A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction

Anna Helgadottir, *et al.*

Science **316**, 1491 (2007);
DOI: 10.1126/science.1142842

associated with the locus. The localization of the risk locus to a region devoid of known genes implicates a previously unrecognized gene or regulatory element that can substantially affect CHD independently of established risk factors. Further studies will be required to elucidate the mechanism by which the locus modulates CHD risk.

Comparison of the Yoruba and Centre d' Etude du Polymorphisme Humain (CEPH) data from the International HapMap Project (www.hapmap.org) revealed notable ethnic differences in allele frequencies in the risk interval (table S6). Of the 10 alleles that were significantly associated with CHD in whites, 3 were virtually absent from the Yoruba population, and 6 others were much 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 (table S7). The apparent ethnic differences in association between these SNPs and CHD in ARIC may reflect differences in statistical power in ARIC but cannot explain the ethnic differences observed in DHS, where African-Americans are the largest group. 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 among ethnic groups (*9*–*11*). 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 prespecified significance threshold than would be predicted by chance alone, as indicated by permutation testing (table S2). 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 because of 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 (*12*). Furthermore, analysis of 50 randomly selected regions of 500 kb each indicated that the 72,864 informative SNPs used in the initial scan provided 30 to 40% of the power that would be obtained by assaying all phase II Hapmap SNPs. Therefore, scans with denser SNP panels and larger sample sizes may reveal further loci associated with CHD risk.

References and Notes

1. C. J. Murray, A. D. Lopez, *Lancet* **349**, 1436 (1997).
2. C. D. Mathers, D. Loncar, *PLoS Med.* **3**, e442 (2006).
3. S. F. Saccone *et al.*, *Hum. Mol. Genet.* **16**, 36 (2007).
4. The ARIC Study Investigators, *Am. J. Epidemiol.* **129**, 687 (1989).
5. P. Schnohr, G. Jensen, H. Scharling, M. Appleyard, *Eur. Heart J.* **3** (suppl. H), H1 (2001).
6. R. G. Victor *et al.*, *Am. J. Cardiol.* **93**, 1473 (2004).
7. A. S. Agatston *et al.*, *J. Am. Coll. Cardiol.* **15**, 827 (1990).
8. E. Pasmant *et al.*, *Cancer Res.* **67**, 1 (2007).
9. J. C. Cohen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1810 (2006).
10. J. C. Cohen, E. Boerwinkle, T. H. Mosley, H. H. Hobbs, *N. Engl. J. Med.* **354**, 34 (2006).
11. S. Romeo *et al.*, *Nat. Genet.* **39**, 513 (2007).
12. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III), *Circulation* **106**, 3143 (2002).

13. We thank T. Hyatt, H. Doelle, T. Naing, L. Nie, K. Moller Rasmussen, M. Refstrup, W. S. Schackwitz, J. Martin, and A. Ustaszewska for excellent technical assistance and J. Schageman, C. Lee, K. Lawson, and K. Williams for statistical analyses. We are indebted to the staff and participants of the OHS, the DHS, the ARIC study, and the CCHS for their important contributions. Supported by grants from Foundation Le Ducq, the Donald W. Reynolds Foundation, NIH (HL-082896), the National Heart, Lung, and Blood Institute Program for Genomic Applications (no. HL-066681), the U.S. Department of Energy (contract no. DE-AC02-05CH11231), the Canadian Institutes of Health Research (no. 44360), the Canadian Foundation for Innovation, the Heart and Stroke Foundation of Ontario (no. NA-5413) the Danish Medical Research Council, the Danish Heart Foundation, and the Research Fund at Rigshospitalet, Copenhagen University Hospital. D.R.C. and D.A.H. have equity interest in Perlegen Sciences.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1142447/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S7

12 March 2007; accepted 24 April 2007
Published online 3 May 2007;
10.1126/science.1142447
Include this information when citing this paper.

