A common allele on chromosome 9 associated with coronary heart disease

Ruth McPherson[1*], Alexander Pertsemlidis[2*], Nihan Kavaslar[1], Alexandre Stewart[1], Robert Roberts[1], David R. Cox[3], David Hinds[3], Len Pennachio[4], Anne Tybjaerg-Hansen[5], Aaron R. Folsom[6], Eric Boerwinkle[7], Helen H. Hobbs[2,9] Jonathan C. Cohen[2,8]

[1]Division of Cardiology, University of Ottawa Heart Institute, Ottawa, Canada, K1Y4W7; [2]Donald W. Reynolds Cardiovascular Clinical Research Center and the Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX, 75390; [3]Perlegen Sciences, Mountain View, California; 94043; [4]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720 USA & U.S. Department of Energy Joint Genome Institute, Walnut Creek, California 94598 USA; [5]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen, DK-2100, Denmark; [6]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, 55454; [7]Human Genetics Center and Institute for Molecular Medicine, University of Texas Health Science Center, Houston, TX, 77030; [8]Center for Human Nutrition and the [9]Howard Hughes Medical Institute at the University of Texas Southwestern Medical Center, Dallas, TX. 75390

*Ruth McPherson and Alexander Pertsemlidis contributed equally to this work.

Running Head: A common allele associated with CHD

[†]To whom correspondence should be addressed:

E-mail: jonathan.cohen@utsouthwestern.edu or rmcphers@ottawaheart.ca

One sentence summary:

An allele on chromosome 9 is systematically associated with a 20-40% increased risk of coronary heart disease.

Coronary heart disease (CHD) is a major cause of death in Western countries. Here we used genome-wide association scanning to identify a 58 kb interval on chromosome 9 that was consistently associated with CHD in six independent samples. The interval contains no annotated genes and is not associated with established CHD risk factors such as plasma lipoproteins, hypertension or diabetes. Homozygotes for the risk allele comprise 20-25% of Caucasians and have a ~30-40% increased risk of CHD. These data indicate that the susceptibility allele acts through a novel mechanism to increase CHD risk in a large fraction of the population.

Coronary heart disease (CHD) is the single largest cause of death worldwide (*1, 2*).
Numerous risk factors influence the development and severity of CHD, such as
hypercholesterolemia, smoking, hypertension and diabetes (*3*). Several of these risk factors
have a major heritable component, but the nature of the DNA sequence variations that confer
cardiovascular risk remain largely unknown. Mutations responsible for Mendelian forms of
hypercholesterolemia greatly increase the risk of coronary atherosclerosis, but these diseases
account for only a small fraction of individuals with heart disease (*4*). Single nucleotide
polymorphism (SNP) association studies (*3*) and genome-wide linkage scans have reported
common alleles associated with CHD risk (*5-10*), but none has been validated by replication
in multiple cohorts.

To identify sequence variants associated with CHD, we undertook a genome-wide
association study using 100,000 SNPs. To minimize false positive associations, without
unduly sacrificing statistical power, we used an iterative case-control design comprising three
sequential comparisons between cases and controls that were performed at a nominal
significance threshold of P<0.025 (Figure 1). Assuming independence, the probability of any
single SNP achieving this level of significance in all three studies with the associations being
in the same direction was 3.9×10^{-6} ($0.025^3 \times 0.5^2$). Thus, none of the 100,000 SNPs would
be expected by chance to replicate consistently in all three comparisons. For the initial
genome-wide scan, cases and controls were white men and women recruited from Ottawa,
Canada to participate in the Ottawa Heart Study (OHS). Cases had symptomatic CHD with
documented onset before the age of 60 years and culminating in coronary artery
revascularization (bypass grafting or percutaneous coronary intervention). To limit
confounding by known major risk factors that strongly predispose to premature CHD,
individuals with diabetes, or with plasma cholesterol levels consistent with monogenic
hypercholesterolemia (>280 mg/dL) were excluded. Controls were healthy white men (>65

y) and women (>70 y) recruited at the same institution as the cases who had no symptoms or history of CHD.

Custom oligonucleotide arrays (*11*) were used to assay 100,000 SNPs arranged at approximately 30 kb intervals throughout the genome in 322 cases and 312 controls (OHS-1). Of these, 9,636 SNPs deviated strongly from Hardy-Weinberg equilibrium (P<0.001) or did not meet quality-control criteria (*11*) and 17,500 were sub-polymorphic (minor allele frequency <1%) in the sample. The remaining 72,864 SNPs were entered into the analysis and 2,586 were associated with CHD in OHS-1 at a nominal significance threshold of 0.025 (Table 1). These 2,586 SNPs were genotyped in a second group of 311 cases and 326 controls (OHS-2), selected from the same population using the same criteria as OHS-1. Of these, 50 were associated with CHD at a nominal significance threshold of 0.025, with the same direction of effect.

To limit attrition of true positive associations due to inadequate statistical power, the third case-control comparison was performed in the Atherosclerosis Risk in Communities (ARIC) study, a large prospective study of CHD risk that enrolled and followed 11,478 European-Americans (*12*). For this analysis, cases were defined as individuals with a documented CHD event (defined as myocardial infarction, coronary artery revascularization, or coronary death) during the 15 yr follow-up period of the study; individuals with prevalent disease at the baseline visit were excluded. The study sample delineated by these criteria provides > 90% power to detect common alleles (minor allele frequency >0.1) that differ in frequency by 0.06 or more between cases and controls. Nonetheless, only two of the 50 SNPs identified in the Ottawa cohorts were significantly associated with incident CHD in the ARIC population (Table 1). These two SNPs, rs10757274 and rs2383206, were located within 20 kb of each other on chromosome 9 and were in strong linkage disequilibrium (r^2=0.89). Both SNPs involve the substitution of G for A. At both loci the A and G alleles were present at

almost equal frequencies among the controls in the three samples. The presence of a G allele was associated with an increased risk of CHD in all three groups (OHS-1, OHS-2, and ARIC) (Table 2).

