Published in final edited form as:
Pharmacogenet Genomics. 2008 March ; 18(3): 169–179. doi:10.1097/FPC.0b013e3282f44d99.

Molecular population genetics of *PCSK9*: a signature of recent positive selection

Keyue Ding and **Iftikhar J. Kullo***
Division of Cardiovascular Diseases, Mayo Clinic and Foundation, Rochester MN, USA

Abstract

Objective—Proprotein convertase subtilisin-like kexin type 9 (*PCSK9*) is a newly discovered serine protease that plays a key role in regulating plasma low-density lipoprotein (LDL) cholesterol levels. Both rare mutations and common variants in the coding regions of *PCSK9* affect LDL cholesterol levels and coronary heart disease risk, as well as response to lipid-lowering therapy.

Methods—We characterized the patterns of variation at the *PCSK9* locus in African-Americans and European-Americans using resequenced data from the SeattleSNPs database (pga.gs.washington.edu). We performed a test of population differentiation and the long range haplotype (LRH) test to detect signatures of recent position selection on *PCSK9*.

Results—A significantly high F_{ST} (a measure of population differentiation) between African-Americans and European-Americans was noted for SNP rs505151 (F_{ST} = 0.309). The LRH test was suggestive of non-neutral evolution of two single nucleotide polymorphisms (SNPs) (rs505151 and rs562556) in *PCSK9* that are associated with elevated LDL cholesterol levels ('gain-of-function' mutations), with differential modes of selection in African-Americans and European-Americans. We observed signals of recent positive selection on the ancestral allele of nonsynonymous SNP rs505151 (E670G, P = 0.0227 and P = 0.0001 in theoretical and empirical distribution, respectively) and the derived allele of nonsynonymous SNP rs562556 (I474V, P = 0.0227 and 0.0001) in African-Americans, whereas in European-Americans the ancestral allele of SNP rs562556 (P = 0.1320 and 0.0370) appeared to be under positive selection.

Conclusions—Our findings suggest that evolutionary dynamics may underlie the gain-of-function mutations in *PCSK9* that influence inter-individual variation in LDL cholesterol levels.

Keywords

PCSK9; *low-density lipoprotein cholesterol*; natural selection

Introduction

Proprotein convertase subtilisin-like kexin type 9 (*PCSK9*, OMIM 607786) is a newly discovered serine protease that plays a key role in low-density lipoprotein (LDL) cholesterol homeostasis by mediating LDL receptor (LDL-R) breakdown through a post-transcriptional mechanism [1-4]. *PCSK9* may also regulate apolipoprotein B-containing lipoprotein production and apoB secretion [5,6], and promote production of nascent very low-density lipoprotein (VLDL) in the fasting state [7]. Adenoviral-mediated over-expression of human *PCSK9* in mice promotes the accumulation of LDL cholesterol in the plasma but this response

Correspondence: Dr. Iftikhar J. Kullo, 200 First Street SW, Rochester MN, 55905, USA, kullo.iftikhar@mayo.edu, Tel.: 507-284-9049, Fax: 507-266-1702.

Competing interest statement: The authors declare that they have no competing financial interests.

is absent in LDL receptor-deficient animals [2,4,8]. Recent studies show that *PCSK9* binds directly to the extracellular domain of the LDL receptor [9,10] and increases its degradation [9]. *PCSK9* expression has been detected in tissues other than the liver and intestine, such as the cerebellum, where *LDLR* expression is not prominent [11]. *PCSK9* may enhance degradation of other receptor types or proteins during the development of cerebellum and telencephalon [11] and promote cerebellar cortical neurogenesis, possibly by increased recruitment of undifferentiated neural progenitor cells into the neuronal lineage [12].

The characterization of 'gain-of-function' versus 'loss-of-function' alleles of *PCSK9* is based on the phenotype (plasma LDL cholesterol levels), and not on defined biochemical alterations. Missense mutations that increase *PCSK9* activity (i.e., gain-of-function mutations) are associated with hypercholesterolemia and coronary heart disease (CHD) [6,13-15]; mutations that inactivate *PCSK9* (i.e., loss-of-function mutations) have the opposite effect, lowering LDL cholesterol levels and reducing risk of CHD [16,17]. Kotowski et al. [18] described a spectrum of nonsense/missense mutations in *PCSK9* that were associated with low or elevated LDL levels, in both non-Hispanic black and non-Hispanic white subjects. The frequency spectrum of mutations, however, varied significantly among non-Hispanic blacks and non-Hispanic whites. For example, two nonsense loss-of-function mutations (Y142X and C679X) are rare in whites but present in approximately 2% of blacks [16,17]. However, the missense loss-of-function R46L mutation is more common in whites (3.2%) than in blacks (0.6%) [16,18,19]. In addition to rare mutations, common genetic variations of *PCSK9* [e.g., C(-161)T in intron 1 and I474V in exon 9] have been linked to plasma LDL cholesterol levels in the Japanese [20], and the E670G SNP has been associated with LDL cholesterol levels and the severity of atherosclerosis in participants of the Lipoprotein Coronary Atherosclerosis Study [21].

By virtue of its role as a major inhibitor of *LDLR*, *PCSK9* is a promising therapeutic target [22-24]. The cholesterol-lowering effect of statins is increased in subjects carrying loss-of-function mutations in *PCSK9* [19], suggesting that lipid-lowering by *PCSK9* inhibitors may be synergistic to that achieved by statins [24-26]. Statins activate a pathway that leads to both up-regulation (increased transcription of *LDLR*) and down-regulation (increased transcription of *PCSK9*) of *LDLR*. Four missense mutations in *PCSK9* (i.e., R46L, G106R, N157K, and R237W) have been associated with hypocholesterolemia and possibly increased response to statin therapy [19]. Kotowski et al. [18] estimated that such loss-of-function nonsense mutations could lead to a 88% reduction in coronary heart disease (CHD) over a 15-year-period, suggesting that inhibition of *PCSK9* may represent a safe and effective strategy for the control of hyperlipidemia. Gain-of-function mutations in *PCSK9* that lead to decreased clearance of plasma LDL were associated with higher pretreatment LDL cholesterol levels and attenuated statin-mediated reduction of LDL cholesterol [21,27]. This newly discovered element of the lipoprotein regulatory system may influence response to statins [28].

Decreased free cholesterol in the hepatocytes due to a diet low in saturated fat and cholesterol or as a result of lipid-lowering therapy (i.e., statins), leads to activation of the sterol regulatory element binding protein (SREBP), a key transcription factor for *LDLR*. Activation of SREBP not only increases expression of *LDLR* mRNA and *LDLR* protein, but also *PCSK9*, the latter acting as a counter-regulatory mechanism to prevent excessive uptake of cholesterol into cells. Several evolutionary modes have been noted in *LDLR*, a gene that is co-regulated with *PCSK9*, indicating that natural selection has shaped the variation in *LDLR* and regulated its expression at different time-scales. These include primate-specific positive selection on the *LDLR* 5′ enhancer [29] and balancing selection on the 3′-UTR regions [30]. Since *PCSK9* triggers the degradation of LDLR and the genes are co-regulated [25], *PCSK9* may be involved in this regulatory network and itself be a target of natural selection.

