



## Multiple Rare Alleles Contribute to Low Plasma Levels of HDL Cholesterol

Jonathan C. Cohen, *et al.*  
*Science* **305**, 869 (2004);  
DOI: 10.1126/science.1099870

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of April 13, 2007):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/305/5685/869>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/cgi/content/full/305/5685/869/DC1>

This article **cites 13 articles**, 7 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/305/5685/869#otherarticles>

This article has been **cited by** 131 article(s) on the ISI Web of Science.

This article has been **cited by** 48 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/305/5685/869#otherarticles>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

the Epo-TB complex so as to engage epothilone (Fig. 3). A similar arginine displacement (4 Å) has been observed in the binding of two epothilones to cytochrome P450-EpoK (32). The double arginine relocation reflects a subtle reorganization of M-loop residues not previously seen with taxanes, but both epothilone and Taxol bridge the M-loop and helix H7 adjacent to the nucleotide-binding site and thereby promote tubulin polymerization and microtubule stability.

Instead of a common pharmacophore (8–11), tubulin displays a promiscuous binding pocket with the bound molecules exploiting contacts with an optimal subset of binding pocket residues. Although this provides a unique challenge for “rational” ligand design, it can be anticipated that the promiscuity principle will apply to the binding of other ligands that occupy the taxane site on TB, namely, discodermolide, eleutherobin, and the sarcodictyins (33).

#### References and Notes

1. A. K. Singla, A. Garg, D. Aggarwal, *Int. J. Pharm.* **235**, 179 (2002).
2. M. M. Gottesman, *Annu. Rev. Med.* **53**, 615 (2002).

3. J. P. Guastalla, V. Diéras, *Br. J. Cancer* **89**, S16 (2003).
4. G. Höfle, N. Bedorf, K. Gerth, H. Reichenbach, *Chem. Abstr.* **120**, 52841 (1993).
5. D. Bollag et al., *Cancer Res.* **55**, 2325 (1995).
6. M. Wartmann, K. H. Altmann, *Curr. Med. Chem. Anti-Cancer Agents* **2**, 123 (2002).
7. S. Kar, J. Fan, M. J. Smith, M. Goedert, L. A. Amos, *EMBO J.* **22**, 70 (2003).
8. L. He et al., *Biochemistry* **39**, 3972 (2000).
9. P. Giannakakou et al., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2904 (2000).
10. M. Wang et al., *Org. Lett.* **1**, 43 (1999).
11. I. Ojima et al., *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4256 (1999).
12. Materials and methods are available as supporting material on Science Online.
13. J. P. Snyder, J. H. Nettles, B. Cornett, K. H. Downing, E. Nogales, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5312 (2001).
14. T. Carlomagno et al., *Angew. Chem. Int. Ed. Engl.* **42**, 2511 (2003).
15. R. E. Taylor, Y. Chen, G. M. Galvin, P. K. Pabba, *Org. Biomol. Chem.* **2**, 127 (2004).
16. L. He, C.-P. H. Yang, S. B. Horwitz, *Mol. Cancer Ther.* **1**, 3 (2001).
17. N. M. Verrills et al., *Chem. Biol.* **10**, 597 (2003).
18. P. Giannakakou et al., *J. Biol. Chem.* **272**, 17118 (1997).
19. M. Wang, B. Cornett, J. Nettles, D. C. Liotta, J. P. Snyder, *J. Org. Chem.* **65**, 1059 (2000).
20. J. Lowe, H. Li, K. H. Downing, E. Nogales, *J. Mol. Biol.* **313**, 1045 (2001).
21. K. C. Nicolaou, F. Roschangar, D. Vourloumis, *Angew. Chem. Int. Ed. Engl.* **37**, 2014 (1998).