A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction

Anna Helgadottir,[1]* Gudmar Thorleifsson,[1]* Andrei Manolescu,[1]* Solveig Gretarsdottir,[1] Thorarinn Blondal,[1] Aslaug Jonasdottir,[1] Adalbjorg Jonasdottir,[1] Asgeir Sigurdsson,[1] Adam Baker,[1] Arnar Palsson,[1] Gisli Masson,[1] Daniel F. Gudbjartsson,[1] Kristinn P. Magnusson,[1] Karl Andersen,[2] Allan I. Levey,[3] Valgerdur M. Backman,[1] Sigurborg Matthiasdottir,[1] Thorbjorg Jonsdottir,[1] Stefan Palsson,[1] Helga Einarsdottir,[1] Steinunn Gunnarsdottir,[1] Arnaldur Gylfason,[1] Viola Vaccarino,[3] W. Craig Hooper,[3] Muredach P. Reilly,[4] Christopher B. Granger,[5] Harland Austin,[3] Daniel J. Rader,[4] Svati H. Shah,[5] Arshed A. Quyyumi,[3] Jeffrey R. Gulcher,[1] Gudmundur Thorgeirsson,[2] Unnur Thorsteinsdottir,[1] Augustine Kong,[1]† Kari Stefansson[1]†

The global endemic of cardiovascular diseases calls for improved risk assessment and treatment. Here, we describe an association between myocardial infarction (MI) and a common sequence variant on chromosome 9p21. This study included a total of 4587 cases and 12,767 controls. The identified variant, adjacent to the tumor suppressor genes *CDKN2A* and *CDKN2B*, was associated with the disease with high significance. Approximately 21% of individuals in the population are homozygous for this variant, and their estimated risk of suffering myocardial infarction is 1.64 times as great as that of noncarriers. The corresponding risk is 2.02 times as great for early-onset cases. The population attributable risk is 21% for MI in general and 31% for early-onset cases.

Coronary artery disease (CAD), including acute myocardial infarction (MI), is the leading cause of death worldwide (*1*). Identification of the underlying genetic architecture of heart disease may provide improved risk assessment and better measures for prevention and treatment.

To this end, we conducted a genome-wide association study on Icelandic patients with MI, using the Illumina Hap300 chip. After quality filtering, 305,953 single-nucleotide polymorphisms (SNPs) were tested for association with MI in a sample of 1607 cases, with age at onset before 70 in males and before 75 in females; 6728 patients without a history of CAD were used as controls (*2*). The results were adjusted for relatedness between individuals and potential population stratification with the use of a method of genomic control (*3*). Although none of the SNPs were significant after adjusting for the number of tests performed, more signals with P values of less than 10^{-5} were observed than expected by chance (fig. S1). Hence, we further explored the SNPs with P values that were closest to genome-wide significance.

The strongest association with MI was observed with three correlated SNPs—rs1333040,

[1]deCODE genetics, Sturlugata 8, IS-101 Reykjavik, Iceland. [2]Landspitali University Hospital, Reykjavik, Iceland. [3]Emory University School of Medicine, Atlanta, GA 30322, USA. [4]University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA. [5]Duke University School of Medicine, Durham, NC 27710, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: augustine.kong@decode.is (A.K.); kstefans@decode.is (K.S.)

rs2383207, and rs10116277. Each had an odds ratio (OR) around 1.22 for the risk allele and a P value of approximately 1×10^{-6} (table S1). All three SNPs are located within a 190-kb linkage disequilibrium (LD) block on chromosome 9p21 (fig. S2). Apart from these 3 SNPs, 11 other SNPs in the same LD block showed nominally significant association with MI. The associations with these SNPs tended to become weaker after accounting for the association with the three SNPs mentioned above (table S1). After adjustment, a few remained nominally significant ($P < 0.05$), but none had a $P < 0.01$.