The Y142X and C679X mutations in Africans and the haplotype structure of neighboring chromosomal regions describe ancient alleles that may have conferred a selective advantage [17,31]. It has been speculated that inactivation of *PCSK9* may have been beneficial, for example by interfering with the life cycle of the malaria parasite or leading to increased LDLR activity in the liver thereby reducing the exposure of peripheral tissues to viruses or other infectious agents that circulate in association with lipoproteins [32]. However, the high frequency of *PCSK9* nonsense mutations in Africans, but not in Europeans, may simply be a result of genetic drift. In the present study, we performed a molecular population genetics study of *PCSK9* variants in African-Americans and European-Americans to characterize the common patterns of variation and investigate whether there is a signature of positive selection in the gene, especially the common variations. Resequenced data for African-Americans and European-Americans from SeattleSNPs (pga.gs.washington.edu) were obtained. We performed tests of nucleotide diversity, estimated population differentiation (F_{ST} statistic), and performed the long range haplotype (LRH) test to assess for signatures of selection.

Materials and Methods

Genotype data

Resequenced data for *PCSK9* – from 24 African-Americans and 23 European Americans – were downloaded from the SeattleSNPs website (pga.gs.washington.edu) [33] on Oct 16, 2006. There were 265 *PCSK9* genetic variants including 247 diallelic SNPs (229 in African-Americans and 125 in European-Americans), 17 indels, and one triallelic SNP (Figure 1b). We also retrieved the genotype data for *PCSK9* SNPs/variants from the HapMap database (Phase II, www.hapmap.org). There were 35 co-segregating SNPs identified in Yoruba in Africa (YRI), Europeans in CEPH (CEU), and Han Chinese in Beijing / Japanese in Tokyo. The DNA sequence of a chimpanzee (*Pan troglodyte*) (genome.ucsc.edu) was used as an outgroup to define the ancestral alleles.

Data analysis

We used the ‘genetics’ package implemented in *R* to perform population genetics analyses, including description of allele frequencies, and Fisher's exact test for Hardy-Weinberg equilibrium (HWE). The false discovery rate (FDR) method was used to correct for multiple testing using the package ‘QVALUE’ in *R* [34]. A measure of linkage disequilibrium (LD, D ′) was estimated for common SNPs (minor allele frequency (MAF) > 0.10 in at least one population) using the ‘LDheatmap’ package in *R*. Haplotype reconstruction was performed using the Bayesian method implemented in the PHASE program [35].

F_{ST}, a measure of population differentiation, quantifies variance of allele frequency between and within populations, and is used to detect local adaptation in the human genome [36-38]. We calculated an unbiased estimate of F_{ST} from pairwise population comparisons. A randomization method (1000 permutations) was used to test the statistical significance of F_{ST} in each pairwise population comparison using the ‘fstat’ software [39].

We used the long-range haplotype test to assess for recent positive selection [40,41]. We refer to the common SNPs as ‘core’ SNPs in the present study. Haplotype homozygosity (HH) was calculated in a stepwise manner – extended HH (EHH) – to assess how LD breaks down with increasing distance from a specific core SNP. HH was calculated between a distance *x* and the specified core SNP for a chromosome population carrying a single allele of the core SNP (ancestral and derived allele, respectively). Distance *x* increases stepwise to the most outlying SNP. The patterns of haplotype homozygosity were estimated on both sides of each allele for a specific core SNP. EHH is on a scale of 0 (no homozygosity, all extended haplotypes are different) to 1 (complete homozygosity, all extended haplotypes are the same). Relative EHH

(REHH) is the ratio of the EHH on the tested core haplotype (core allele here) compared with the EHH of the grouped set of core haplotypes at the region not including the core haplotype tested [40]. The details of calculating HH and variance of HH have been previously described [41].

Population demographic history can also result in the rejection of the null hypothesis of neutrality. To test for statistically significant evidence of selection on core SNPs, we used coalescent theory to obtain the expected distribution of EHH under a calibrated demographic model for African-American and European-American populations [42]. Using the program 'cosi' (www.broad.mit.edu/~sfs/cosi), we simulated a one megabase (MB) region 1,000 times under the 'best-fitting' population parameters for two populations. The 'best-fitting' set of parameters yielded good agreement with all aspects (including allele frequency spectrum, fraction of alleles that are ancestral, linkage disequilibrium, and F_{ST}) of the observed data in the human genome [42]. This model has been previously used to assess adaptive evolution in *TCF7L2* [43] and *NAT* [44] genes in the context of the LRH test. In addition, we obtained the empirical distribution of core haplotype frequencies versus REHH by screening the entire chromosome 1 HapMap data (release #16) in Yoruban (YRI), and European-descent populations (CEU).

We followed the method of Voight et al [45] to obtain an integral haplotype score (iHS). First, we calculated the integral of the decay of EHH away from a specified core allele until EHH reached 0.05. The integrated EHH (iHH) (summed over both directions away from the core SNP) is denoted iHH_A or iHH_D for the ancestral and derived core allele, respectively. Second, the test statistic of unstandardized iHS was denoted as: $\ln(iHH_A/iHH_D)$. If unusually long haplotypes carry the derived allele, a large negative value of unstandardized iHS would exist; a large positive value indicates long haplotypes carrying the ancestral allele. Finally, the unstandardized iHS can be adjusted using the expectation (E_p) and standard deviation (SD_p) of $\ln(iHH_A/iHH_D)$ regardless of allele frequency at the core SNP:

$$iHS=\frac{\ln(iHH_A/iHH_D)-\mathrm{E_p}[\ln(iHH_A/iHH_D)]}{SD_p[\ln(iHH_A/iHH_D)]}$$

. The E_p and SD_p of $\ln(iHH_A/iHH_D)$ were estimated from the empirical distribution at SNPs whose derived allele frequency p matches the frequency at the core SNP from phased data of chromosome 1 in the HapMap database (Phase I, release 16a). The values of iHS were assigned as 'not available' if (1) iHH_A or $iHH_D = 0$, and (2) there was no matched frequency at the core SNP from phased data of chromosome 1 in the HapMap database.

Results

Data summary and sequence variations

We analyzed sequence variation in *PCSK9* based on re-sequenced data from SeattleSNPs database (shown in Figure 1b) for African-Americans and European-Americans, respectively (Table 1). Only 8 out of 229 SNPs in African-Americans and 3 out of 126 SNPs in European-Americans departed significantly from Hardy-Weinberg equilibrium (Fisher's exact test, $P < 0.05$). However, after correction for multiple testing using false discovery rate, these departures were not statistically significant. Three measures of nucleotide diversity – θ_w, π, and θ_H – were calculated for the resequenced regions. The average nucleotide diversity (π) in the whole genomic sequence (i.e., 15.06×10^{-4} in African-Americans and 10.31×10^{-4} in European-Americans, respectively) was larger than that reported previously [46]. Nucleotide diversity was higher in coding regions (9.13×10^{-4}) than the 5′ flanking regions (7.44×10^{-4}) in African-Americans, whereas in European-Americans, nucleotide diversity was greater in 5′ flanking regions (6.98×10^{-4}) than in coding regions (4.08×10^{-4}).

The classical tests of evolutionary neutrality test based on nucleotide diversity – Tajima's D [47] and Fay and Wu's H [48] – were calculated for coding regions, 5′ flanking regions, and the entire gene, respectively. A coalescent simulation based on the neutral model and the population structure model [49] was used to assess the statistical significance of Tajima's D and Fay and Wu's H. In African-Americans, the 5′ flanking region of *PCSK9* had a negative Tajima's D and positive Fay and Wu's H ($P < 0.05$). In European-Americans, the 5′ flanking region of *PCSK9* had a positive Tajima's D and Fay and Wu's H ($P < 0.05$).