22. R. E. Taylor, Y. Chen, A. Beatty, D. C. Myles, Y. Zhou, *J. Am. Chem. Soc.* **125**, 26 (2003).
23. K. C. Nicolaou, A. Ritzen, K. Namoto, *Chem. Commun.* **2001**, 1523 (2001).
24. J. Johnson et al., *Org. Lett.* **2**, 1537 (2000).
25. K. C. Nicolaou et al., *J. Am. Chem. Soc.* **123**, 9313 (2001).
26. A. Regueiro-Ren et al., *Org. Lett.* **4**, 3815 (2002).
27. A. Rivkin et al., *J. Am. Chem. Soc.* **125**, 2899 (2003).
28. I. H. Hardt et al., *J. Nat. Prod.* **64**, 847 (2001).
29. K. C. Nicolaou et al., *Chem. Biol.* **7**, 593 (2000).
30. K. C. Nicolaou et al., *Angew. Chem. Int. Ed. Engl.* **42**, 3515 (2003).
31. A. Regueiro-Ren et al., *Org. Lett.* **3**, 2693 (2001).
32. S. Nagano et al., *J. Biol. Chem.* **278**, 44886 (2003).
33. J. Jimenez-Barbero, F. Amat-Guerri, J. P. Snyder, *Curr. Med. Chem. Anti-Cancer Agents* **2**, 91 (2002).
34. Supported in part by NIH and the Department of Energy. J.N., B.C., and J.P.S. thank D. Liotta (Emory University) for encouragement, discussion, and support during the course of the work. We thank P. Giannakakou, J. Gallivan, and D. Lynn (Emory University) for helpful discussions. Structure has been deposited in the Protein Data Bank under accession number 1TVK.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/305/5685/866/DC1

Materials and Methods

Figs. S1 to S8

References

15 April 2004; accepted 12 July 2004

## Multiple Rare Alleles Contribute to Low Plasma Levels of HDL Cholesterol

Jonathan C. Cohen,<sup>1,2,3,\*†</sup> Robert S. Kiss,<sup>5\*</sup>  
Alexander Pertsemlidis,<sup>1</sup> Yves L. Marcel,<sup>5†</sup> Ruth McPherson,<sup>5</sup>  
Helen H. Hobbs<sup>1,3,4</sup>

Heritable variation in complex traits is generally considered to be conferred by common DNA sequence polymorphisms. We tested whether rare DNA sequence variants collectively contribute to variation in plasma levels of high-density lipoprotein cholesterol (HDL-C). We sequenced three candidate genes (*ABCA1*, *APOA1*, and *LCAT*) that cause Mendelian forms of low HDL-C levels in individuals from a population-based study. Nonsynonymous sequence variants were significantly more common (16% versus 2%) in individuals with low HDL-C (<fifth percentile) than in those with high HDL-C (>95th percentile). Similar findings were obtained in an independent population, and biochemical studies indicated that most sequence variants in the low HDL-C group were functionally important. Thus, rare alleles with major phenotypic effects contribute significantly to low plasma HDL-C levels in the general population.

Many clinically important quantitative traits are highly heritable, but progress in the elucidation of their genetic architecture has been limited. Because quantitative traits do not segregate in Mendelian fashion in most families, their distribution in the population is presumed to reflect the cumulative contribution of multiple common DNA sequence variants that each has a small effect (1). Sequence variants with strong phenotypic effects may also contribute to variation in complex traits (2). Although these variants

are likely to be rare individually, they may be sufficiently common in aggregate to contribute to variation in common traits in the population. Whereas most Mendelian disorders are caused by a spectrum of different mutations in a gene (or genes) (3), the contribution of rare alleles to more common, quantitative traits has not been systematically evaluated.

