically distinctive in this population is APOA1 which codes for ApoA-1 protein, a potentially important contributing factor in the development of CAD.

MATERIALS AND METHODS: DNA sequencing was performed to determine the status of the seven single-nucleotide polymorphisms (SNPs) in the APOA1 gene from 94 unrelated SAI adults. Genotypes, allelic frequencies, and intragenic linkage disequilibrium of the APOA1 SNPs were calculated.

RESULTS: Several polymorphisms and patterns were common among persons of south Asian ethnicity. Frequencies for SNPs T655C, T756C and T1001C were found to be different than those reported in European Caucasian individuals. Linkage disequilibrium was found to be present between most (13 of 15) SNP pairings indicating common inheritance patterns.

CONCLUSIONS: SAIs showed variability in the sequence of the APOA1 gene and linkage disequilibrium for most SNPs. This pattern of APOA1 SNPs may contribute to decreased levels of HDL cholesterol reported in SAIs, leading to an increased risk for developing CAD in this population.

Key words: APOA1, coronary artery disease, linkage disequilibrium, single nucleotide gene polymorphisms, south Asian immigrants.

Introduction

Genetic haplotypes and single nucleotide polymorphisms (SNPs) in the human genomic DNA sequence frequently differ among ethnic groups. It is imperative to study and document genetic variation among ethnic groups, especially as diagnosis and treatment of disease moves increasingly toward more personalized, specific approaches to medical care. Studying the genetics of geographically diverse groups of people has revealed important differences at the molecular level. A classic example is a single base pair change in the hemoglobin gene which causes sickle cell disease and occurs more often in persons of direct African descent. Many of these genetic discoveries have played critical roles in the diagnosis and treatment of a variety of conditions from classically inherited diseases to more complex disorders such as cancer and heart disease. For example, South Asian immigrants (SAIs-people with ancestors from the Indian subcontinent, i.e. India, Pakistan, Bangladesh, Nepal, Bhutan, and Sri Lanka) living abroad in regions with more Westernized cultures such as the United Kingdom, Canada and the United States have experienced higher rates of heart disease than other ethnic groups. Additionally, in the SAI population, the highest rates of mortality occur from...
coronary artery disease (CAD) and at a younger age than in other ethnic groups.\textsuperscript{[6-9]} SAI s constitute the second largest Asian immigrant population as well as the fastest growing ethnic immigrant population in the United States.

Many factors contribute to CAD. Clearly, an individual’s environment and behaviors contribute, but genetic factors play a large role especially in the immigrant population. Factors that may be partially influenced by an individual’s behavior, such as BMI, cholesterol levels and blood pressure are also largely determined by genetics.\textsuperscript{[10]} However, the factor that seems to be of major concern among the population of SAI s is the triglyceride level and high-density lipoprotein (HDL) to low-density lipoprotein (LDL) ratios. When compared to similarly age-matched Caucasians in the United States, South Asian men have lower levels of hypertension, tobacco usage, and obesity.\textsuperscript{[11]} But, the same study demonstrated that SAI s in the United States have dyslipidemia (high LDL and triglyceride levels, accompanied by low HDL levels) and a greater prevalence of noninsulin-dependent diabetes mellitus, or type 2 diabetes.\textsuperscript{[11]} High levels of HDL cholesterol and a high ratio of HDL to LDL cholesterol are protective against heart disease. Low HDL levels alone or accompanied by high triglyceride and high lipoprotein levels have been documented in the SAI s Indian population.\textsuperscript{[12]} The size of HDL particles has been shown to be smaller in South Asian men; larger HDL particles may be more effective at scavenging cellular cholesterol and facilitating its removal from circulation. Smaller HDL particles are not as useful in facilitating reverse uptake of circulating cholesterol.\textsuperscript{[13,14]} As cholesterol and specifically low HDL cholesterol levels appear to play a central role in the development of heart disease in South Asian immigrants, this article explores the underlying genetics of this phenomenon and the role of \textit{APOA1} gene polymorphisms in our SAI study population.