Population differentiation

We calculated F_{ST}, a measure of population differentiation, to test whether there are significant differences in allele frequencies between African-Americans and European-Americans for genetic variants in *PCSK9*. In the SeattleSNPs data for *PCSK9*, 18 out of 247 loci showed significantly high F_{ST} ($P < 0.05$) between African-Americans and European-Americans (Figure 2a). A significantly high F_{ST} was noted in one nonsynonymous SNP in the C-terminal domain (rs505151, E670G) ($F_{ST} = 0.309$). We also used the genotype data for three populations (i.e., YRI, CEU, and CHB+JPT) from the HapMap project (www.hapmap.org) to calculate F_{ST}. *PCSK9* showed significantly high levels of population differentiation in the HapMap data: 22 out of 35 loci in YRI vs. CEU, 31 out of 35 in YRI vs. CHB+JPT, and 19 out of 35 in CEU vs. CHB+JPT. The SNP rs505151 also showed a significantly high F_{ST} among Africans and non-Africans: 0.234 in YRI vs. CEU, and 0.242 in YRI vs. CHB+JPT.

Linkage disequilibrium patterns

We estimated a measure of linkage disequilibrium (LD, D′) of common SNPs (minor allele frequency ≥ 0.10 in at least one population: 89 in African-Americans and 69 in European-Americans) in the *PCSK9* region to determine LD patterns in African-Americans and European-Americans, respectively (Figure 3). In the *PCSK9* region, a third of SNP pairs (1370/3916, 35.0%) in African-Americans and half of the SNP pairs (1159/2346, 49.4%) in European-Americans showed significant LD. The LD pattern for *PCSK9* in European-Americans indicated 3 continuous haplotype blocks. Within each haplotype block in European-Americans, there was strong and significant LD: 70.7% in block A (from rs17111503 to rs572512), 81.2% in block B (from rs572512 to rs7525407), and 90.2% in block C (from rs7525407 to rs597387). Within each haplotype block, more than half SNP pairs showed nearly complete LD (e.g., $|D'| = 1$). The pattern of three-haplotype blocks in *PCSK9* was less obvious in African-Americans. For example, in block C (from rs7525407 to rs597387), 62.2% SNP pairs showed nearly complete LD and 50.3% were significant.

Long-range haplotype test for positive selection

In order to identify mutations in the *PCSK9* gene potentially targeted by positive selection, we performed the long-range haplotype (LRH) test. The common SNPs (MAF ≥ 0.05) from SeattleSNPs data were selected as ‘core’ SNPs, including 89 SNPs in African-Americans, and 69 SNPs in European-Americans. We tested the relative extended haplotype homozygosity (REHH) in *PCSK9*, using the theoretical distribution based on the calibrated population model for African-Americans and Europeans, as well as the empirical distribution of core haplotype frequencies versus REHH through screening the entire chromosome 1 HapMap data in Yoruban (YRI) and European-descent populations (CEU). Based on both the theoretical and empirical distribution of REHH, several SNPs in both coding and regulatory regions deviated significantly from evolutionary neutrality in specific populations (Figure 4). Tables 2 and 3 show SNPs with a significantly high value of REHH in African-Americans and European-Americans, respectively. Significantly high REHH was noted in a regulatory region SNP (rs2479409) and 2 nonsynonymous SNPs [rs562556 (I470V) and rs505151 (E670G)]. The derived allele (G) of SNP rs2479409 in African-Americans deviated significantly from

neutrality (P = 0.0856 in simulated distribution, and P = 0.0269 in empirical distribution). In the coding regions, the ancestral allele (A) of rs505151 (P = 0.0227 and 0.0001), and the derived allele (A) of rs562556 (P = 0.0222 and 0.0061) showed statistical evidence of positive selection in African-Americans. We also noted that the ancestral allele (G) of rs562556 (P = 0.1320 and 0.0370) in European-Americans deviated significantly from neutrality.

We next plotted the EHH decay versus physical distance for SNPs rs2479409, rs562556 and rs505151 (Figure 5). The overall EHH of the derived allele of rs2479409 decayed more slowly than the ancestral allele in African-Americans. It should be noted that slow decay with distance for the ancestral allele of rs562556 led to a significantly high REHH in European-Americans, whereas in African-Americans it was the derived allele that had significantly high REHH. The EHH decay with distance of rs505151 was similar for the ancestral allele and the derived allele.

Calculation of the iHS

The iHS was calculated for SNPs in the *PCSK9* locus. If there is positive selection on a specific allele, the plot of EHH versus physical distance shows that the area under the EHH curve is much greater for the selected allele than for a neutral allele [45]. To test this, we calculated iHS for every SNP in African-Americans and European-Americans separately, treating each SNP in turn as a core SNP (Figure 6). The iHS at each SNP provides a measure of the strength of evidence for selection acting at or near that SNP and an |iHS| > 2.5 the value corresponds to the most extreme 1% of iHS values in the empirical distribution [45]. Thirty out of the 89 common SNPs (MAF $\geq$ 0.10) in African-Americans but none in European-Americans exceeded a threshold of 2.5 for |iHS| (7 SNPs > 2.5, and 23 SNPs < -2.5). The largest iHS value was noted for SNP rs728474 (iHS = 5.07), indicating that unusually long haplotypes carry the ancestral allele of this SNP. The iHS values for SNPs rs2479409, rs562556, and rs505151 were -0.897, 2.261, and -0.876 in African-Americans, and -0.698 and 2.208 for rs2479409 and rs562556 in European-Americans.

Discussion

Resolving the underlying allelic architecture and searching for signatures of natural selection in the human genome has been hailed as an attractive indirect strategy for implementation and interpretation of genetic analyses of drug response (e.g., statins) and complex disorders (e.g., dyslipidemia) [50,51]. The relatively high frequency (~2%) of 'loss-of-function' non-sense mutations in *PCSK9* in Africans, with large effects on plasma cholesterol levels, has fueled speculation that natural selection may have shaped variation in *PCSK9* [17]. Our study of molecular population genetics of *PCSK9* in 24 African-Americans and 23 European-Americans reveals signatures of positive selection in the gain-of-function mutations of *PCSK9*. The results of our study provide insight into the evolutionary history of mutations in *PCSK9* that contribute to inter-individual variation in plasma LDL cholesterol levels in the general population.

A series of statistical tests for evolutionary neutrality was performed to detect population adaptations which may have occurred in more recent evolutionary times (i.e., tests based on nucleotide diversity, population differentiation, and REHH). The information based on Tajima's D and Fay and Wu's H (P < 0.05) in the 5′ flanking regions (Table 1) suggested positive selection in the ancestral allele (H is significantly positive) [48]. Additionally, a significantly greater F_{ST} was noted in a nonsynonymous SNP in the carboxyl-terminal domain (rs505151, E670G, gain-of-function mutation) (F_{ST} = 0.309) (Figure 2), indicating possible positive selection.

REHH is a comparison of the EHH of one haplotype against the EHH of other haplotypes in the same sample [40]. A high REHH (>1.0) indicates that a haplotype displays increased

homozygosity at greater distances compared with other haplotypes. We determined the statistical significance of the REHH value for the *PCSK9* core SNPs in two ways – by 1,000 coalescent simulations under a calibrated population model and by empirical distribution from HapMap database. Three common potentially functional SNPs (i.e., rs505151, rs2479409, and rs562556) showed an unusually extended homozygosity (Tables 2 and 3,Figure 4), indicating positive selection may have operated on either the ancestral or derived alleles in different populations. We used the empirical distribution of YRI from HapMap database for EHH, but it should be noted that there is an estimated 20~25% European admixture in African-Americans [52]. Since population admixture may mask the signature of selection [53], further work is needed to confirm the signature of positive selection in a sample from an African population.

