Table 2. Fifty Best *P* Values Obtained for the Comparison between Rapid Progressors and Control Subjects

This table is available in its entirety in the online version of the *Journal of Infectious Diseases*.

and/or the alleles in linkage disequilibrium are associated with prevention of rapid progression, with an odds ratio (OR) of 0.24 [95% confidence interval [CI], 0.12–0.46] (Table 1). Another association was identified on chromosome 1 in *RXRG* gene ($P = 3.86 \times 10^{-6}$, OR, 3.29 [95% CI, 2.08–5.20]). Finally, SNPs modulating rapid progression were also found in *SOX5* ($P = 1.80 \times 10^{-6}$; OR, 0.45 [95% CI, 0.32–0.63]) and *TGFBRAP1* ($P = 7.04 \times 10^{-6}$; OR, 0.34 [95% CI, 0.20–0.57]). Two other independent SNPs, rs1360517 and rs3108919, met the FDR threshold of 25% (Table 1) but were not close to any known gene (distance, <20 kb).

To complete our analysis, we explored the influence of covariables—such as *CCR5*-Δ32 and *CCR5-P1* haplotypes [7], the HIV-1 infection mode (mucosal or parenteral), age at seroconversion, and sex—on all the associations presented in Table 1. None of them was found to affect the observed signals.

To deepen our genomic analysis, we also tried to combine our results with previously published data from the EuroCHAVI cohort, which assessed the viral load end point [3]. We combined their *P* values with ours using the classical Fisher method. Unfortunately, no combined *P* values met the FDR threshold of 25% (data not shown; best combined *P* value, 6.23×10^{-5}). The lack of significant common signals between the 2 studies may likely stem from the difference in inclusion criteria, in particular the use of viral load versus the use of