The \textit{APOA1} gene encodes the protein, ApoA-1, which is the principle protein component of HDL cholesterol. Low serum levels of ApoA-1 correlate with the occurrence of heart disease.\textsuperscript{[15]} The \textit{APOA1} gene is contained within the \textit{APOA1/C3/A4/A5} cluster on chromosome 11q23, with more than 182 distinct SNPs identified in this gene cluster region.\textsuperscript{[16]} Analysis of the \textit{APOA1/C3/A4/A5} gene cluster has shown linkage disequilibrium within a 150 kb span of the region.\textsuperscript{[17]} Previously SAI s have been reported to carry six specific SNPs in the \textit{APOA1} gene (Table 1) located in the intronic regions of the \textit{APOA1} gene.\textsuperscript{[18]} Although these SNPs are not located in the coding regions of the gene, they potentially can impact \textit{APOA1} gene expression. Additionally, we analyzed a separate SNP in the 5’ UTR (-75G>A) region that is present in an estimated 11-35\% of the population, in a variety of racial and ethnic groups, and shown to decrease \textit{APOA1} gene expression and ApoA-1 protein translation.\textsuperscript{[16,19,20]}

The characterization of the \textit{APOA1} gene is of importance due to the role of HDL cholesterol and ApoA-1 levels in the development of heart disease in the SAI population. Thus, we determined the frequency and pattern of occurrence of SNPs in the \textit{APOA1} gene in the SAI population. Certain polymorphic alleles were found to be more common than the so called wild-type alleles, using the National Center for Biotechnology Information (NCBI) database \textit{APOA1} sequence (NC_000011.9), highlighting the fact that the consensus genomic DNA reference sequences do not always reflect every ethnic population studied.

### Materials and Methods

#### Sample collection and demographics: Adult SAI males and females (n=74, age range of 35-65 years) representing first and second generations living in the United States were recruited in the Kansas City Metro area at local places of worship following the distribution of flyers and announcements outlining the purpose,

### Table 1: SNPs in the human \textit{APOA1} gene present in the SAI population

<table>
<thead>
<tr>
<th>SNP nucleotide locus</th>
<th>Wild-type allele</th>
<th>Polymorphic allele</th>
<th>Location in gene</th>
<th>dbSNP reference number</th>
</tr>
</thead>
<tbody>
<tr>
<td>-75</td>
<td>G</td>
<td>A</td>
<td>5’ UTR</td>
<td>rs670</td>
</tr>
<tr>
<td>319</td>
<td>C</td>
<td>T</td>
<td>Intron 2</td>
<td>rs5070</td>
</tr>
<tr>
<td>655</td>
<td>C</td>
<td>T</td>
<td>Intron 3</td>
<td>rs2070665</td>
</tr>
<tr>
<td>756</td>
<td>C</td>
<td>T</td>
<td>Intron 3</td>
<td>rs5072</td>
</tr>
<tr>
<td>938</td>
<td>T</td>
<td>C</td>
<td>Intron 3</td>
<td>rs12718464</td>
</tr>
<tr>
<td>1001</td>
<td>C</td>
<td>T</td>
<td>Intron 3</td>
<td>rs7116797</td>
</tr>
<tr>
<td>1149</td>
<td>T</td>
<td>C</td>
<td>Intron 3</td>
<td>rs5076</td>
</tr>
</tbody>
</table>

Wild-type allele is determined according to the most recent human sequence assembly (GRCh37). Nucleotide locus refers to the \textit{APOA1} NCBI reference sequence (NC_000011.9).
Table 2: Frequency of genotypes and alleles at each APOA1 SNP