To validate the association between rs10757274 and rs2383206 and CHD, both SNPs were assayed in three additional independent cohorts: the Copenhagen City Heart Study (CCHS), a prospective study of ischemic heart disease in 10,578 Danish men and women (*13*); the Dallas Heart Study (DHS), a population-based probability sample of Dallas County residents (*14*); and a third group of 647 cases and 847 controls from the Ottawa Heart Study population who had not been included in the genome-wide scan or in the first replicate analysis (OHS-3). In the CCHS, cases were individuals who experienced an ischemic cardiovascular event during the 20 yr follow-up period while controls were those who did not develop incident disease. In the DHS, electron-beam computer tomography was used to measure coronary artery calcium (CAC), an index of coronary atherosclerosis (*15*). Each individual underwent two consecutive scans. The distribution of CAC scores is extremely skewed and inter-scan variability is high for scores below 10 Agatston units (*16*), therefore we excluded individuals with CAC scores between 2 and 10 units and divided the population into controls (CAC scores of ≤ 2 units, n=575) and cases (those with CAC scores ≥ 10 units, n=166), as previously described (*16*). In OHS-3, cases had documented CHD before the age of 55 (men) or 65 (women), whereas controls were men and women aged > 70 years who did not have symptoms of CHD. In all three validation studies, both SNPs were significantly associated with CHD (Table 2). These data further support the association between the G alleles of rs10757274 and rs2383206 and CHD.

The magnitude of incident CHD associated with the G allele was determined by Cox proportional-hazards modeling in the ARIC and CCHS cohorts. The hazard ratios associated with the risk alleles were comparable in the two populations, and indicated a graded increase

in risk from noncarriers to heterozygotes to homozygotes (Table 3). The two SNPs (rs10757274 and rs2383206) define an allele that is associated with a ~15-20% increase in risk in the 50% of individuals who are heterozygous for the allele and ~30-40 % increase in CHD in the 25% of individuals of European descent who are homozygotes. The population attributable risk associated with the risk allele was 12.5-15% in the ARIC population and 10-13% in the CCHS cohort.

The mechanistic basis for the association of the allele defined by rs10757274 and rs2383206 and CHD is not known. The risk allele may promote CHD by increasing the development of atherosclerotic plaques in the coronary arteries, by promoting thrombogenesis, or by increasing the tendency of atherosclerotic plaques to rupture. The finding that the risk allele was associated with increased coronary artery calcification in the DHS suggests that it promotes CHD by increasing the atherosclerotic burden. This notion is supported by the observation that the prevalence of the risk allele was greatest among cases in OHS-1, who were ascertained for severe premature atherosclerosis rather than for cardiac events. To determine if the atherogenic effects of the risk allele are mediated via established risk factors for coronary artery atherosclerosis, we tested for association between the risk allele and intermediate phenotypes in the ARIC and CCHS populations. In both cohorts, the risk allele was not significantly associated with systolic or diastolic blood pressure, fasting blood glucose or diabetes, or plasma levels of lipids or lipoprotein-cholesterol (Supplementary Tables 1 and 2). The risk allele was associated with a slight but statistically significant decrease in body mass index in the ARIC cohort, but not in the CCHS population. Thus, the effect of the chromosome 9 risk allele on CHD is not mediated by any of the established major risk factors for the disease.

To fine-map the interval associated with CHD, SNPs spaced at ~5 kb intervals across the region extending 175 kb upstream and downstream of rs10757274 and rs2383206 were

assayed in 500 cases and 500 controls from OHS-1 and OHS-2. Eight additional SNPs at the locus spanning a 58 kb region (extending from 22,062,301 to 22,120,389) were significantly associated with CHD (Figure 2). All eight were in strong linkage disequilibrium with each other and with rs10757274 and rs2383206. The region demarcated by these SNPs was flanked on both sides by ~50 kb regions in which none of the 30 SNPs tested were associated with CHD, even at a nominal significance threshold of 0.025. Two of 65 SNPs in the 350 kb region surrounding the 58 kb risk locus were associated with CHD at the nominal significance threshold, but neither of these SNPs was in linkage disequilibrium with rs10757274 and rs2383206. These data indicate that the risk allele comprises a single haplotype that spans ~58 kb.

Inspection of the UCSC Genome Browser and BLAST searches against the NCBI nr nucleotide sequence database revealed no annotated genes or microRNAs within the 58 kb interval. Thus, the nucleotide changes associated with the risk allele do not alter the coding sequence of a known gene. A number of spliced ESTs map within the interval, but none contain open reading frames that extend more than a few amino acids. As a first step towards identifying functional elements responsible for the observed association, we sequenced the 58 kb interval in two homozygotes for the risk allele, and in one individual who was homozygous for the reference allele. Of the 66 polymorphisms (SNPs plus small insertions or deletions) identified in the 6 alleles that were sequenced, 35 were specific to the risk allele (Supplementary Table 3). Only one of these variants, a copy number variation in a run of 9 consecutive CAT repeats (extending from nucleotide 22110787 to 22110814, NCBI build 36.1) mapped to a spliced transcript (BI765545) that appears to be part of a large noncoding RNA, DQ485453 (*17*). PCR amplification of cDNAs confirmed expression of the transcript in placenta and transformed lymphocytes (data not shown). Whereas the function of

DQ485453 is not known, it is possible that variation in the expression or function of this transcript may be associated with risk of CHD.