To replicate the observed associations, we genotyped the three SNPs—rs1333040, rs2383207 and rs10116277—in an additional 665 Icelandic MI cases and 3533 controls, and in three case-control sample sets of European descent from three cities from the United States: Philadelphia, Atlanta, and Durham (*2*). For consistency, we used the same age-at-onset criteria in the association analysis as for the discovery group. The association with MI was replicated with significance in all four groups (table S2). When we combined the replication sets using a Mantel-Haenszel model (*4*), all three SNPs showed highly significant association with MI ($P < 1 \times 10^{-8}$), with ORs comparable to those observed in the Icelandic discovery samples. When all groups were combined, rs2383207 showed the most-significant association ($P = 2.0 \times 10^{-16}$), with an OR of 1.25 [95% confidence interval (CI) 1.18 to 1.31] for the risk allele G. Notably, rs2383207 and rs10116277 are highly correlated ($r^2 = 0.90$) and their effects could not be reliably distinguished from each other in these data. The SNP rs1333040 is also substantially correlated with rs2383207 and rs10116277 ($r^2 = 0.57$ and 0.67, respectively). In an attempt to refine this association signal, we identified the SNPs that are substantially correlated with rs2383207 ($r^2 > 0.5$) based on the Hapmap CEU data and that are not part of the Illumina Hap300 chip. Among the 36 such SNPs, we selected 8 to be genotyped. Each of the 36 SNPs was either one of the eight or had a very good surrogate among them ($r^2 > 0.90$) (table S7). With data from all case-control groups combined, allele G of the refinement SNP rs10757278 showed the strongest association with the disease (OR = 1.28, $P = 1.2 \times 10^{-20}$; Table 1 and table S2). Furthermore, whereas rs2383207 was no longer significant after adjusting for rs10757278 ($P = 0.25$), rs10757278 remained significant with adjustment for rs2383207 ($P = 2.0 \times 10^{-5}$). Among the SNPs in this region that showed very significant association with the disease when tested individually, none was significant after adjustment for rs10757278 with the exception of the refinement SNP rs1333046, which was marginally significant ($P = 0.044$) with adjustment (table S8). Henceforth, for simplicity of presentation, we focused on the most significant SNP, rs10757278. Additional results for other SNPs in the region are provided in tables S2 to S6.

Table 1. Association results for rs2383207 (G) and rs10757278 (G), in 9p21 in Iceland and the United States. Results are shown for the initial Icelandic discovery MI case-control group (Iceland A), an independent Icelandic replication group (Iceland B), and for the three U.S. replication groups of European descent. Also included are the results for the MI case-control groups combined. Study population includes the number of MI cases (*n*) and controls (*m*). To combine the Icelandic cohorts, we analyzed them together and adjusted the results for relatedness in the combined group. For the combined groups, we calculated OR and *P* value using a Mantel-Haenszel model, and the frequency in cases and controls is a simple average over the frequency in the individual groups. When combining Icelandic and U.S. groups, the frequency in cases and controls is the average over the two populations.

Study population (*n*/*m*) Variant (allele)	Frequency		OR (95% CI)	*P*
	Controls	Cases		
Iceland A (1607/6728)				
rs2383207 (G)	0.455	0.506	1.22 (1.13–1.33)	1.4×10^{-6}
rs10757278 (G)	0.434	0.489	1.25 (1.15–1.36)	1.5×10^{-7}
Iceland B (665/3533)				
rs2383207 (G)	0.462	0.525	1.29 (1.15–1.45)	2.6×10^{-5}
rs10757278 (G)	0.436	0.503	1.31 (1.16–1.47)	1.4×10^{-5}
Atlanta (596/1284)				
rs2383207 (G)	0.541	0.593	1.23 (1.07–1.42)	3.0×10^{-3}
rs10757278 (G)	0.484	0.551	1.31 (1.14–1.50)	1.5×10^{-4}
Philadelphia (582/504)				
rs2383207 (G)	0.524	0.602	1.37 (1.16–1.63)	2.6×10^{-4}
rs10757278 (G)	0.470	0.550	1.38 (1.17–1.64)	1.9×10^{-4}
Durham (1137/718)				
rs2383207 (G)	0.513	0.559	1.20 (1.05–1.37)	6.0×10^{-3}
rs10757278 (G)	0.460	0.521	1.28 (1.12–1.46)	2.7×10^{-4}
Combined				
Iceland (2272/10,261)				
rs2383207 (G)	0.458	0.511	1.24 (1.16–1.33)	3.3×10^{-10}
rs10757278 (G)	0.435	0.493	1.26 (1.18–1.35)	5.3×10^{-12}
U.S. groups (2315/2506)				
rs2383207 (G)	0.526	0.585	1.25 (1.15–1.36)	1.1×10^{-7}
rs10757278 (G)	0.471	0.541	1.31 (1.21–1.43)	1.5×10^{-10}
Replication groups (2980/6039)				
rs2383207 (G)	0.494	0.555	1.27 (1.18–1.36)	1.4×10^{-11}
rs10757278 (G)	0.454	0.522	1.31 (1.22–1.40)	1.0×10^{-14}
All groups (4587/12,767)				
rs2383207 (G)	0.492	0.548	1.25 (1.18–1.31)	2.0×10^{-16}
rs10757278 (G)	0.453	0.517	1.28 (1.22–1.35)	1.2×10^{-20}

Table 2. Genotype-specific OR for the risk allele of rs10757278. The risk for heterozygous carriers (0X) and homozygous carriers (XX) is compared with the risk for noncarriers (00), together with 95% CI and the population attributable risk (PAR). The lower part of the table includes the corresponding values when the analysis is restricted to early-onset MI cases. Study population includes the number of MI cases (*n*) and controls (*m*). For the Icelandic groups, *P* values and OR were adjusted for relatedness using simulations.