The E670G cSNP (rs505151) is located in the cysteine-rich C-terminal domain and appears to be involved in regulation of autoprocessing, since deletion of this domain leads to accumulation of processed PCSK9 [54]. A common haplotype containing the E670G variant was associated with plasma LDL cholesterol levels and severity of coronary atherosclerosis in African-Americans and Whites [21]. Evans and Beil [55], in a European population, also noted that the E670G polymorphism is associated with polygenic hypercholesterolemia in men (but not in women). The LRH test indicated that positive selection simultaneously shaped the ancestral allele of SNPs rs505151 (A) and the derived allele of rs562556 (A) in African-Americans. However, in European-Americans, the ancestral allele of rs562556 (G) appeared to be under positive selection. The positive selection on ancestral and derived allele in different populations suggested that molecular function may have adapted to varying environments.

The iHS measures how unusual the haplotypes around a given SNP (i.e., core SNP) are relative to the genome as a whole [45]. In African-Americans, the direction of iHS values for SNPs rs562556 (iHS = 2.261) and rs505151 (iHS = -0.876) were consistent with the results obtained from LRH test. However, the iHS for SNP rs562556 in European-Americans was also positive without a suitable interpretation. The positive iHS value suggested the area of EHH curve under the ancestral allele is greater than that under the derived allele. In the context of the core SNP used in LRH test and iHS, different strategies were used to count the haplotype homozygosity at a given distance. In the LRH test, we simply looked at carefully matched genetic distances (marker homozygosity used as a proxy for genetic distance) [40]. However, in iHS, the integral of the decay of EHH away from a specified core allele until EHH reaches 0.05 was calculated for ancestral and derived allele [45]. It remains unclear which test is statistically more powerful to detect positive selection in an empirical study.

Several recent studies used the HapMap [45,56], Perlegen [56,57], or whole-genome association data sets (i.e., Affymetrix 100 K) [58] to detect the signature of natural selection in the whole genome. No signature of recent positive selection was noted in *PCSK9* in these whole-genome studies. This may be because the coverage of *PCSK9* in these genotyping data is lower, in contrast to the re-sequenced genomic data used in the present study.

The forces driving positive selection on these two gain-of-function mutations remain unclear. In addition to the liver and small intestine, *PCSK9* is transiently expressed during embryonic development in the telencephalon and cerebellum [11,12]. It has been suggested that *PCSK9* has a role in the differentiation of cortical neurons and neural development [12,59], and had measurable proapoptotic effects in cerebellar granule neurons [59]. The close association between differentiation and apoptosis during neurogenesis may confer increased sensitivity to cellular alterations, particularly those affecting the cell cycle and metabolism [60]. Knockout of *PCSK9* results in a lethal phenotype in zebrafish [11] but not in the mouse [26].The role of *PCSK9* in apoptosis of neuronal cell and neurogenesis suggests a novel biological function of *PCSK9* in affecting developmental processes. Whether the effects of *PCSK9* on neuronal cell differentiation and apoptosis confer a selective advantage remains speculative. It is also unclear

whether gain-of-function mutations in *PCSK9* confer novel biochemical and biological functions.

Several evolutionary studies have shown evidence of positive selection on gain-of-function mutations. A striking example is positive selection on gain-of-function mutations in *P53* [61, 62], that appear to lead to distinct novel function in different tumors [61]. Further studies indicated that positive selection for gain-of-function in tumor suppressor genes is an important aspect of tumorigenesis [63]. Another example is a mutation in *FGFR2* (fibroblast growth factor receptor 2), which confer a selective advantage on spermatogonial cells but leads to Apert syndrome (i.e., a characteristic combination of craniosynostosis and syndactyly) [64].

Cohen et al. [17] speculated that the high frequency of nonsense mutations (i.e., loss-of-function mutations) of *PCSK9* in Africans confers a selective advantage. They speculated that *PCSK9* inactivation may interfere with the life cycle of the malaria parasite [32], leading to selective pressure that maintains nonsense mutations in Africans. In the present study, the C679X mutation was a singleton in African-Americans and we did not find a signature of positive selection.

In conclusion, based on the long-range haplotype test, a signature of recent positive selection was noted on the two gain-of-function mutations of *PCSK9*: SNPs rs562556 (I474V) and rs505151 (E670G), which showed differential selective modes among African-Americans and European-Americans. The ancestral allele of SNP rs562556 and the derived allele of SNP rs505151 were under positive selection in African-Americans, whereas the derived allele of SNP rs562556 appeared to be under positive selection in European-Americans. A significant population differentiation was also noted for the allele frequency of SNP rs505151 between African-Americans and European-Americans. Our findings suggest that evolutionary dynamics may underlie the gain-of-function mutations in *PCSK9* and variation in LDL cholesterol metabolism, and thereby influence susceptibility to coronary heart disease, as well as response to statin therapy.

Acknowledgments

We acknowledge the technical support of the Supercomputing Institute of University of Minnesota, Minneapolis and fundings by NIH grants K23-RR17720 and HL75794.

References

1. Maxwell KN, Breslow JL. Adenoviral-mediated expression of Pcsk9 in mice results in a low-density lipoprotein receptor knockout phenotype. Proc Natl Acad Sci U S A 2004;101:7100–7105. [PubMed: 15118091]
2. Park SW, Moon YA, Horton JD. Post-transcriptional regulation of low density lipoprotein receptor protein by proprotein convertase subtilisin/kexin type 9a in mouse liver. J Biol Chem 2004;279:50630–50638. [PubMed: 15385538]
3. Benjannet S, Rhainds D, Essalmani R, Mayne J, Wickham L, Jin W, et al. NARC-1/PCSK9 and its natural mutants: zymogen cleavage and effects on the low density lipoprotein (LDL) receptor and LDL cholesterol. J Biol Chem 2004;279:48865–48875. [PubMed: 15358785]
4. Lalanne F, Lambert G, Amar MJ, Chetiveaux M, Zair Y, Jarnoux AL, et al. Wild-type PCSK9 inhibits LDL clearance but does not affect apoB-containing lipoprotein production in mouse and cultured cells. J Lipid Res 2005;46:1312–1319. [PubMed: 15741654]
5. Ouguerram K, Chetiveaux M, Zair Y, Costet P, Abifadel M, Varret M, et al. Apolipoprotein B100 metabolism in autosomal-dominant hypercholesterolemia related to mutations in PCSK9. Arterioscler Thromb Vasc Biol 2004;24:1448–1453. [PubMed: 15166014]