Alternatively the risk allele may alter the sequence of a regulatory element that affects the expression of a gene(s) located outside of the 58 kb interval. Cross-species sequence alignments using VISTA revealed several segments within the 58 kb interval that are conserved among humans, dogs and mice that may contain such regulatory elements (Supplementary Figure 1). It is also possible that the risk allele extends beyond the 58 kb interval defined in this study, and that the functional sequence variants that confer risk of CHD are located outside of the interval. Resequencing the coding regions of the two genes most proximal to the risk locus, CDKN2A and CDKN2B, which lie 70-100 kb telomeric to the interval (Figure 2), revealed only a single variant (A158S in CDKN2A) that was present in 6 of the 96 apparently healthy white men and women examined. The localization of the risk locus to a region devoid of known genes implicates a previously unrecognized gene that can substantially affect CHD risk independently of the established CHD risk factors. Further studies will be required to elucidate the mechanism by which allelic variation at the locus modulates risk of coronary heart disease.

The greater haplotype diversity found in populations of African descent (*18*) may be useful for fine localization of functional sequence variants responsible for genetic associations. Comparison of the Yoruba and CEPH data from the HapMap revealed striking ethnic differences in the frequencies of alleles in the risk interval (Supplementary Table 4). Of the 10 alleles that were significantly associated with CHD in whites, 3 were virtually absent from the Yoruba population, and 6 others significantly less common. Both rs10757274 and rs2383206 were present at appreciable frequencies among African-Americans in ARIC and DHS, but neither SNP was associated with CHD in either population (Supplementary Table 5). It is possible that the apparent ethnic differences in association

between these SNPs and CHD in ARIC are a spurious consequence of differences statistical power: there are significantly fewer African-Americans than whites in the ARIC study. However inadequate statistical power cannot explain the ethnic differences observed in DHS, since African-Americans are the largest group in this study. Accordingly, it seems more likely that the functional sequence variants associated with the risk allele in whites are less common in African-Americans. This notion is consistent with our finding that the frequencies of several alleles associated with CHD risk factors differ widely between these ethnic groups (*19-21*). Comprehensive analysis of the locus in African-Africans may allow further refinement of the risk interval.

The results of this study illustrate both the perils and the promise of whole-genome association. The initial scan and the first replicate screen both generated substantially more SNPs that achieved the pre-specified significance threshold than would be predicted by chance alone, as indicated by permutation testing. Yet only two of these SNPs (comprising one allele) survived further replication, despite the use of a large sample (i.e. ARIC) with high statistical power. This finding highlights the necessity for adequate replication to protect against artifacts that may occur due to population stratification, multiple testing, or other factors to which whole-genome association studies are particularly susceptible. The consistent replication of the chromosome 9 risk allele in six independent study samples indicates that the approach can be productively applied to conditions as complex as CHD, which is known to be influenced by a variety of environmental and genetic factors (*22*). Analysis of 50 randomly selected regions of 500 kb each indicated that the 72,864 informative SNPs used in the initial scan provided 30-40% of the power that would be obtained by assaying all Phase II Hapmap SNPs. Therefore scans with denser SNP panels may reveal further loci associated with CHD risk.

References

1.   C. J. Murray, A. D. Lopez, *Lancet* **349**, 1436 (1997).

2.   C. D. Mathers, D. Loncar, *PLoS Med* **3**, e442 (2006).

3.   A. J. Lusis, R. Mar, P. Pajukanta, *Annu Rev Genomics Hum Genet* **5**, 189 (2004).

4.   D. J. Rader, J. Cohen, H. H. Hobbs, *J Clin Invest* **111**, 1795 (2003).

5.   P. Pajukanta *et al.*, *Am J Hum Genet* **67**, 1481 (2000).

6.   D. Shiffman *et al.*, *Am J Hum Genet* **77**, 596 (2005).

7.   S. Francke *et al.*, *Hum Mol Genet* **10**, 2751 (2001).

8.   E. R. Hauser *et al.*, *Am J Hum Genet* **75**, 436 (2004).

9.   U. Broeckel *et al.*, *Nat Genet* **30**, 210 (2002).

10.   S. B. Harrap *et al.*, *Arterioscler Thromb Vasc Biol* **22**, 874 (2002).

11.   S. F. Saccone *et al.*, *Hum Mol Genet* **16**, 36 (2007).

12.   T. A. S. Investigators, *Am J Epidemiol* **129**, 687 (1989).

13.   P. Schnohr, G. Jensen, H. Scharling, M. Appleyard, *European Heart Journal* **3, Suppl H.**, H1 (2001).

14.   R. G. Victor *et al.*, *Am J Cardiol* **93**, 1473 (2004).

15.   A. S. Agatston *et al.*, *J Am Coll Cardiol* **15**, 827 (1990).

16.   T. Jain *et al.*, *J. Am Coll Cardiol* **44**, 1011 (2004).

17.   E. Pasmant, Laurendeau,I., Heron,D., Vidaud,M.,Vidaud,D., Bieche,I., *Cancer Res* **In Press** (2007).

18.   D. A. Hinds *et al.*, *Science* **307**, 1072 (2005).

19.   J. Cohen *et al.*, *Proc Natl Acad Sci USA* **103**, 1810 (2006).

20.   J. C. Cohen, E. Boerwinkle, T. H. Mosley, H. HH, *N. Engl. J. Med.* **354**, 34 (2006).

21.   S. Romeo *et al.*, *Nat Genet* (2007).

22.   National Cholesterol Education Program (NCEP) Expert Panel on Detection.

        Circulation **106**, 3143 (2002).

Table 1. SNP associations in a genome-wide association scan in OHS-1 and replication studies in OHS-2 and ARIC.

Values in the table are numbers of SNPs. P-values for the number of SNPs observed relative to the number expected were derived empirically from 1000 permutations in which case-control status was randomized. SNPs in the ARIC cohort were tested for association with incident CHD. Individuals with prevalent CHD at baseline were excluded.

OHS, Ottawa Heart Study; ARIC, Atherosclerosis Risk in Communities Study; SNP, single nucleotide polymorphism

Cohort	SNPs Assayed	SNPs with P<0.025 Observed	SNPs with P<0.025 Expected	Observed-Expected	P
OHS-1	72,864	2,586	2,066	520	<0.001
OHS-2	2,083	50	26	24	<0.001
ARIC	50	2	0	2	NA

.