In this study, we evaluated the hypothesis that rare sequence variations contribute significantly to low plasma levels of high-density lipoprotein cholesterol (HDL-C), a

major risk factor for coronary atherosclerosis. If this hypothesis is correct, then mutations that impair HDL production or enhance HDL catabolism should be significantly more common among individuals with low plasma levels of HDL-C than among those with high plasma levels of HDL-C. Furthermore, sequence variants with major phenotypic effects are likely to be found exclusively in one extreme or the other, whereas alleles found in both the high HDL-C and the low HDL-C groups are likely to be neutral with respect to plasma HDL-C levels. The prevalence of mutations with major effects on plasma HDL-C levels is not known. Molecular defects causing rare genetic forms of HDL deficiency have been identified in the genes encoding apolipoprotein AI (*APOA1*), the major protein component of HDL (4); the adenosine triphosphate binding cassette (ABC) transporter A1 (*ABCA1*), required for the efflux of cholesterol from cells to HDL particles (5); and lecithin cholesterol acyltransferase (*LCAT*), the enzyme that catalyzes the formation of

<sup>1</sup>Donald W. Reynolds Cardiovascular Clinical Research Center and McDermott Center for Human Growth and Development, <sup>2</sup>Center for Human Nutrition, <sup>3</sup>Department of Internal Medicine and Department of Molecular Genetics, <sup>4</sup>Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA. <sup>5</sup>Lipoprotein and Atherosclerosis Research Group, University of Ottawa Heart Institute, Ottawa, Ontario K7Y4W7, Canada.

\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: jonathan.cohen@utsouthwestern.edu (J.C.C.); ymarcel@ottawaheart.ca (Y.L.M.)

REPORTS

cholesteryl esters in HDL (6). Homozygotes for mutations in these genes have virtually no circulating HDL-C, whereas heterozygotes have about half the normal plasma level of HDL-C. Mutations in *ABCA1* have also been associated with a less severe form of familial hypoalphalipoproteinemia (7, 8), but systematic screening of these genes in population-based samples has not been reported.

To determine whether sequence variations in these three candidate genes contribute to low plasma levels of HDL-C in the general population, we sequenced the coding regions and consensus splice sites of each gene in 32 individuals from each of four gender and race groups [black men, white men, black women, and white women (table S1)] constituting the upper and lower 5% of the distribution of plasma HDL-C levels in a population-based study of Dallas County residents (9). Nonsynonymous sequence variations were significantly more common in the low HDL-C group than in the high HDL-C group (Table 1). Of the 128 individuals with low plasma levels of HDL-C, 21 (16%) had sequence variants not present in the high HDL-C group (Table 2). In contrast, only 3 (2%) of the individuals in the high HDL-C group had sequence variants not found in the low HDL-C group ( $P < 0.0001$ , Fisher's exact test). A total of 20 of the 21 mutant alleles found only in the low HDL-C group were in *ABCA1* (fig. S1) and two of these (W590S and N1800H) were previously identified in individuals with hypoalphalipoproteinemia (6).

Thus, one of six individuals with HDL-C levels below the fifth percentile in the Dallas Heart Study had a rare mutation in *ABCA1* or *APOA1*. Similar analyses were performed in a second sample comprising white Canadians who had either low (10 to 34 mg/dl) or high (58 to 116 mg/dl) levels of HDL-C (Table 3). A total of 21 (14%) of the 155 Canadians in the low HDL-C group had sequence variants that were not present in the 108 Canadians with a high plasma level of HDL-C. In contrast, only three individuals in the high HDL-C group (3%) had sequence variants not present in the low HDL-C group ( $P < 0.001$ , Fisher's exact test). Four sequence variants found in the Canadian low HDL-C group (*ABCA1* P85L, N1800H, S1731C, and R1851X) had been previously identified in subjects with hypoalphalipoproteinemia (10). One mutation (*ABCA1* N1800H) was found in both the Canadian and Dallas low HDL-C groups. In both Canadian and Dallas groups, almost all of the excess sequence variations in the low HDL-C group were in *ABCA1*. This may reflect the size of the coding sequence of *ABCA1*, which is three times larger than those of the other two genes; it is also possible that sequence changes occur more frequently at this locus or that mutations arising in *ABCA1* are more likely to persist in the population.