6. Sun XM, Eden ER, Tosi I, Neuwirth CK, Wile D, Naoumova RP, et al. Evidence for effect of mutant PCSK9 on apolipoprotein B secretion as the cause of unusually severe dominant hypercholesterolaemia. Hum Mol Genet 2005;14:1161–1169. [PubMed: 15772090]
7. Lambert G, Jarnoux AL, Pineau T, Pape O, Chetiveaux M, Laboisse C, et al. Fasting induces hyperlipidemia in mice overexpressing proprotein convertase subtilisin kexin type 9: lack of modulation of very-low-density lipoprotein hepatic output by the low-density lipoprotein receptor. Endocrinology 2006;147:4985–4995. [PubMed: 16794006]
8. Maxwell KN, Breslow JL. Proprotein convertase subtilisin kexin 9: the third locus implicated in autosomal dominant hypercholesterolemia. Curr Opin Lipidol 2005;16:167–172. [PubMed: 15767856]
9. Zhang DW, Lagace TA, Garuti R, Zhao Z, McDonald M, Horton JD, et al. Binding of PCSK9 to EGF-A repeat of LDL receptor decreases receptor recycling and increases degradation. J Biol Chem 2007;282:18602–18612. [PubMed: 17452316]
10. Cunningham D, Danley DE, Geoghegan KF, Griffor MC, Hawkins JL, Subashi TA, et al. Structural and biophysical studies of PCSK9 and its mutants linked to familial hypercholesterolemia. Nat Struct Mol Biol 2007;14:413–419. [PubMed: 17435765]
11. Poirier S, Prat A, Marcinkiewicz E, Paquin J, Chitramuthu BP, Baranowski D, et al. Implication of the proprotein convertase NARC-1/PCSK9 in the development of the nervous system. J Neurochem 2006;98:838–850. [PubMed: 16893422]
12. Seidah NG, Benjannet S, Wickham L, Marcinkiewicz J, Jasmin SB, Stifani S, et al. The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): liver regeneration and neuronal differentiation. Proc Natl Acad Sci U S A 2003;100:928–933. [PubMed: 12552133]
13. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, Devillers M, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat Genet 2003;34:154–156. [PubMed: 12730697]
14. Timms KM, Wagner S, Samuels ME, Forbey K, Goldfine H, Jammulapati S, et al. A mutation in PCSK9 causing autosomal-dominant hypercholesterolemia in a Utah pedigree. Hum Genet 2004;114:349–353. [PubMed: 14727179]
15. Leren TP. Mutations in the PCSK9 gene in Norwegian subjects with autosomal dominant hypercholesterolemia. Clin Genet 2004;65:419–422. [PubMed: 15099351]
16. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 2006;354:1264–1272. [PubMed: 16554528]
17. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet 2005;37:161–165. [PubMed: 15654334]
18. Kotowski IK, Pertsemlidis A, Luke A, Cooper RS, Vega GL, Cohen JC, et al. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. Am J Hum Genet 2006;78:410–422. [PubMed: 16465619]
19. Berge KE, Ose L, Leren TP. Missense mutations in the PCSK9 gene are associated with hypocholesterolemia and possibly increased response to statin therapy. Arterioscler Thromb Vasc Biol 2006;26:1094–1100. [PubMed: 16424354]
20. Shioji K, Mannami T, Kokubo Y, Inamoto N, Takagi S, Goto Y, et al. Genetic variants in PCSK9 affect the cholesterol level in Japanese. J Hum Genet 2004;49:109–114. [PubMed: 14727156]
21. Chen SN, Ballantyne CM, Gotto AM Jr, Tan Y, Willerson JT, Marian AJ. A common PCSK9 haplotype, encompassing the E670G coding single nucleotide polymorphism, is a novel genetic marker for plasma low-density lipoprotein cholesterol levels and severity of coronary atherosclerosis. J Am Coll Cardiol 2005;45:1611–1619. [PubMed: 15893176]
22. Lambert G. Unravelling the functional significance of PCSK9. Curr Opin Lipidol 2007;18:304–309. [PubMed: 17495605]
23. Seidah NG, Prat A. The proprotein convertases are potential targets in the treatment of dyslipidemia. J Mol Med 2007;85:685–696. [PubMed: 17351764]
24. Brown MS, Goldstein JL. Biomedicine. Lowering LDL--not only how low, but how long? Science 2006;311:1721–1723. [PubMed: 16556829]
25. Dubuc G, Chamberland A, Wassef H, Davignon J, Seidah NG, Bernier L, et al. Statins upregulate PCSK9, the gene encoding the proprotein convertase neural apoptosis-regulated convertase-1

implicated in familial hypercholesterolemia. Arterioscler Thromb Vasc Biol 2004;24:1454–1459. [PubMed: 15178557]

26. Rashid S, Curtis DE, Garuti R, Anderson NN, Bashmakov Y, Ho YK, et al. Decreased plasma cholesterol and hypersensitivity to statins in mice lacking Pcsk9. Proc Natl Acad Sci U S A 2005;102:5374–5379. [PubMed: 15805190]
27. Naoumova RP, Tosi I, Patel D, Neuwirth C, Horswell SD, Marais AD, et al. Severe hypercholesterolemia in four British families with the D374Y mutation in the PCSK9 gene: long-term follow-up and treatment response. Arterioscler Thromb Vasc Biol 2005;25:2654–2660. [PubMed: 16224054]
28. Mangravite LM, Thorn CF, Krauss RM. Clinical implications of pharmacogenomics of statin treatment. Pharmacogenomics J 2006;6:360–374. [PubMed: 16550210]
29. Wang QF, Prabhakar S, Wang Q, Moses A, Chanan S, Brown M, et al. Primate-Specific Evolution of an LDLR Enhancer. Genome Biol 2006;7:R68. [PubMed: 16884525]
30. Fagundes NJ, Salzano FM, Batzer MA, Deininger PL, Bonatto SL. Worldwide genetic variation at the 3′-UTR region of the LDLR gene: possible influence of natural selection. Ann Hum Genet 2005;69:389–400. [PubMed: 15996168]
31. Attie AD, Seidah NG. Dual regulation of the LDL receptor--some clarity and new questions. Cell Metab 2005;1:290–292. [PubMed: 16054075]
32. Horton JD, Cohen JC, Hobbs HH. Molecular biology of PCSK9: its role in LDL metabolism. Trends Biochem Sci 2007;32:71–77. [PubMed: 17215125]
33. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 2004;74:106–120. [PubMed: 14681826]
34. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003;100:9440–9445. [PubMed: 12883005]
35. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978–989. [PubMed: 11254454]
36. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. Genome Res 2002;12:1805–1814. [PubMed: 12466284]
37. Kayser M, Brauer S, Stoneking M. A genome scan to detect candidate regions influenced by local natural selection in human populations. Mol Biol Evol 2003;20:893–900. [PubMed: 12717000]
38. Kullo IJ, Ding K. Patterns of population differentiation of candidate genes for cardiovascular disease. BMC Genet 2007;8:48. [PubMed: 17626638]
39. Goudet J. FSTAT (Version 1.2): a computer program to calculate F-Statistics. Journal of Heredity 1995;86:485–486.
40. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature 2002;419:832–837. [PubMed: 12397357]
41. Wang H, Ding K, Zhang Y, Jin L, Kullo IJ, He F. Comparative and evolutionary pharmacogenetics of ABCB1: complex signatures of positive selection on coding and regulatory regions. Pharmacogenet Genomics 2007;17:667–678. [PubMed: 17622943]
42. Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. Calibrating a coalescent simulation of human genome sequence variation. Genome Res 2005;15:1576–1583. [PubMed: 16251467]
43. Helgason A, Palsson S, Thorleifsson G, Grant SF, Emilsson V, Gunnarsdottir S, et al. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. Nat Genet 2007;39:218–225. [PubMed: 17206141]
44. Patin E, Barreiro LB, Sabeti PC, Austerlitz F, Luca F, Sajantila A, et al. Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. Am J Hum Genet 2006;78:423–436. [PubMed: 16416399]
45. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol 2006;4:e72. [PubMed: 16494531]
46. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001;409:928–933. [PubMed: 11237013]

47. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989;123:585–595. [PubMed: 2513255]
48. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. Genetics 2000;155:1405–1413. [PubMed: 10880498]
49. Ding K, Kullo IJ. Molecular evolution of 5′ flanking regions of 87 candidate genes for atherosclerotic cardiovascular disease. Genet Epidemiol 2006;30:557–569. [PubMed: 16799961]
50. Bamshad M, Wooding SP. Signatures of natural selection in the human genome. Nat Rev Genet 2003;4:99–111. [PubMed: 12560807]
51. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage. Science 2006;312:1614–1620. [PubMed: 16778047]
52. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, et al. Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet 1998;63:1839–1851. [PubMed: 9837836]
53. Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, et al. Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol 2004;2:e286. [PubMed: 15361935]
54. Naureckiene S, Ma L, Sreekumar K, Purandare U, Lo CF, Huang Y, et al. Functional characterization of Narc 1, a novel proteinase related to proteinase K. Arch Biochem Biophys 2003;420:55–67. [PubMed: 14622975]
55. Evans D, Beil FU. The E670G SNP in the PCSK9 gene is associated with polygenic hypercholesterolemia in men but not in women. BMC Med Genet 2006;7:66. [PubMed: 16875509]
56. Tang K, Thornton KR, Stoneking M. A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. PLoS Biol 2007;5:e171. [PubMed: 17579516]
57. Wang ET, Kodama G, Baldi P, Moyzis RK. Global landscape of recent inferred Darwinian selection for Homo sapiens. Proc Natl Acad Sci U S A 2006;103:135–140. [PubMed: 16371466]
58. Zhang C, Bailey DK, Awad T, Liu G, Xing G, Cao M, et al. A whole genome long-range haplotype (WGLRH) test for detecting imprints of positive selection in human populations. Bioinformatics 2006;22:2122–2128. [PubMed: 16845142]
59. Bingham B, Shen R, Kotnis S, Lo CF, Ozenberger BA, Ghosh N, et al. Proapoptotic effects of NARC 1 (= PCSK9), the gene encoding a novel serine proteinase. Cytometry A 2006;69:1123–1131. [PubMed: 17051583]
60. Furutani-Seiki M, Jiang YJ, Brand M, Heisenberg CP, Houart C, Beuchle D, et al. Neural degeneration mutants in the zebrafish, Danio rerio. Development 1996;123:229–239. [PubMed: 9007243]
61. Koonin EV, Rogozin IB, Glazko GV. p53 gain-of-function: tumor biology and bioinformatics come together. Cell Cycle 2005;4:686–688. [PubMed: 15846083]
62. Glazko GV, Koonin EV, Rogozin IB. Mutation hotspots in the p53 gene in tumors of different origin: correlation with evolutionary conservation and signs of positive selection. Biochim Biophys Acta 2004;1679:95–106. [PubMed: 15297143]
63. Glazko GV, Babenko VN, Koonin EV, Rogozin IB. Mutational hotspots in the TP53 gene and, possibly, other tumor suppressors evolve by positive selection. Biol Direct 2006;1:4. [PubMed: 16542006]
64. Goriely A, McVean GA, van Pelt AM, O'Rourke AW, Wall SA, de Rooij DG, et al. Gain-of-function amino acid substitutions drive positive selection of FGFR2 mutations in human spermatogonia. Proc Natl Acad Sci U S A 2005;102:6051–6056. [PubMed: 15840724]
65. Yue P, Averna M, Lin X, Schonfeld G. The c.43_44insCTG variation in PCSK9 is associated with low plasma LDL-cholesterol in a Caucasian population. Hum Mutat 2006;27:460–466. [PubMed: 16619215]

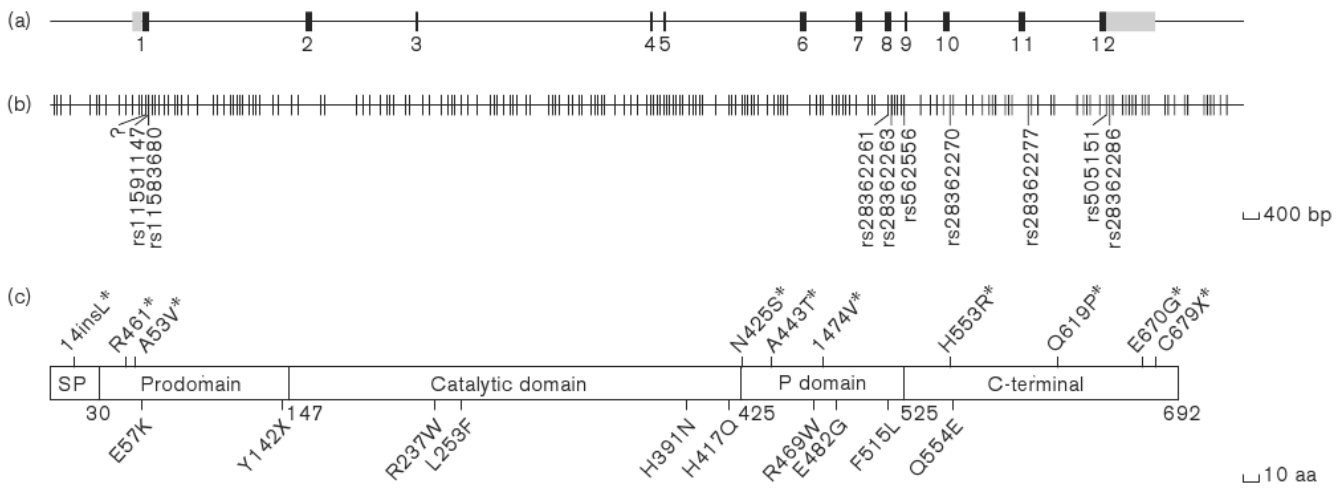

Figure 1.

(a) Structure of *PCSK9*. There are 12 exons; coding regions and untranslated regions are represented by 'black' and 'gray' boxes, respectively. (b) Distribution of variants found in the SeattleSNPs database. There are 265 variants including SNPs and in/dels. Reference SNP numbers show the 8 non-synonymous SNPs, and one insertion-frame codon is labeled '?'. Known nonsynonymous sequence variations in *PCSK9* identified by Kotowski et al. [18]. 14ins Leu was identified by Yue et al. [65]. (c) There are five domains in the PCSK9 protein [3,12, 54]: 1) a signal peptide (SP) (1~30 aa), 2) a prodomain (31 ~ 147 aa), 3) a catalytic domain (148 ~ 425 aa), 4) a putative P domain (426 ~ 525 aa), and 5) a C-terminal (526 ~ 692 aa).

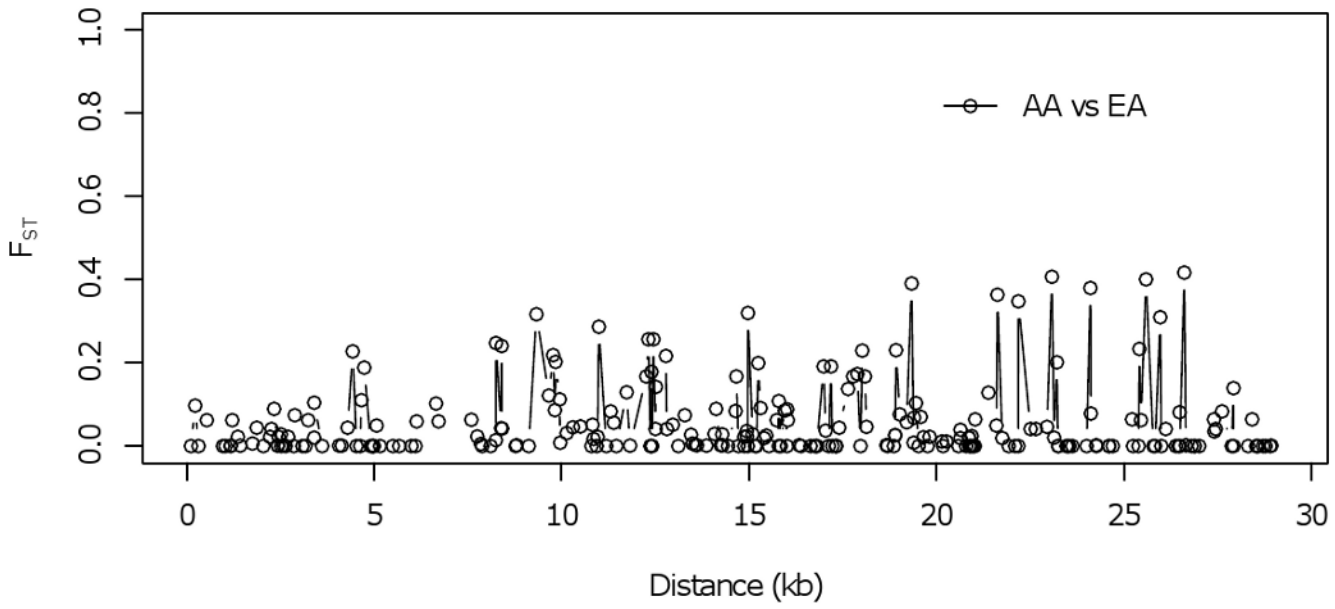

Figure 2.
Distribution of F_{ST} along *PCSK9*. F_{ST} was calculated using SeattleSNPs data. AA, African-Americans; EA, European-Americans