Table 2. Association between SNPs rs10757274 and rs2383206 and CHD.

Values are numbers of individuals in each genotype group. P-values were calculated by Chi-Square tests on allele counts.

OHS, Ottawa Heart Study, ARIC, Atherosclerosis Risk in Communities Study; CCHS, Copenhagen City Heart Study, DHS, Dallas Heart Study,

SNP, single nucleotide polymorphism.

Cohort	*rs10757274*							*rs2383206*						
	Controls			Cases				Controls			Cases			
	AA	AG	GG	AA	AG	GG	P	AA	AG	GG	AA	AG	GG	P
OHS-1	85	149	78	49	148	125	4.8×10^{-6}	77	147	88	45	140	137	8.1×10^{-6}
OHS-2	85	161	80	56	140	108	0.002	80	160	86	50	141	113	0.001
ARIC	2063	3822	1858	230	525	282	0.001	2140	4161	2231	230	600	324	0.001
CCHS	2752	4543	1758	393	792	340	0.0001	2489	4583	1981	372	782	371	0.005
DHS	161	283	131	28	92	46	0.025	144	284	147	25	91	50	0.023
OHS-3	228	418	201	121	333	193	0.0001	197	416	229	115	327	209	0.004

Table 3. Risk of CHD as a function of rs10757274 and rs2383206 in the Atherosclerosis Risk in Communities Study and the Copenhagen City

Heart Study.

		Atherosclerosis Risk in Communities Study					*Copenhagen City Heart Study*			
		Number of Events					**Number of Events**			
	n (%)	**Observed**	**Expected[1]**	**Incidence[2]**	**Hazard Ratio**	**n (%)**	**Observed**	**Expected**	**Incidence**	**Hazard Ratio**
rs10757274										
AA	2,293 (26)	255[3]	295	79 (70-89)	1	3,145 (30)	393[5]	473	61(55-68)	1
AG	4,347 (50)	564	553	93 (86-101)	1.18 (1.02-1.37)	5,335 (50)	792	755	73 (68-79)	1.26 (1.12-1.42)
GG	2,140 (24)	298	269	101 (90-114)	1.29 (1.09-1.52)	2,098 (20)	340	296	80 (72-89)	1.38 (1.19-1.60)
rs2383206										
AA	2,370 (25)	259[4]	310	78 (69-88)	1	2,861 (27)	372[6]	425	64 (58-71)	1
AG	4,761 (49)	643	610	97 (90-105)	1.26 (1.09-1.46)	5,365 (51)	782	772	72 (67-77)	1.16 (1.02-1.31)
GG	2,555 (26)	345	327	97 (88-108)	1.26 (1.07-1.48)	2,352 (22)	371	327	78 (71-87)	1.29 (1.12-1.50)

[1]Based on the Log-rank test. [2]Incidence rate measured in number of events per 10,000 person years of follow-up. [3]p<0.0111

[4]p<0.0041. [5]p<0.00001. [6]p<0.005

Figure Legends

Figure 1. Study design for identification and validation of sequence variants associated with coronary heart disease.

Figure 2. Fine mapping of the genomic interval on chromosome 9 associated with coronary heart disease. SNPs spaced ~5 kb apart in the interval extending 175 kb upstream and downstream of rs10757274 and rs2383206 were assayed in 500 cases and 500 controls from the Ottawa Heart Study population. Bars represent P values (determined using Chi-Square tests) for differences in allele frequency between cases and controls. Arrowheads indicate rs10757274 and rs2383206. The shaded box represents the risk interval. The linkage disequilibrium map indicates pairwise r^2 values. Blocks are shaded on a continuous scale where white represents an r^2 of 0 and black represents an r^2 of 1.

Figure 1

Screening

Genome-wide Association Scan (75,000 SNPs/person)
Ottawa Heart Study-1 (OHS-1)
322 Cases : 312 controls

Replicate Association Study 1: SNPs with P <0.025
Ottawa Heart Study-2 (OHS-2)
311 cases : 326 controls

Replicate Association Study 2: SNPs with P <0.025
Atherosclerosis Risk in Communities Study (ARIC)
1,347 cases : 9,054 controls

rs10757274 and rs2383206

Validation

Copenhagen City Heart Study (CCHS)	*Dallas Heart Study (DHS)*	*Ottawa Heart Study-3 (OHS-3)*
1,525 cases	**154 cases**	**618 cases**
9,053 controls	**527controls**	**782 controls**