Study population (*n*/*m*) Variant (allele)	Genotype-specific OR			PAR
	00	0X (95% CI)	XX (95% CI)	
Iceland (2272/10,261)				
rs10757278 (G)	1	1.25 (1.12–1.39)	1.58 (1.38–1.81)	0.19
U.S. groups (2315/2506)				
rs10757278 (G)	1	1.28 (1.14–1.45)	1.72 (1.45–2.03)	0.23
All groups (4587/12,767)				
rs10757278 (G)	1	1.26 (1.16–1.36)	1.64 (1.47–1.82)	0.21
Early-onset MI (<50 for males; <60 for females)				
Iceland (621/10,261)				
rs10757278 (G)	1	1.38 (1.13–1.69)	1.94 (1.53–2.46)	0.27
U.S. groups (1080/2506)				
rs10757278 (G)	1	1.56 (1.32–1.85)	2.08 (1.69–2.58)	0.34
All groups (1701/12,767)				
rs10757278 (G)	1	1.49 (1.31–1.69)	2.02 (1.72–2.36)	0.31

Table 3. Association of the G allele of rs10757278 with CAD. The association results are shown for CAD, both including and excluding known MI cases. Results are shown for the Icelandic case-control group (except for the initial discovery group individuals, which have been excluded both from cases and controls), for two of the U.S. groups, and for the groups combined. For the combined groups, the allelic frequency in cases and controls is a simple average over the individual groups or, when combining Icelandic and U.S. groups, the average over the two populations. Study population includes the number of all cases (*n1*), cases excluding MI patients (*n2*), and controls (*m*).

Study population (*n1*/*n2*/*m*) Variant (allele)	Control frequency	All CAD cases			Excluding MI cases		
		Case frequency	OR (95% CI)	*P*	Case frequency	OR (95% CI)	*P*
Iceland (1563/773/3533)							
rs10757278 (G)	0.439	0.496	1.26 (1.15–1.37)	1.9×10^{-7}	0.490	1.22 (1.09–1.37)	5.0×10^{-4}
Atlanta (724/128/1284)							
rs10757278 (G)	0.484	0.552	1.31 (1.15–1.50)	3.6×10^{-5}	0.557	1.34 (1.04–1.73)	2.6×10^{-2}
Philadelphia (709/126/504)							
rs10757278 (G)	0.470	0.547	1.36 (1.16–1.60)	1.9×10^{-4}	0.528	1.26 (0.96–1.66)	1.0×10^{-1}
Combined							
U.S. groups (1433/254/1788)							
rs10757278 (G)	0.477	0.550	1.33 (1.20–1.47)	2.7×10^{-8}	0.542	1.30 (1.08–1.57)	5.9×10^{-3}
All groups (2996/1027/5321)							
rs10757278 (G)	0.458	0.523	1.29 (1.21–1.38)	3.6×10^{-14}	0.525	1.24 (1.13–1.37)	1.1×10^{-5}

To investigate the mode of inheritance in more detail, we computed genotype-specific ORs for rs10757278. With results from all groups combined, relative to noncarriers, the ORs for heterozygous and homozygous carriers of the risk allele G were 1.26 and 1.64, respectively (Table 2). Assuming a frequency of 45.3% for the allele, the average of the frequencies in Iceland and the United States, the corresponding population attributable risk is 21%.

Because the impact of genetic factors on CAD has been shown to be greater at early ages (*5*), we investigated the correlation of allele G of rs10757278 to age at onset of MI. In this analysis, we used all cases with a known age at onset, including those who had onset after the age of 70 or 75 for males and females, respectively. This added a total of 973 cases to the study groups compared with the number of cases used in the case-control analyses. Regressing the age at onset on the number of risk alleles showed that, for each copy of the risk allele, the age at onset of MI was on average reduced by approximately 1 year ($P = 2.9 \times 10^{-7}$) (table S4). Alternatively, restricting the case-control analysis to early-onset MI, defined as an MI before the age of 50 for males and before the age of 60 for females, the allelic OR for rs10757278 G in all groups combined increased to 1.42 (95% CI 1.31 to 1.53) (table S5). Relative to noncarriers, genotype-specific OR for early-onset MI is 1.49 and 2.02 for heterozygous and homozygous carriers of the risk allele, respectively (Table 2).