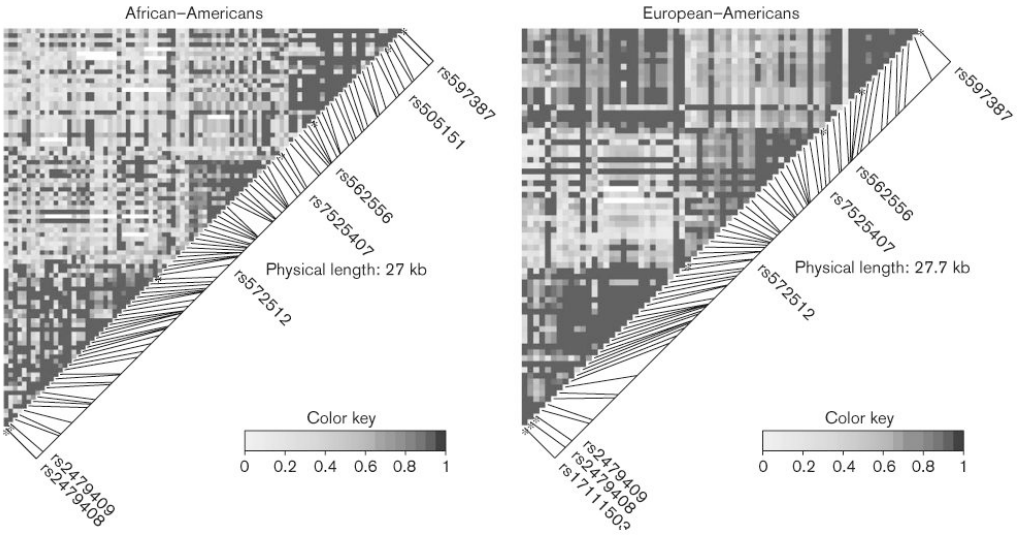

Figure 3.

LD patterns (D′) of common SNPs (minor allele frequency > 0.10 at least in one population) along *PCSK9* in African-Americans and European-Americans. Common non-synonymous SNPs (rs562556 and rs505151) and regulatory SNPs (rs17111503, rs2479408, and rs2479409) are shown. SNPs rs572512, rs7525407, and rs597387 were used to delineate the boundary of the 3 haplotype blocks in European-Americans.

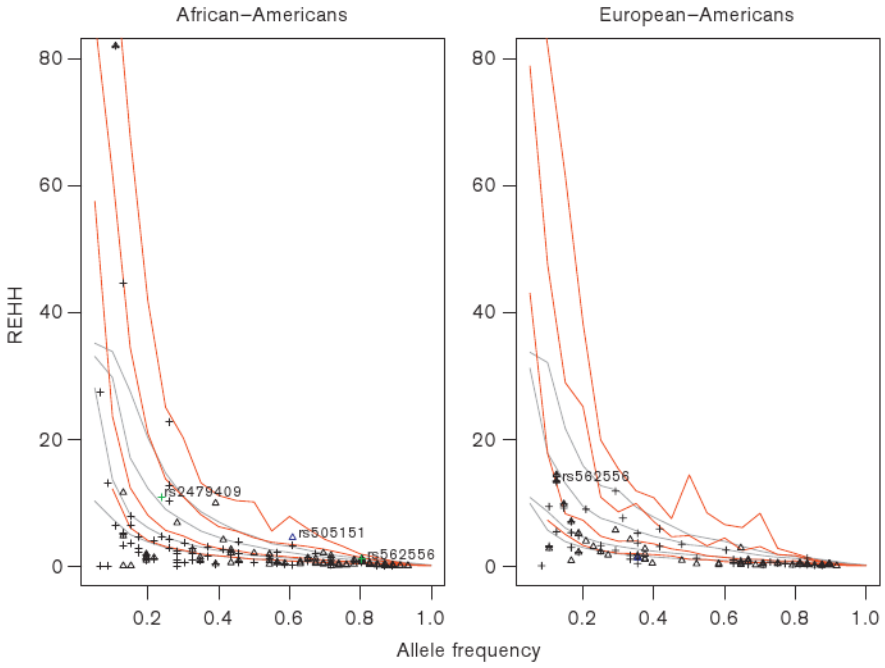

Figure 4.

Relative extended haplotype homozygosity (REHH) tests for common SNPs in *PCSK9* (REHH versus allele frequency) based on both empirical distribution and theoretical distribution in African-Americans and European-Americans. The points of REHH against allele frequency were calculated from phased data of chromosome 7 in HapMap database, as well as simulated data, using the 'sweep' program (see Methods section, data points not shown in the current figure). The 50^{th}, 75^{th}, 95^{th}, and 99^{th} percentile lines of the distribution of empirical data or simulation data were obtained by binning the data by allele frequency into 20 bins of equal size with intervals of 5% (from bottom to upper): gray, empirical data; red, simulated data. The triangles and circles indicate the REHH against frequency of the ancestral and derived alleles, respectively. SNPs rs2479409, rs562556 and rs505151 are shown in green, blue, and green, respectively.

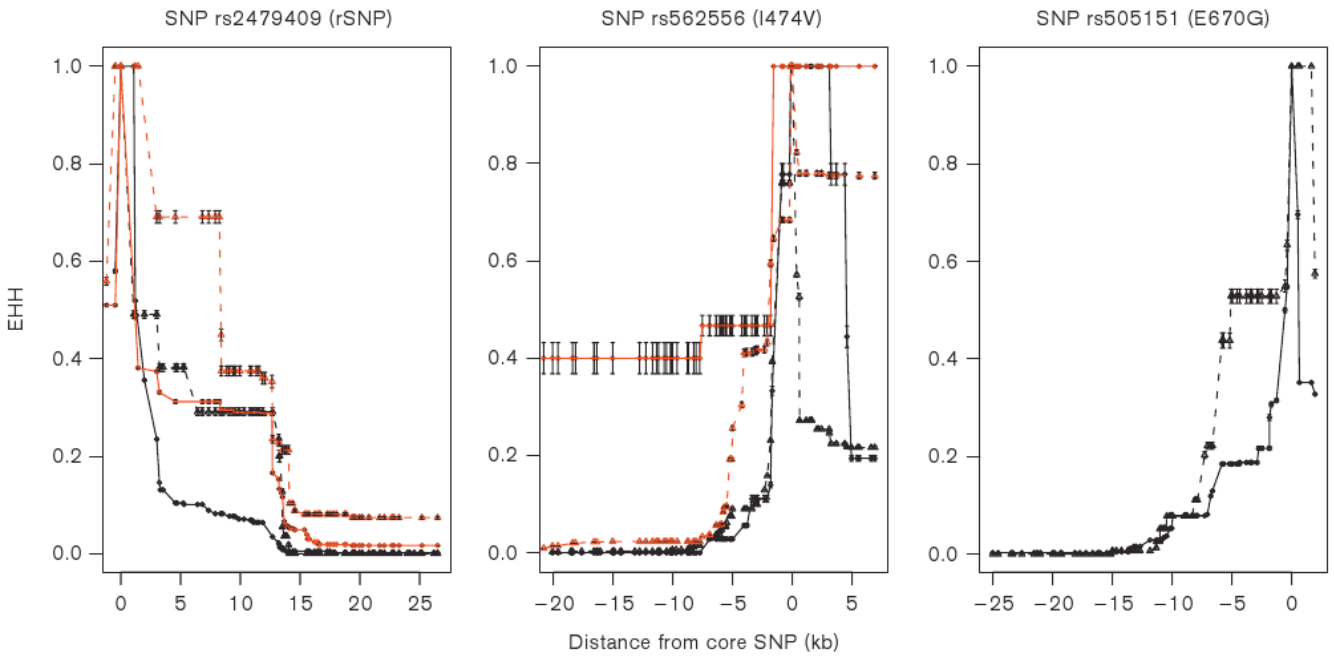

Figure 5.