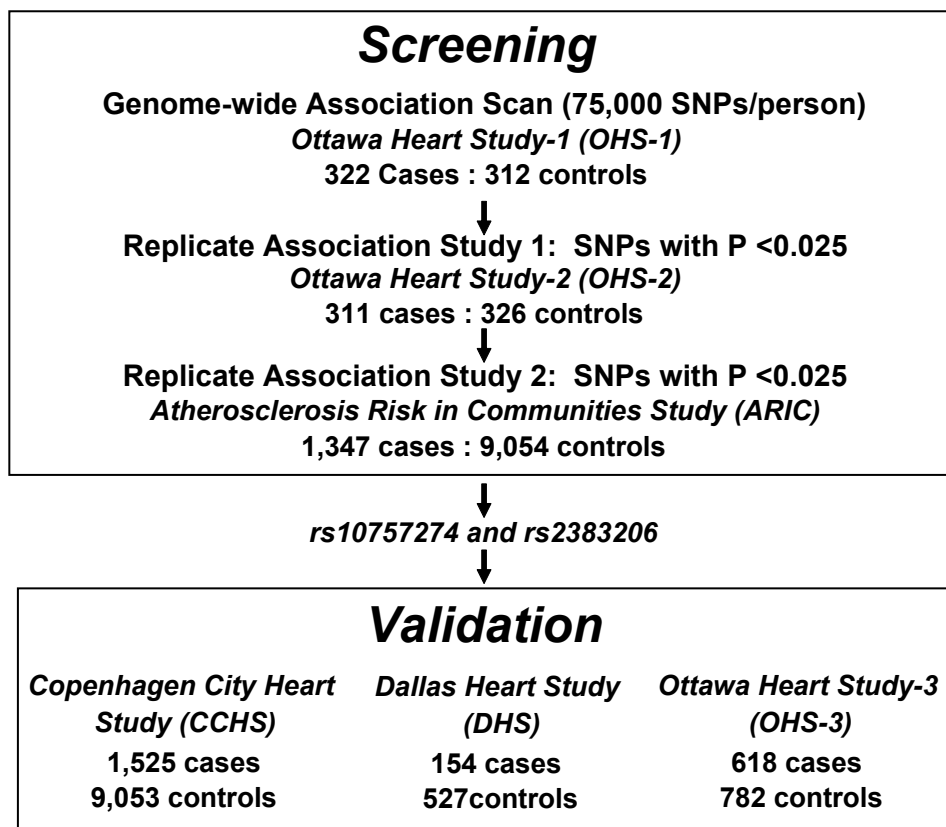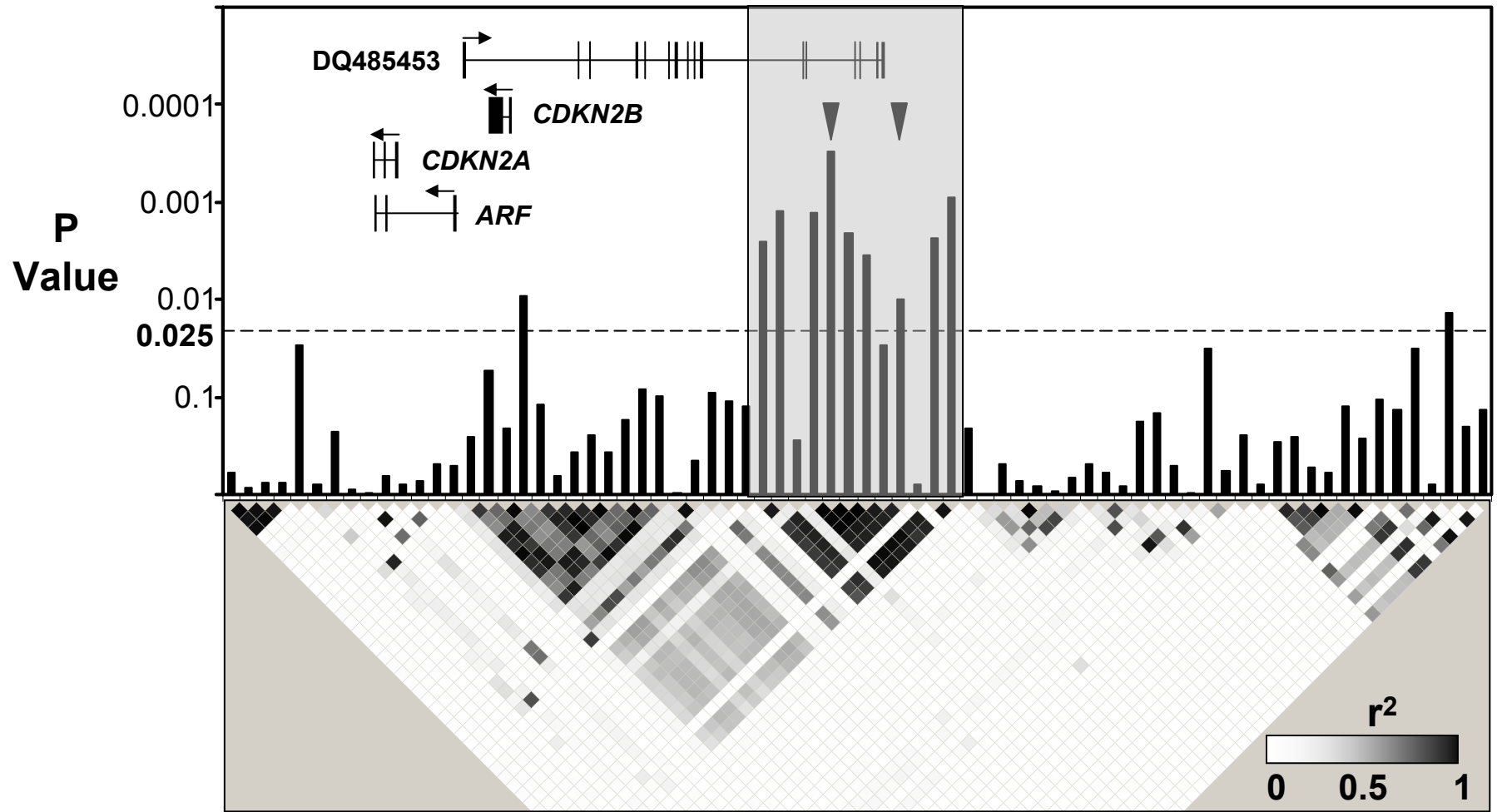Figure 2

Supplementary Online Material

Materials and Methods:

Study Populations

1) The Ottawa Heart Study. White men and women with premature CAD defined as > 50% stenosis in > 1 coronary artery and age of onset < 55y (M) and <65 y (F) were recruited from the patient population of the University of Ottawa Heart Institute in Ottawa, Canada.. Individuals with advanced disease requiring coronary artery bypass grafting or percutaneous coronary intervention who did not have a history of diabetes mellitus or severe dyslipidemia, suggestive of a monogenic lipid disorder (TC>280 mg/dl/7.0 mmol/L) were included in the initial genome-wide scan (OHS1) and in the first replication (OHS 2). The mean age of onset of CAD in these individuals was 47.8+7.5(SD) years. Healthy elderly men (> 65 y) and women (> 70y) of the same ethnic background as the CAD cases were recruited from the Ottawa region. Clinical history and systems review revealed no indicators of clinical CAD. The mean age of the control subjects was 74.9+4.8(SD) years.