To test whether differences in population substructure between the low HDL-C and high

HDL-C groups contributed to the observed differences in the prevalence of rare sequence variants, we compared the frequency of synonymous substitutions in the high and low HDL-C groups. The frequency of synonymous substitutions was similar in the high HDL-C and low HDL-C groups in both the U.S. and Canadian populations (Table 1). Furthermore, the racial

composition of the low HDL-C and high HDL-C groups was matched. Thus, it is unlikely that the excess sequence variations in the low HDL-C group is attributable to either population stratification or chance variations in allele frequencies.

Analysis using a computer program [PolyPhen (11)] that predicts effects of amino acid changes on protein function (12) indicated that

**Table 1.** Sequence variations in the coding regions of *ABCA1*, *APOA1*, and *LCAT*. Values represent the numbers of sequence variants identified in 256 individuals from the Dallas Heart Study (DHS) (128 with low HDL-C and 128 with high HDL-C) and 263 Canadians (155 with low HDL-C and 108 with high HDL-C) (17). NS, nonsynonymous (nucleotide substitutions resulting in an amino acid change); S, synonymous (coding sequence substitutions that do not result in an amino acid change). GenBank accession numbers for DHS *ABCA1*, *APOA1*, and *LCAT* sequences are NM\_005502, NM\_000039, and NM\_000229, respectively.

	Sequence variants unique to one group				Sequence variants common to both groups	
	Low HDL-C		High HDL-C		NS	S
	NS	S	NS	S		
DHS						
<i>ABCA1</i>	14	6	2	5	10	19
<i>APOA1</i>	1	0	0	1	0	1
<i>LCAT</i>	0	1	1	0	1	1
Canadians						
<i>ABCA1</i>	14	2	2	3	7	5
<i>APOA1</i>	0	1	0	0	2	0
<i>LCAT</i>	6	1	0	0	0	0

**Table 2.** Nonsynonymous sequence variations in *ABCA1*, *APOA1*, and *LCAT* in Dallas Heart Study participants with low ( $n = 128$ ) or high ( $n = 128$ ) plasma HDL-C levels. The effect of each amino acid (18) substitution on protein function was predicted with the use of PolyPhen (11, 19). Substitutions affecting known functional residues in the protein or that resulted in large changes in surface accessibility or side chain volume of the affected amino acid were scored as probably damaging. Substitutions that negatively affected predicted transmembrane regions or that were predicted to have smaller effects on surface accessibility or side chain volume of the affected amino acid on the basis of inferred protein structure were scored as possibly damaging. Conservative substitutions at poorly conserved residues were scored as benign. NA, not applicable.

Nucleotide change	Amino acid change (18)	<i>n</i>	Race	Predicted effect
Low HDL-C <i>ABCA1</i>				
c.593C>A	S198X	1	Black	NA
c.742G>A	P248A	1	Black	Benign
c.1201A>C	K401Q	1	White	Benign
c.1769G>C	W590S*	1	Black	Probably damaging
c.1913G>A	R638Q	1	Black	Possibly damaging
c.2320A>T	T774S*	4	Black	Benign
c.2320A>T	E815G	1	White	Probably damaging
c.2444A>G	S1181F	1	White	Possibly damaging
c.3542C>T	R1341T	1	Black	Possibly damaging
c.4022G>C	S1376G	1	Black	Benign
c.4126A>G	R1615Q	1	Black	Possibly damaging
c.4844G>A	A1670T	1	Black	Possibly damaging
c.5008G>A	N1800H*	1	White	Possibly damaging
c.5398A>C	D2243E	4	Black	Benign
Low HDL-C <i>APOA1</i>				
c.152G>C	R51T	1	Black	Possibly damaging
Low HDL-C <i>LCAT</i>				
None				
High HDL-C <i>ABCA1</i>				
c.1486C>T	R496W	1	White	Probably damaging
c.5039G>A	R1680Q	1	White	Possibly damaging
High HDL-C <i>APOA1</i>				
None				
High HDL-C <i>LCAT</i>				
c.340G>A	V114M	1	Black	Benign

\*These sequence variations have been identified previously.