Extended haplotype homozygosity (EHH) versus physical distance for 3 core SNPs (two nonsynonymous and one regulatory) in *PCSK9*: SNP rs2479409 is in the regulatory region, and SNPs rs562556 (I474V) and rs505151 (E670G) are in the coding regions. In each plot, black represents African-Americans; red, European-Americans. The solid and broken lines indicate the ancestral and derived allele, respectively. SNP rs505151 is not a common SNP in European-Americans.

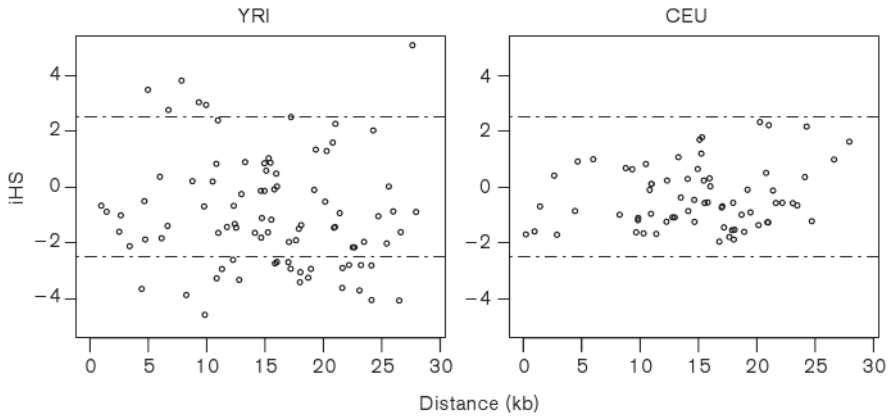

Figure 6.
The pattern of integral haplotype score (iHS) along the *PCSK9* genomic sequence. The horizontal lines represent iHS = 2.5 and -2.5, respectively.

Table 1

Summary statistics for *PCSK9* derived from the SeattleSNPs resequenced data (indels are not included)

Measures	African-Americans			European-Americans		
	Coding (2,079 bp)	**5′ FR (2,290 bp)**	**Overall (29,276 bp)**	**Coding (2,079 bp)**	**5′ FR (2,290 bp)**	**Overall (29,276 bp)**
SS (singletons)[a]	14 (7)	14 (7)	229 (86)	8 (5)	6 (1)	125 (27)
π	9.13	7.44	15.06	4.08	6.98	10.31
θ_W	15.23	13.87	17.84	8.78	5.96	9.81
θ_H	8.15	1.21	15.88	7.70	2.34	17.99
Tajima's D	-1.226	-1.425	-0.572	-1.489	0.442	0.002
Prob[b]	0.182	0.106	0.316	0.090	0.551	0.959
Prob[c]	0.270	0.164	0.561	0.133	0.469	0.644
Fay and Wu's H	0.98	6.34	-0.82	-3.62	4.64	-7.68
Prob[b]	0.717	0.000*	0.799	0.059	0.000*	0.332
Prob[c]	0.762	0.000*	0.766	0.067	0.000*	0.301

5′ FR, 5′ flanking regions; π, the average number of pairwise differences per site between two sequences chosen at random from a sample of sequences; θ_W, the proportion of segregating sites in a sample, corrected for the size of the sample; θ_H, an estimator of θ weighted by the homozygosity of the derived variants

[a] SS, segregating sites

[b] The reference mRNA sequence of *PCSK9* in the SeattleSNPs database includes 2,079 bp with an insertion polymorphism of 'CTG' in its 5′ terminal, compared with NM_174936 from NCBI.

[c] *P* values were calculated from coalescent simulations under a standard neutral model (*Prob*[b]) and a population structure model (*Prob*[c]) [49], given the recombination rate of 1.3 cM/MB in the *PCSK9* locus (D1S2652).

Table 2

The relative extended haplotype homozygosity (REHH) test for common SNPs (MAF ≥ 0.10) in African-Americans based on resequencing data from SeattleSNPs

SNPs (allele)	Frequency	REHH	*P* value	
			simulation	empirical
Ancestral allele				
rs728474 (G)	0.109	82.000	0.0057	0.0000
rs499718 (C)	0.391	9.882	0.0158	0.0011
rs505151 (A)	0.609	4.452	0.0277	0.0001
rs12067569 (T)	0.696	2.018	0.0715	0.0072
rs11206517 (T)	0.717	1.625	0.0731	0.0044
rs6656066 (G)	0.804	0.919	0.0443	0.0162
rs533375 (G)	0.804	1.135	0.0184	0.0038
rs585131 (T)	0.804	1.081	0.0222	0.0061
rs540796 (G)	0.804	1.081	0.0222	0.0061
Derived allele				
rs28362278 (T)	0.109	82.000	0.0057	0.0000
NA (A)	0.130	44.571	0.0303	0.0001
rs28362288 (C)	0.109	82.000	0.0057	0.0000
rs11206513 (T)	0.261	22.667	0.0071	0.0007
rs7530425 (T)	0.261	12.750	0.0326	0.0060
rs2479409 (G)	0.239	10.818	0.0856	0.0269
rs10888897 (C)	0.261	10.200	0.0558	0.0126
rs499883 (G)	0.609	3.238	0.0452	0.0020
rs2483205 (T)	0.674	1.919	0.0786	0.0097
rs534347 (T)	0.804	1.135	0.0184	0.0037
rs562556 (A)	0.804	1.081	0.0222	0.0061

Table 3

The relative extended haplotype homozygosity (REHH) test for common SNPs (MAF ≥ 0.10) in European-Americans based on resequencing data from SeattleSNPs

SNPs (allele)	Frequency	REHH	*P* value	
			simulation	empirical
Ancestral allele				
rs534347 (C)	0.125	13.393	0.1385	0.0463
rs562556 (G)	0.125	14.350	0.1320	0.0370
rs631220 (A)	0.125	14.350	0.1320	0.0370
rs7552841 (C)	0.649	2.827	0.0407	0.0148
Derived allele				
rs533375 (A)	0.125	13.393	0.1385	0.0463
rs634272 (T)	0.208	14.350	0.0942	0.0316
rs557435 (A)	0.125	13.240	0.1385	0.0490
rs585131 (C)	0.125	14.350	0.1320	0.0370
rs540796 (A)	0.125	14.350	0.1320	0.0370
rs643257 (C)	0.125	14.350	0.1320	0.0370
rs639750 (G)	0.292	11.769	0.0358	0.0106
rs613855 (G)	0.479	3.388	0.1292	0.0472
rs624612 (G)	0.417	5.826	0.0187	0.0131
rs625619 (A)	0.604	2.474	0.0492	0.0265
rs521662 (C)	0.313	7.543	0.0705	0.0194
rs499883 (G)	0.354	5.128	0.1290	0.0356