<table>
<thead>
<tr>
<th>SNP ID locus in APOA1 gene</th>
<th>Homozygous wild-type (%)</th>
<th>Homozygous polymorphism (%)</th>
<th>Heterozygous (%)</th>
<th>Wild-type allele frequency</th>
<th>Polymorphic allele frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP 1: G-75A</td>
<td>49 (66.22)</td>
<td>5 (6.76)</td>
<td>20 (27.03)</td>
<td>0.797</td>
<td>0.203</td>
</tr>
<tr>
<td>SNP 2: T319C</td>
<td>51 (54.26)</td>
<td>9 (9.57)</td>
<td>34 (36.17)</td>
<td>0.723</td>
<td>0.277</td>
</tr>
<tr>
<td>SNP 3: T655C</td>
<td>12 (12.77)</td>
<td>50 (53.19)</td>
<td>32 (34.04)</td>
<td>0.298</td>
<td>0.702</td>
</tr>
<tr>
<td>SNP 4: T756C</td>
<td>12 (12.77)</td>
<td>51 (54.26)</td>
<td>31 (32.98)</td>
<td>0.293</td>
<td>0.707</td>
</tr>
<tr>
<td>SNP 5: C938T</td>
<td>73 (77.66)</td>
<td>0 (0)</td>
<td>21 (22.34)</td>
<td>0.888</td>
<td>0.112</td>
</tr>
<tr>
<td>SNP 6: T1001C</td>
<td>11 (11.70)</td>
<td>43 (45.74)</td>
<td>40 (42.55)</td>
<td>0.330</td>
<td>0.670</td>
</tr>
<tr>
<td>SNP 7: C1149T</td>
<td>86 (91.49)</td>
<td>0 (0)</td>
<td>8 (8.51)</td>
<td>0.957</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Nucleotide locus refers to the APOA1 NCBI reference sequence (NC_000011.9). (SNP1, n=74 subjects; SNP2-7, n=94 subjects)
polymorphic alleles for these SNPs did not necessarily coincide with SNPs at positions 655, 756 and 1001. The C938T and C1149T SNPs were also far less common, occurring in only 11.2 and 4.3% of subjects, respectively, with no subjects exhibiting homozygosity for these polymorphisms. The C319T SNP was found in 27.7% of subjects. No apparent association was detected with the occurrence of any other SNP. The SNP (G-75A) located in the 5'UTR, near the promoter region was found in 20.3% of 74 SAI subjects studied for the SNP.

Patterns of SNP occurrence in SAI subjects:

Patterns of SNPs were analyzed to discover the most common concurrent groupings of SNPs [Figure 1a]. Slightly more than one half of the subjects (51.2% or 43 of 82) who carried the 655/756 polymorphic alleles also carried the T319C polymorphism. Three patterns of SNP occurrence covered 84.1% (69 of 82) of the subjects studied having more than one polymorphic allele. These patterns included polymorphisms at positions 655/756/1001, 319/655/756/1001 or 655/756/938/1001. The remaining 13 subjects with more than one SNP comprised all other patterns making them far less common. Overall, there were nine separate patterns of polymorphism occurrence in the subjects studied [Figure 1a]. Of the nine patterns, each with six SNPs, 21 of 54 most prevalent alleles were wild-type and 33 were polymorphic, making the polymorphisms more common than wild-type alleles in the SAI population.

Linkage disequilibrium and Hardy-Weinberg analysis: Linkage disequilibrium (LD) analysis was performed on each combination of the six SNPs within the APOA1 gene [Figure 1b]. All alleles were found to be significantly in LD except for two pairings: 319 vs. 938 and 938 vs. 1149. Three SNPs (655, 756 and 1001) were found to have a very high degree of LD as compared to the other pairings. These data indicate that SNP 938 is inherited in a more independent manner than the other SNPs within the APOA1 gene while SNPs 655, 756 and 1001 had the highest degree of LD. Given the observed frequency of each allele, none were significantly different from expected Hardy-Weinberg equilibrium (data not shown).

Discussion

Gene sequencing and SNP assays are becoming easier and more widely available, leading to establishment of personalized medicine approaches in treating diseases. Genetic differences between ethnic groups will most certainly reveal the causes for differences in susceptibility to a variety of diseases. Characterization of polymorphisms in the APOA1 gene may well shed light on why South Asians are disproportionately affected by heart disease with ApoA-1 protein levels being the single most powerful predictor for the development of heart disease. [23-25] This study further shows that the sequence of the APOA1 gene in SAIs is commonly divergent from the wild-type gene sequence, the NCBI reference sequence NC_000011.9, established from the Human Genome Reference Consortium (GRCh37) primary assembly. This most recent sequence assembly was compiled using data collected from multiple individuals, likely from a different (non-South Asian) ethnic background. It should be noted that the Celera Assembly reference for APOA1 indicates different wild-type alleles for the SNPs analyzed in this study: G-75A, C319T, C655T, C756T, C938T, C1001T and C1149T.
difference in reference assembly sequences confirms that these polymorphisms are highly diverse among individuals, underscoring the need to elucidate and characterize the sequence differences among ethnic groups and determine the impact of SNPs on ApoA-1 protein production.