72% of the nonsynonymous sequence variants identified in the low HDL-C group were likely to adversely affect protein function. In a previous study (12), 69% of nonsynonymous mutations associated with a functional disorder were predicted to be damaging with the use of this algorithm, compared to 32% of nonsynonymous sequence variants identified in DNA samples from arbitrarily selected individuals. Thus, the proportion of nonsynonymous sequence variants predicted to be damaging in the low HDL-C groups with the use of PolyPhen was comparable to that obtained for disease-associated sequence variants (12, 13).

Biochemical measurement of cholesterol efflux rates in cells from Canadian subjects with nonsynonymous sequence variants in *ABCA1* provided further evidence that the sequence variants identified in this study were causally related to low plasma levels of HDL-C (Table 3). Eleven of 14 *ABCA1* sequence variants represented in the low HDL-C group were associated with cholesterol efflux rates that were two standard deviations or more below the mean value obtained in subjects with normal HDL-C levels [i.e.,  $\leq 38\%$  of total cellular cholesterol

per 2 hours (Table 3)]. Conversely, the cholesterol efflux was within the normal range in cells from two individuals in the high HDL-C group who had a missense mutation (Thr<sup>774</sup>→Pro<sup>774</sup>) in *ABCA1*. These findings are consistent with the notion that the excess of sequence variants in subjects with low HDL-C reflects an accumulation of damaging alleles in this group.

The results of this study provide direct evidence that rare variants contribute significantly to low plasma levels of HDL-C, a common quantitative trait. To determine whether any of the common sequence variants identified in this study influenced plasma levels of HDL-C, we tested for associations between sequence variations with frequencies  $> 10\%$  in at least one racial group and plasma levels of HDL-C in the Dallas Heart Study (table S1). Previously we showed that functionally significant sequence variations in *APOE* (14) and *APOA5* (15) are associated with plasma lipid levels in all four major gender and race groups in this population. In contrast to those results, we found no sequence variation in *ABCA1* that was systematically associated with plasma levels of HDL-C in men and women of both race groups

(table S2). Two sequence variants were associated with plasma HDL-C levels in two groups (white men and black men), but the effect of these variants on HDL-C levels was modest. For example, plasma HDL-C concentrations were  $52 \pm 17$  mg/dl in black men homozygous for the Met<sup>883</sup> allele of *ABCA1* and  $48 \pm 15$  mg/dl in homozygotes for the common allele (Ile<sup>883</sup>) at this locus. These data are consistent with haplotype analyses of the *ABCA1* gene (16) and with quantitative trait linkage mapping studies, which have not provided consistent support for common variants at any locus contributing to variation in plasma HDL-C levels. Nonetheless, our results do not exclude a role for common sequence variations in hypoalphalipoproteinemia. Because we screened primarily the coding sequences and flanking intronic regions of the three candidate genes, it is possible that common sequence variations in intronic or regulatory regions not sequenced in this study confer susceptibility to low plasma levels of HDL-C. Alternatively, specific haplotypes or combinations of alleles at these loci or common sequence variants at other loci may contribute to low plasma HDL-C levels.

The strategy used in this study can be generalized to analyze the relationships between sequence variations in candidate genes and other quantitative traits in humans. Screening individuals with low or high levels of a trait increases the likelihood of detecting functionally significant sequence variations. Comparison of the numbers of sequence variants found in individuals with trait levels at opposite ends of the distribution provides a method to assess the significance of the variants identified, including rare variants that are unlikely to be detected by haplotype-based association studies. The method is unbiased, because the selection of individuals has its basis solely in their phenotype and has the added advantage of not requiring ascertainment of families.