2) The Atherosclerosis Risk in Communities Study (ARIC). The ARIC study comprised men and women aged 45 to 64 years who were randomly selected from four communities (Jackson, MS; Minneapolis MN; Forsyth County, NC; and Washington County, MD). Race or ethnic group was determined by self-identification; participants described themselves as black or white in response to a questionnaire on which the available categories were "black", "white", "Indian", or "Asian". Plasma lipids, glucose, insulin and lipoproteins were assayed in the ARIC central lipid laboratory with commercial reagents, as previously described [1-3].

3) The Copenhagen City Heart Study (CCHS). The Copenhagen City Heart Study population was randomly drawn from the Copenhagen Population Register in January 1976[4]. The sample was drawn from a population of approximately 90 000 inhabitants 20 years and older living within 10 wards surrounding Rigshospitalet, the National University Hospital of Copenhagen.

A second examination was performed 5 years later (1981-1983), and a third examination was performed after 15 years (1991-1994), at which time blood samples were obtained from 9,259 individuals for isolation of DNA. A self-administered questionnaire was used to obtain information regarding familial history, education and socio-economic status, and smoking and drinking habits. Plasma cholesterol and triglyceride levels were determined enzymatically using commercial reagents, and HDL-C was determined after removal of apoB-containing lipoproteins by precipitation with phosphotungstic acid and magnesium. Participants were white and of Danish descent.

4) The Dallas Heart Study. The Dallas Heart Study is a multi-ethnic, population-based probability sample of Dallas County residents. The study included three phases: an in-home interview, an in-home phlebotomy visit, and a clinic visit during which a variety of imaging examinations were performed. White men and women who underwent electron-beam computer tomography to assess coronary artery calcification were eligible for the present study.

Genotyping:

1) Oligonucleotide hybridization. A total of 100,000 sequence variants spavced at * kb intervals across the genome were assayed by hybridization to oligonucleotides arrays synthesized on glass slides as described[5]. Briefly, each SNP was interrogated by 24 25mer oligonucleotide probes targeting both the reference and alternate alleles on forward and reverse strands. Each allele and strand was represented by six oligonucleotides comprising five perfect match probes and one mismatch.

2) Mass Spectrometry. The 50 sequence variants identified in the genome-wide scan were assayed in the ARIC population by mass spectrometry using the Sequenom MassARRAY system (Sequenom, Inc.; San Diego, CA).

3) Fluorogenic 5'-nucleotidase assays for rs10757274 and rs2383206 were developed with the use of the TaqMan assay system (Applied Biosystems). The assays were performed on a 7900HT Fast Real-Time PCR instrument with probes and reagents purchased from Applied Biosystems.

DNA sequencing. The 58 kb interval between rs12555547 and rs10965244 was sequenced in two individuals homozygous for the risk allele, and in one individual homozygous for the wild-type allele as described[6]. The coding region and flanking intronic sequences of CDKN2A and CDKN2B were sequenced in 96 healthy white men. All sequence variants identified were verified by manual inspection of the chromatograms and missense changes were confirmed by an independent resequencing reaction.

Reverse transcription and PCR Amplification of cDNAs. RNA was isolated from human brain, liver, placenta, and lymphocytes by a modified phenol-chloroform extraction (TRIZOL reagent, Invitrogen Corporation, Carlsbad, CA), and reverse transcribed (SuperScript III First-strand synthesis system, Invitrogen). Aliquots of cDNA were amplified using primers directed against spliced ESTs CN277071, BX100299 and DQ485453. Primer sequences are available from the authors upon request.

Statistical Analysis. For the genome-wide scan, allelic associations were evaluated for each SNP using chi-square tests on 2×2 contingency tables with no adjustments. Calculations were performed independently at Perlegen Sciences and at UT Southwestern and essentially identical results were obtained. To determine the empirical P-values we randomized the sample case/control status 1,000 times, and calculated allelic associations on each permuted dataset. SNPs that were significantly associated with CHD in the genome-wide scan were assayed in a second set of cases and controls and analyzed using Chi-square tests and permutation testing. For the remaining data sets, case-control differences in allele frequencies of rs10757274 and rs2383206 were evaluated using chi-square tests on 3X2

contingency tables. Population attributable risk was calculated using the formula $PAR = I(T)$

$- I(0)$, where $I(T)$ is the total disease incidence in the population, and $I(0)$ is the disease

incidence in unexposed individuals.

References:

1.   Investigators, T.A.S. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
2.   Chambless, L.E. et al. Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study. *J Clin Epidemiol* **56**, 880-90 (2003).
3.   Brown, S.A. et al. Plasma lipid, lipoprotein cholesterol, and apoprotein distributions in selected U.S. communities. *Arteriscler. Thromb* **13**, 1139-1158 (1993).
4.   Schnohr, P., Jensen, G., Scharling, H. & Appleyard, M. The Copenhagen City Heart Study. Osterbroundersogelsen. Tables with data from the third examination 1991-94. *European Heart Journal* **3, Suppl H.**, H1-H83 (2001).
5.   Saccone, S.F. et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet* **16**, 36-49 (2007).
6.   Tartaglia, M. et al. Gain-of-function SOS1 mutations cause a distinctive form of Noonan syndrome. *Nat Genet* **39**, 75-9 (2007).
7.   Yang, J. et al. Decreased lipid synthesis in livers of mice with disrupted Site-1 protease gene. *Proc Natl Acad Sci U S A* **98**, 13607-13612 (2001).