The SNPs analyzed in our study have been examined in a variety of ethnic groups. For example, the SNP located in the 5'UTR (G-75A) has been the most widely studied. A meta-analysis was previously performed for the minor-allele frequency (MAF) determination with G-75A polymorphism across a range of ethnic groups.[19] For our discussion, the term MAF refers to the frequency of the polymorphic allele, although it may not represent the allele with the lower frequency. Caucasians primarily of northern and western European descent have an MAF of 0.11-0.21.[19] Individuals of Asian (Japan and Singapore) descent have an MAF ranging from 0.14 to 0.28.[19] Our study demonstrated that SAIs have an MAF of 0.20 for this SNP.

The other six SNPs within the APOA1 gene have also been studied in multiple ethnic groups. Allele frequencies obtained from the SNP database (dbSNP) (http://www.ncbi.nlm.nih.gov/snp)[20] for studies representing diverse ethnic groups from Europe, Africa, Thailand, China and India have been summarized in Table 3. Notable differences from our study include the higher prevalence of polymorphic allele in persons of European heritage at position 319, and the decreased frequency of the polymorphic allele in Thai individuals at positions 655, 756 and 1001. The alleles analyzed in individuals of Indian ethnicity living in the United States were all highly similar to the SNP allele frequencies observed in our SAI population [Table 3].

In addition to determining the SNP frequencies for the SAI population, the intragenic linkage disequilibrium was also analyzed using established protocols for the six SNPs within the APOA1 gene. Of the 15 pairwise analyses, 13 were found to be in linkage disequilibrium. The SNPs at locations 655, 756 and 1001 had the highest degree of linkage while the SNPs at locations 319, 1149, and especially 938 had a more independent inheritance pattern [Figure 1]. A study by Qi et al.[27] demonstrated that common intragenic haplotypes can be predictive of clinical outcomes. Their study reported a haplotype in the human perilipin gene (PLIN) common to Malays and Indian populations that was significantly correlated with increased obesity risk as compared to individuals of Chinese descent, with all three ethnic populations living in Singapore.[27] A similar study could eventually be carried out with SAI subjects compared with differing ethnic populations, assessing the correlation between APOA1 haplotypes and relative risk for developing CAD.

These data provide an excellent basis for further studies of APOA1 gene expression in the SAI population. Although the G-75A SNP has previously been studied in relation to gene expression, the other SNPs have not. Research has shown that the presence of the polymorphic allele at the G-75A SNP is associated with lower APOA1 gene expression.[20] A study conducted on individuals from Northern India showed that significantly lower levels of both plasma ApoA-1 and HDL cholesterol were present in persons carrying the A allele, as opposed to those with the GG (wild-type) genotype.[28] The other SNPs may have a comparable impact on APOA1 gene expression, and further explorations are currently underway using our study population. The initial study which reported the presence of the six intronic SNPs in the SAI population also demonstrated that the C938T