**Table 3.** Nonsynonymous sequence variations in *ABCA1*, *APOA1*, and *LCAT* in white Canadians with low ( $n = 155$ ) or high ( $n = 108$ ) plasma HDL-C levels. Cholesterol efflux data for the controls is the mean  $\pm$  SD. ND, not done; NA, not applicable. Functional effects of amino acid substitutions were estimated as described in the legend to Table 2.

Nucleotide change	Amino acid change (18)	n	Predicted effect	Cholesterol efflux (% of total cellular cholesterol/2 hours)
Low HDL-C <i>ABCA1</i>				
c.254C>T	P85L*	1	Probably damaging	0.8
c.917G>A	R306H	1	Benign	ND
c.1375A>C	T459P	1	Possibly damaging	0.28
c.1651C>G	H551D	1	Probably damaging	0.32
c.1769G>T	W590L	1	Probably damaging	0.31
c.2893T>C	R965C	1	Probably damaging	0.59
c.3077T>C	L1026P	1	Benign	0.25
c.3966G>A	W1322X	1	NA	0.38
c.4156G>C	E1386Q	1	Benign	0.51
c.4430G>T	C1477F	1	Probably damaging	0.34
c.5116G>A	D1706N	1	Possibly damaging	0.38
c.5192C>G	S1731C*	1	Possibly damaging	0.28
c.5398A>G	N1800H*	1	Possibly damaging	0.27
c.5864C>T	R1851X*	1	NA	0.26
c.6217A>G	T2073A	1	Possibly damaging	0.28
Controls ( $n = 46$ )				0.52 $\pm$ 0.07
Low HDL-C <i>APOA1</i>				
None				
Low HDL-C <i>LCAT</i>				
c.167T>C	L56P	1	Possibly damaging	ND
c.254G>A	W85X	1	NA	ND
c.382G>A	G128S	1	Probably damaging	ND
c.463A>G	N155D	1	Possibly damaging	ND
c.901G>A	D301N	1	Benign	ND
c.1103G>T	G396V	1	Probably damaging	ND
High HDL-C <i>ABCA1</i>				
c.1664A>C	K555T	1	Benign	ND
c.2320A>C	T774P*	2	Benign	0.63/0.57
High HDL-C <i>APOA1</i>				
None				
High HDL-C <i>LCAT</i>				
None				

\*These sequence variations have been identified previously.

References and Notes

1. K. E. Lohmueller, C. L. Pearce, M. Pike, E. S. Lander, J. N. Hirschhorn, *Nat. Genet.* **33**, 177 (2003).
2. J. K. Pritchard, *Am. J. Hum. Genet.* **69**, 124 (2001).
3. D. E. Reich, E. S. Lander, *Trends Genet.* **17**, 502 (2001).
4. G. Assmann, A. von Eckardstein, H. Funke, *Circulation* **87**, III28 (1993).
5. M. R. Hayden et al., *Curr. Opin. Lipidol.* **11**, 117 (2000).
6. J. A. Kuivenhoven et al., *J. Lipid Res.* **38**, 191 (1997).
7. G. K. Hovingh et al., *J. Lipid Res.* **44**, 1251 (2003).
8. S. Mott et al., *Atherosclerosis* **152**, 457 (2000).
9. R. G. Victor et al., *Am. J. Cardiol.* **93**, 1473 (2004).
10. R. R. Singaraja, L. R. Brunham, H. Visscher, J. J. Kastelein, M. R. Hayden, *Arterioscler. Thromb. Vasc. Biol.* **23**, 1322 (2003).
11. Information about this program is available at [www.bork.embl-heidelberg.de/PolyPhen/](http://www.bork.embl-heidelberg.de/PolyPhen/).
12. S. Sunyaev et al., *Hum. Mol. Genet.* **10**, 591 (2001).
13. V. Ramensky, P. Bork, S. Sunyaev, *Nucleic Acids Res.* **30**, 3894 (2002).
14. J. C. Cohen, H. H. Hobbs, unpublished observation.
15. L. A. Pennacchio et al., *Hum. Mol. Genet.* **11**, 3031 (2002).
16. D. A. Tregouet et al., *Arterioscler. Thromb. Vasc. Biol.* **24**, 775 (2004).
17. Details of each sequence variant are provided at <http://pga.swmed.edu>.
18. Single-letter abbreviations for the amino acid resi-

dues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

19. A detailed explanation of PolyPhen scoring criteria is available at [http://tux.embl-heidelberg.de/ramensky/doc/pph\\_help.html](http://tux.embl-heidelberg.de/ramensky/doc/pph_help.html).