Supplementary Figure 1. Evolutionary conservation in the risk interval on Chromosome 9.
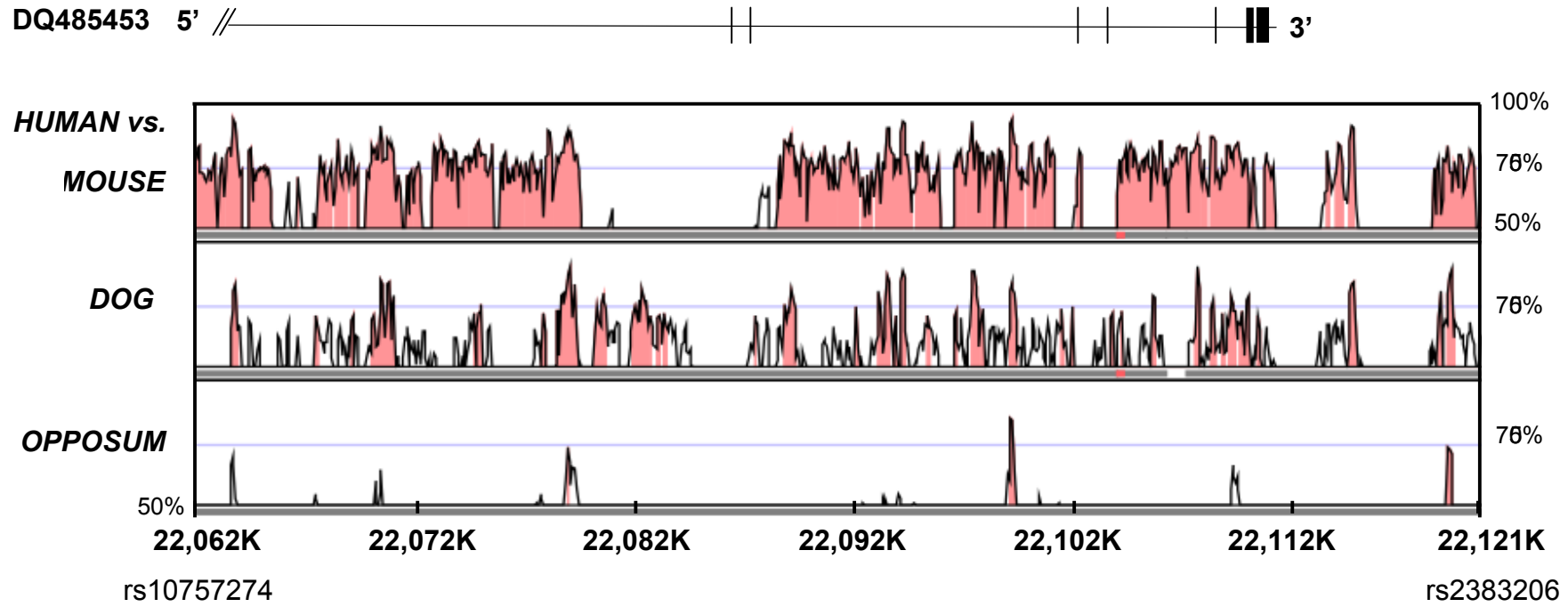
Supplementary Table 1. Clinical and laboratory characteristics of participants in the ARIC Study.

	rs10757274				rs2383206			
	AA	**AG**	**GG**	**P-value**	**AA**	**AG**	**GG**	**P-value**
N	2,293	4,347	2,140	-	2,370	4,761	2,555	-
M/F	1,057/1,236	1,972/2,375	958/1,182	0.67	1,106/1,264	2,160/2,601	1,152/1,403	0.48
Age (years)	54±6	54±6	54±6	0.76	54±6	54±6	54±6	0.09
BMI (kg/m^2)	27.2±5.0	26.9±4.8	26.8±4.7	0.004	27.2±5.0	26.9±4.8	26.9±4.8	0.01
SBP (mm/Hg)	119±17	118±17	118±17	0.44	119±17	118±17	118±17	0.53
DBP (mm/Hg)	72±10	72±10	71±10	0.28	72±10	71±10	72±10	0.06
Glucose (mg/dL)	106±33	105±32	104±25	0.07	105±30	105±32	104±26	0.29
Cholesterol (mg/dL)	215±39	215±41	213±40	0.16	214±39	215±41	213±40	0.17
Triglyceride (mg/dL)	116 (82-167)	112 (80-160)	112 (80-162)	0.16	115 (82-166)	112 (81-160)	113 (81-162)	0.28
LDL-C (mg/dL)	137±37	137±38	135±37	0.23	137±37	137±39	136±37	0.27
HDL-C (mg/dL)	51±17	51±17	51±16	0.56	51±17	51±17	51±16	0.16

Values are means ± standard deviations, except for triglycerides which are medians (interquartile ranges). P-values were calculated using

ANOVA. Variables with highly skewed distributions (triglycerides) were log transformed prior to analysis.

Supplementary Table 2. Clinical and laboratory characteristics of participants in the Copenhagen City Heart Study.

	rs10757274				rs2383206			
	AA	**AG**	**GG**	**P-value**	**AA**	**AG**	**GG**	**P-value**
N	3,145	5,335	2,098	-	2,861	5,365	2,352	-
M/F	1,358/1,787	2,401/2,934	952/1146	0.18	1,233/1,628	2,395/2,970	1,083/1,269	0.10
Age (years)	58±15	57±15	57±15	0.29	58±12	58±15	57±15	0.10
BMI (kg/m^2)	26±4	26±4	26±4	0.57	25±4	26±4	26±4	0.24
SBP (mm/Hg)	139±23	138±22	139±23	0.59	139±23	139±22	139±23	0.96
DBP (mm/Hg)	84±12	84±12	84±12	0.38	84±12	84±12	84±12	0.22
Glucose (mg/dL)	104±34	105±36	105±34	0.35	105±35	105±35	105±34	0.60
Cholesterol (mg/dL)	237±50	238±49	236±52	0.40	237±50	238±49	236±52	0.29
Triglyceride (mg/dL)	135 (96-202)	133 (96-196)	137(96-194)	0.98	135 (96-202)	136(96-196)	137 (96-194)	0.88
LDL-C (mg/dL)	161±48	162±47	161±47	0.43	161±48	162±47	161±47	0.40
HDL-C (mg/dL)	61±19	61±19	60±19	0.64	61±19	61±19	60±19	0.41

Values are means ± standard deviations, except for triglycerides which are medians (interquartile ranges). P-values were calculated using

ANOVA. Variables with highly skewed distributions (triglycerides) were log transformed prior to analysis.