Having established that allele G of rs10757278 is associated with MI, we explored its impact on the broader phenotype of CAD (Table 3). To eliminate bias that could have arisen from the selection of the most-significant variants in the initial genome-wide study, the cases and controls from the Icelandic discovery group (Iceland A) were not included here. If the Icelandic group had been included, there would have been little change to the estimated effects, but the results would have become more significant due to the larger sample sizes. Also, the group from Durham did not have CAD cases without MI. As expected, rs10757278 was associated with CAD with high significance (OR = 1.29, $P = 3.6 \times 10^{-14}$ for the groups combined). After removal of MI cases from the analyses, the associations remained significant for the groups from Iceland and Atlanta but not for the Philadelphia group. Combining results from the three groups gave an OR of 1.24 ($P = 0.000011$).

The variants on chromosome 9p21 associated with MI are located in an LD block that contains the *CDKN2A* and *CDKN2B* genes. The proteins encoded by these genes—called $p16^{INK4a}$, ARF, and $p15^{INK4b}$—play a critical role in regulating cell proliferation, cell aging and senescence, and apoptosis in many cell types (*6*). These are all important features of atherogenesis, the underlying cause of MI and CAD (*7*, *8*). Sequencing of 93 early-onset MI patients across exons, exon-intron junctions, and regulatory regions of *CDKN2A* and *CDKN2B* did not reveal obvious candidates for functional variants or other variants that could account for the observed association with rs10757278 (tables S11 and S12). In addition to *CDKN2A* and *CDKN2B* genes, the LD block contains two exons of the mRNA transcript AF109294, a hypothetical methylthioadenosine phosphorylase fusion protein mRNA, and several expressed sequence tags that are expressed in various tissues (*2*). The functional relevance of the variants of this genomic region to MI and CAD remains to be elucidated.

We show that a common genetic variant located in the vicinity of the tumor suppressor genes *CDKN2A* and *CDKN2B* on chromosome 9p21 associates with MI. This is the first common variant discovered to consistently confer substantial risk (OR > 1.20) of MI in multiple case-control groups of European descent. Due to its high frequency, the population attributable risk of the variant is approximately 21% for MI in general and approximately 31% for early-onset cases, which is substantial from a public health point of view. However, as the relative risks are not extremely high, it explains only a small fraction of the familial clustering of the disease and would not generate large linkage scores. Hence, other susceptibility variants remain to be identified and some could be located in candidate regions identified by genome-wide linkage scans (*9*–*11*). There is evidence to suggest that the variants identified here could increase the risk of CAD in general in addition to their impact on MI, an observation that warrants further investigation. The mechanism whereby the genetic variants exert their effects in the pathogenesis of MI remains to be elucidated.

References and Notes

1. T. Thom *et al.*, *Circulation* **113**, e85 (2006).
2. Materials and methods are available as supporting material on *Science* Online.
3. B. Devlin, K. Roeder, *Biometrics* **55**, 997 (1999).
4. N. Mantel, W. Haenszel, *J. Natl. Cancer Inst.* **22**, 719 (1959).
5. S. Zdravkovic *et al.*, *J. Intern. Med.* **252**, 247 (2002).
6. W. Y. Kim, N. E. Sharpless, *Cell* **127**, 265 (2006).
7. A. J. Lusis, *Nature* **407**, 233 (2000).
8. T. Minamino, I. Komuro, *Circ. Res.* **100**, 15 (2007).
9. E. Zintzaras, G. Kitsios, *J. Hum. Genet.* **51**, 1015 (2006).
10. Q. Wang *et al.*, *Am. J. Hum. Genet.* **74**, 262 (2004).
11. N. J. Samani *et al.*, *Am. J. Hum. Genet.* **77**, 1011 (2005).

12. We thank the participants whose contribution made this study possible, the nurses at the Icelandic Heart Association, and personnel at the deCODE core facilities. The authors from deCODE genetics declare competing financial interests. deCODE genetics has applied for a patent related to this work.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1142842/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
Tables S1 to S13
References

21 March 2007; accepted 26 April 2007
Published online 3 May 2007;
10.1126/science.1142842
Include this information when citing this paper.