20. We thank R. Wilson, S. Niu, and J. Horton for DNA sequencing; J. Schageman, N. Prikhodko, and A.

Patel for analysis of the sequence data; and M. Brown, J. Goldstein, and S. Sunyaev for helpful discussions. This study was supported by grants from NIH (HL53917), National Heart, Lung, and Blood Institute Program for Genomic Applications (UO1-HL66880), Fondation Le Ducq, Heart and Stroke Foundation of Canada, and Canadian Institutes of Health Research (64519, CIHR-RD, and DOP-51714).

**Supporting Online Material**  
[www.sciencemag.org/cgi/content/full/305/5685/869/DC1](http://www.sciencemag.org/cgi/content/full/305/5685/869/DC1)  
 Materials and Methods  
 Fig. S1  
 Tables S1 and S2

3 May 2004; accepted 12 July 2004

# HLA and NK Cell Inhibitory Receptor Genes in Resolving Hepatitis C Virus Infection

Salim I. Khakoo,<sup>1\*</sup> Chloe L. Thio,<sup>2</sup> Maureen P. Martin,<sup>3</sup> Collin R. Brooks,<sup>1</sup> Xiaojiang Gao,<sup>3</sup> Jacquie Astemborski,<sup>2</sup> Jie Cheng,<sup>3</sup> James J. Goedert,<sup>4</sup> David Vlahov,<sup>5</sup> Margaret Hilgartner,<sup>6</sup> Steven Cox,<sup>7</sup> Ann-Margeret Little,<sup>7</sup> Graeme J. Alexander,<sup>8</sup> Matthew E. Cramp,<sup>9</sup> Stephen J. O'Brien,<sup>10</sup> William M. C. Rosenberg,<sup>1</sup> David L. Thomas,<sup>2</sup> Mary Carrington<sup>3\*</sup>

Natural killer (NK) cells provide a central defense against viral infection by using inhibitory and activation receptors for major histocompatibility complex class I molecules as a means of controlling their activity. We show that genes encoding the inhibitory NK cell receptor KIR2DL3 and its human leukocyte antigen C group 1 (HLA-C1) ligand directly influence resolution of hepatitis C virus (HCV) infection. This effect was observed in Caucasians and African Americans with expected low infectious doses of HCV but not in those with high-dose exposure, in whom the innate immune response is likely overwhelmed. The data strongly suggest that inhibitory NK cell interactions are important in determining antiviral immunity and that diminished inhibitory responses confer protection against HCV.

Natural killer (NK) cells are key components of the innate antiviral immune response. In vivo, they are under the constitutively dominant influence of inhibitory receptors for self-MHC class I ligands (1, 2), such that effector functions occur only when activating signals overcome inhibitory signals (3, 4). The killer cell immunoglobulin-like receptors (KIR) represent a diverse family of activating and inhibitory receptors that are integral in this model. As with their MHC class I ligands, the population diversity and rapid evolution of

the KIR genes strongly suggests that they are under pathogen-mediated selection (5–7). KIR haplotypes vary in number and map to separate chromosomes, and because HLA and KIR map to separate chromosomes, some individuals lack specific KIR-HLA receptor-ligand pairings. To date, only activating KIR have been associated with disease outcome (8–10), whereas the influence of inhibitory KIR on disease is undetermined. Hepatitis C virus (HCV) is a common infection worldwide, causing cirrhosis and hepatocellular carcinoma. About 20% of individuals

resolve acute infection, an outcome associated with specific components of the adaptive immune system (11), including HLA class I (12). Because resolution of HCV infection may also involve the innate immune system, including NK cells (13, 14), we examined the possible synergistic influence that corresponding KIR-HLA combinations might have on the outcome of HCV infection.