Supplementary Table 3. Sequence variations identified in the 58 kd risk interval on chromosome 9 in two individuals homozygous for the risk allele and one homozygo for the alternate allele.

Position	Type	Allele		Risk Genotype		Referent Genotype
		Major	Minor	1	2	1
22062264	SNP	A	G	G/G	G/G	A/A
22062301	SNP	G	C	C/C	C/C	G/G
22062638	SNP	G	A	A/A	A/A	G/G
22062719	SNP	A	G	G/G	G/G	A/A
22063996	SNP	T	G	G/G	G/G	N/N
22067543	SNP	C	T	T/T	T/T	N/N
22071397	SNP	G	T	T/T	T/T	G/G
22071850	SNP	C	T	T/T	T/T	C/C
22072375	SNP	A	C	A/C	A/A	A/A
22073209	SNP	C	T	C/C	C/T	N/N
22073400	SNP	A	T	A/T	A/A	N/N
22073404	SNP	C	T	T/T	T/T	N/N
22074310	SNP	C	T	T/T	T/T	N/N
22075598	SNP	T	C	C/C	C/C	N/N
22077473	SNP	T	C	C/C	C/C	C/T
22078090	SNP	A	T	T/T	T/T	A/A
22078094	SNP	A	G	G/G	G/G	A/A
22078260	SNP	C	T	T/T	T/T	C/C
22078465	DEL	CA	--	-/-	-/-	CA/CA
22089568	SNP	C	A	A/A	A/A	N/N
22089755	INS	-	A	A/A	A/A	-/-
22090176	SNP	G	C	C/C	C/C	G/G
22091702	SNP	T	C	C/C	C/C	T/T
22092165	SNP	C	T	T/T	T/T	C/C
22092437	SNP	G	A	G/G	G/G	A/G
22093183	SNP	G	T	T/T	T/T	G/G
22093341	SNP	T	G	G/G	G/G	T/T
22093813	SNP	A	G	G/G	G/G	A/A
22095927	SNP	T	C	C/C	C/C	T/T
22096225	SNP	G	A	A/A	A/A	N/N
22096271	SNP	A	G	G/G	G/G	N/N
22096400	SNP	G	A	A/G	G/G	N/N
22096731	SNP	T	A	A/A	A/A	T/T
22097238	SNP	A	T	A/A	A/A	A/T
22100131	SNP	T	C	C/C	C/C	T/T
22101587	INS	-----	TTGAT	TTGAT/ TTGAT	TTGAT/ TTGAT	-/-
22102241	SNP	A	C	C/C	C/C	A/A
22102427	SNP	A	G	G/G	G/G	A/A
22102599	SNP	T	C	C/C	C/C	N/N
22104469	SNP	G	C	C/C	C/C	G/G
22104495	SNP	A	G	G/G	G/G	A/A
22105026	SNP	A	G	G/G	G/G	A/A
22105286	SNP	T	C	C/C	C/C	T/T

22105589	SNP	A	T	T/T	T/T	N/N
22105959	SNP	A	G	G/G	G/G	A/G
22106046	SNP	A	G	G/G	G/G	A/A
22106071	SNP	T	C	C/C	C/C	C/T
22106220	SNP	T	C	C/C	C/C	T/T
22107781	SNP	C	T	C/T	C/C	C/C
22110491	INS	-	T	T/T	T/T	-/-
22110813	DEL	CAT	---	-/-	-/-	CAT/-
22113766	SNP	A	C	C/C	C/C	A/A
22114123	SNP	T	A	A/A	A/A	T/T
22114140	SNP	A	T	T/T	T/T	A/A
22115347	SNP	A	C	C/C	C/C	A/A
22115503	SNP	G	C	C/C	C/C	G/G
22115913	SNP	C	T	T/T	T/T	N/N
22117613	SNP	C	T	C/T	C/C	C/C
22117641	SNP	G	A	A/G	G/G	G/G
22117879	SNP	A	G	A/G	A/G	A/A
22118166	INS	--	AT	AT/-	AT/-	-/-
22118600	SNP	G	A	A/G	A/G	N/N
22118730	SNP	C	G	C/G	C/C	N/N
22119594	SNP	G	C	C/C	C/C	C/C
22119724	INS	-	T	T/T	T/-	T/T
22120389	SNP	A	T	A/T	A/T	A/A

Supplementary Table 4. Allele frequencies of chromosome 9 sequence variants associated with Coronary Heart Disease in whites.

		Minor Allele Frequency	
SNP ID	Position	Whites	African-Americans
rs9632884	22062301	0.48	0.008
rs6475606	22071850	0.5	0.008
rs10757272	22078260	0.5	0.15
rs10757274	22086055	0.49	0.21
rs4977574	22088574	0.5	0.08
rs2891168	22088619	0.5	0.08
rs1333042	22093813	0.49	0.008
rs2383206	22105026	0.49	0.41
rs1333048	22115347	0.49	0.25
rs1333049	22115503	0.49	0.175

Supplementary Table 5. Association between rs10757274 and 2383206 and CHD in African-Americans.

Cohort	rs10757274							rs2383206						
	Controls			Cases				Controls			Cases			
	AA	AG	GG	AA	AG	GG	P	AA	AG	GG	AA	AG	GG	P
ARIC	1857	971	143	187	92	14	0.90	1019	1532	541	114	161	46	0.32
DHS	447	251	32	175	66	21	0.64	231	371	129	90	123	48	0.67

Values in the table are numbers of individuals. P values were calculated using Chi-square tests.