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>G-75A (%)</th>
<th>T319C (%)</th>
<th>T655C (%)</th>
<th>T756C (%)</th>
<th>C938T (%)</th>
<th>T1001C (%)</th>
<th>C1149T (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Caucasian††</td>
<td>G (69)</td>
<td>C (66)</td>
<td>C(89-92)</td>
<td>C(89-92)</td>
<td>C(94-96)</td>
<td>C (92)</td>
<td>C (98)</td>
</tr>
<tr>
<td>African American††</td>
<td>G (83)</td>
<td>T (59)</td>
<td>C (93)</td>
<td>C (93)</td>
<td>C (97)</td>
<td>C (69)</td>
<td>C (73)</td>
</tr>
<tr>
<td>Thai†</td>
<td>G (87)</td>
<td>T (66)</td>
<td>T (53)</td>
<td>T (57)</td>
<td>C (97)</td>
<td>T (67)</td>
<td>C (94)</td>
</tr>
<tr>
<td>Chinese†</td>
<td>N/A</td>
<td>T (67)</td>
<td>C(58-70)</td>
<td>C(58-70)</td>
<td>C (96)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Indian†</td>
<td>N/A</td>
<td>N/A</td>
<td>C (69)</td>
<td>C (69)</td>
<td>C (90)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>South Asian immigrant (current study population)</td>
<td>G (80)</td>
<td>T (72)</td>
<td>C (70)</td>
<td>C (71)</td>
<td>C (89)</td>
<td>C (67)</td>
<td>C (96)</td>
</tr>
</tbody>
</table>

Data provided by the NCBI SNP database (http://www.ncbi.nlm.nih.gov/snp). Shaded sections indicate that the polymorphic allele is most prevalent. (PARC-Rochester European Panel, n=24; †HapMap-CEU Panel, n=180; †PARC-Jackson African American Panel, n=24; †HapMap-ASW Panel, n=90; Thais32-16Pools Panel, n=32; †HapMap-HCB Panel, n=43; ††HapMap-GIH Panel, n=88)
SNP is significantly correlated with low plasma HDL cholesterol (<40 mg) levels.[18] Furthermore, a study performed comparing first generation American-born adults of South Asian heritage with American-born adults of European heritage demonstrated that the South Asian individuals had lower HDL levels, but higher LDL and triglyceride levels; yet the South Asians had lower BMI.[20]

A recently published study on this cohort of SAI subjects revealed a statistically significant correlation between three of the APOA1 SNPs (C655T, C756T, and C1001T) and metabolic syndrome (MS) with positive odds ratios ranging from 2.5 to 3.0 (CI>95%).[21] Negative odds ratios (CI>95%) were also observed for HDL cholesterol, Apo-A1 protein and lipoprotein(a), indicating that high levels of these proteins, all of which are influenced by APOA1 gene expression patterns and protective against MS.[21] These findings may be of even greater relevance given the data reported herein that the three significantly correlated SNPs have the highest degree of LD, indicating a conserved hereditary pattern. The occurrence of MS was also significantly correlated with the presence of pro-inflammatory HDL (dys-HDL) as well as increased common carotid artery intima-media thickness (CCA-IMT), a surrogate marker for CAD.[21] Given that SAI are known to carry a disproportionately high risk for CAD, there is need to explore and understand gene-environment interactions.[6-9] This increased risk seems to be primarily due to dyslipidemia, as opposed to other common risk factors such as obesity and high blood pressure. Apo-A1 is a key component of HDL and genetic changes that affect APOA1 expression may well play an important role in CAD as well as MS.

Clearly, genetic factors play a large role in determining cholesterol levels and ratios in all human populations. Family and twin studies have shown that genetic heritability accounts for nearly two-thirds of the variability observed between individuals' HDL and ApoA-1 levels.[20,30] In the SAI population, these genetic factors may lead to an increased incidence and mortality resulting from heart disease. While living in the countries and cultures of South Asia, the propensity toward dyslipidemia may not lead to a selective disadvantage. The diet and level of activity more common in individuals living in southern Asia may not interplay with the genetic predisposition to lead to common occurrence of heart disease, as opposed to the more Western lifestyle of the United States, Canada and the United Kingdom with greater access to animal products, leading to more consumption of cholesterol in the diet. This research and future related studies should lend crucial insight into the causes of heart disease in South Asians living in Westernized societies, and certainly will provide the knowledge needed for more targeted and effective prevention and treatment options.

**Acknowledgements**

We would like to thank the participants who contributed their valuable time to further our knowledge. We are also thankful to the Hindu temples in the Kansas City Metro area for supporting our study and helped with subject recruitment. We thank the study coordinator Ms. Kuntal Shastri for administrative assistance, as well as Yan Bin Dong and Haidong Zhu at the Medical College of Georgia for their assistance in the preparation and sequencing of previous SAI DNA samples.

**References**