Individuals who were exposed to HCV (685 with persistent and 352 with resolved infection) (table S1, A to C) were categorized according to their KIR-binding motifs based on HLA-B and -C genotyping data (15). Group 1 HLA-C (HLA-C1) allotypes have asparagine at residue 80 and are ligands for the inhibitory receptors KIR2DL2 and KIR2DL3, which segregate as alleles of a single locus (Table 1). The remaining HLA-C allotypes (group 2, HLA-C2) have

<sup>1</sup>Liver Group, Division of Infection, Inflammation, and Repair, Southampton University, Southampton 5016 6YD, UK. <sup>2</sup>Department of Medicine, Johns Hopkins Medical Institutions, Baltimore, MD 21231, USA. <sup>3</sup>Basic Research Program, Scientific Applications International Corporation Frederick, Inc., Laboratory of Genomic Diversity, NCI Frederick, Frederick, MD 21702, USA. <sup>4</sup>Viral Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20852, USA. <sup>5</sup>New York Academy of Medicine, New York, NY 10029, USA. <sup>6</sup>Department of Pediatrics, New York Presbyterian Hospital–Cornell Medical Center, New York, NY 10021, USA. <sup>7</sup>Anthony Nolan Research Institute, The Royal Free Hospital, London NW3 2PF, UK. <sup>8</sup>Department of Medicine, University of Cambridge, Cambridge CD2 2QQ, UK. <sup>9</sup>Department of Gastroenterology, Derriford Hospital, Plymouth PL6 8DH, UK. <sup>10</sup>Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702, USA.

\*To whom correspondence should be addressed: E-mail: [carringt@ncifcr.gov](mailto:carringt@ncifcr.gov) (M.C.); [sik@soton.ac.uk](mailto:sik@soton.ac.uk) (S.I.K.)

**Table 1.** Frequency of KIR and HLA receptor-ligand pairings in the population studied, stratified by race, study site, and route of infection.

	N	KIR2DL1- HLA-C2 N (%)	KIR2DL2- HLA-C1 N (%)	KIR2DL3- HLA-C1 N (%)	KIR2DS1- HLA-C2 N (%)	KIR2DS2- HLA-C1* N (%)	KIR3DL1- HLA-Bw4 N (%)	KIR3DS1- HLA-Bw4* N (%)
All	1037	689 (66.4)	591 (57.0)	754 (72.7)	231 (22.3)	441 (42.5)	635 (61.2)	216 (20.8)
				Race†				
UK Caucasian	340	219 (64.4)	144 (42.4)	271 (79.7)	78 (22.9)	144 (42.3)	205 (60.1)	83 (24.4)
USA Caucasian	355	220 (62.0)	163 (45.9)	265 (74.6)	88 (24.8)	167 (47.0)	205 (57.7)	89 (25.1)
USA African-American	271	205 (75.6)	108 (39.9)	166 (61.3)	47 (17.3)	99 (36.5)	188 (69.4)	31 (11.4)
USA other	69	44 (63.8)	30 (43.5)	52 (75.4)	17 (24.6)	30 (43.5)	35 (50.7)	12 (17.4)
				Route				
No blood products	543	372 (68.5)	229 (42.2)	382 (70.3)	115 (21.2)	222 (40.9)	344 (63.3)	105 (19.3)
Blood products	494	317 (64.2)	217 (43.9)	372 (75.3)	116 (23.5)	219 (44.3)	291 (59.0)	111 (22.5)

\*Receptor ligand pairing inferred by protein sequence but not directly demonstrated. †Excludes two non-Caucasian individuals from the United Kingdom.