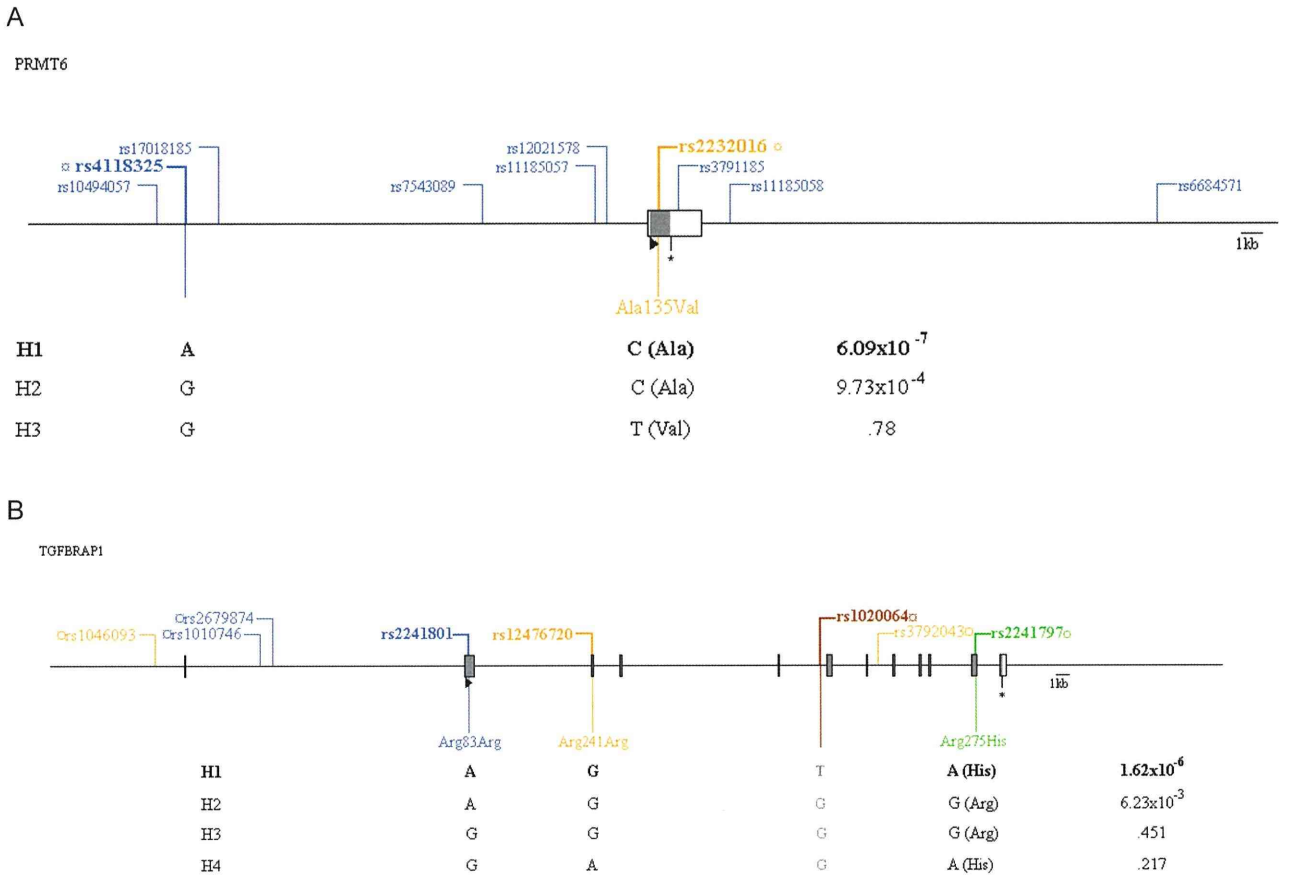

Figure 4. Haplotype maps for *PRMT6* (*A*) and *TGFBRAP1* (*B*). Exons and untranslated regions are symbolized by shaded and unshaded boxes, respectively. The positions of the ATG and stop codons are indicated by a triangle (▶) and by an asterisk (*), respectively. Single-nucleotide polymorphisms (SNPs) in high linkage disequilibrium ($r^2 > 0.9$) are represented with the same color, and the genotyped tagSNPs are marked with the symbol ¤. For greater clarity, only the exonic polymorphisms in linkage disequilibrium with the genotyped tagSNPs ($r^2 > 0.9$) are specified in panel B. The numbers in the right column correspond to the *P* values obtained when comparing the allelic frequency of each haplotype between rapid progressors and control subjects.

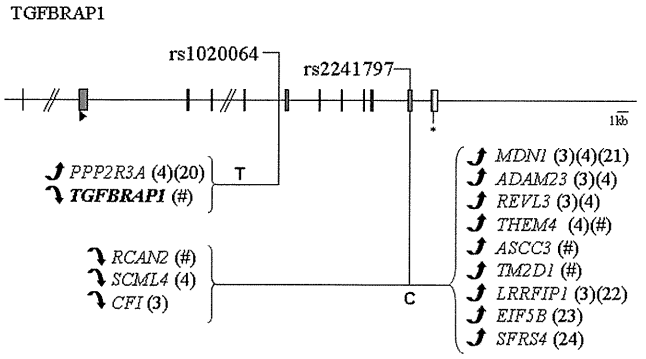

Figure 5. Correlation between some *TGFBRAP1* polymorphisms and differential gene expression according to the Genevar database [18] and the Dixon database [19]. Exons and untranslated regions are symbolized by shaded and unshaded boxes, respectively. The positions of the ATG and stop codons are indicated by a triangle (▶) and by an asterisk (*), respectively. The modulation of gene expression is indicated by the arrow direction: increased (↑) or decreased (↓) expression. The numbers correspond to the bibliographic references of previous works linking these genes to AIDS. The present genomewide association study of rapid progression is indicated by a pound sign (#).

CD4 T cell count, and also from the elimination of very rapid progressors in the Euro-CHAVI study, since such patients could not exhibit a sufficiently prolonged viral load set point [3].

In the past, we have often observed that haplotypes may be more informative than individual SNPs; this was notably shown for *CCR5* [7], *CXCR1* [6], and *HLA* [5]. We thus decided to explore the haplotype patterns for the genes exhibiting the best signals in our rapid progression GWAS, limiting the investigation to exonic and promoter SNPs (Figure 4). For *PRMT6*, we demonstrated that the effect of the *PRMT6* SNP rs4118325 could be tracking a haplotype (composed of rs4118325 and rs2232016 [Ala135Val]) through linkage disequilibrium (r^2 = 0.12, D' = 1) (Figure 4*A*). No differential messenger RNA (mRNA) expression could be significantly associated with either of these SNPs using the Genevar database [18] or the Dixon database [19].

For *TGFBRAP1*, the rs1020064 SNP was in high linkage disequilibrium (r^2 = 0.97) with a haplotype containing 3 exonic SNPs (rs2241801 [Arg83Arg], rs12476720 [Arg241Arg], and rs2241797 [Arg275His]) (Figure 4*B*). The SNP rs2241797 (Arg275His) appeared to be essential, since the signal disappeared when that SNP was removed from the haplotype ($P > 5.0 \times 10^{-2}$) and remained identical when either of the 2 synonymous SNPs were removed ($P = 1.62 \times 10^{-6}$). Interestingly, rs1020064 has been associated with differential expression of *TGFBRAP1* in the Genevar database and of *PPP2R3A* [20] in the Dixon database (Figure 5). More strikingly, in the Dixon database the nonsynonymous SNP rs2241797 was significantly associated with the differential expression of several proteins (Figure 5), among which 4 have been independently described to interact with HIV-1 (*MDN1* [21], *LRRFIP1* [22], *EIF5B* [23], and *SFRS4* [24]), and several have exhibited a positive association in an AIDS GWAS. Of note, no differential expression could be found for the 2 synonymous SNPs rs2241801 and rs12476720. For the *SOX5* rs1522232 and *RXRG* rs10800098 SNPs, no haplotype involving exonic or promoter SNPs and no differential mRNA expression could be found.

DISCUSSION

For the first time, the extreme HIV-1 rapid progression phenotype was specifically examined in a GWAS. The power of using the GRIV extreme design has been demonstrated by previous candidate gene studies [5–7, 25] and by the previous GWAS of nonprogressors [4]. Here we have identified 6 novel associations with rapid progression with ORs as high as 4, emphasizing the power of this extreme design in spite of a relatively modest sample size. As a comparison, there was only 1 independent ($r^2 < 0.5$) signal passing the FDR threshold of 25% in the GWAS of the nonprogressor GRIV cohort [4]. No replication was readily available, since this specific rapid progression design is rather unique in the world. The biological relevance of the rapid progression associations and their statistical validation through simulations with a second control group (<1% odds of finding as many independent signals) are nevertheless compelling. The use of an extreme population may provide a strong contrast and indeed help unravel new genetic factors, behaving as a magnifying glass [25, 26].

Four of the 6 SNPs associated with rapid progression to AIDS were clearly linked to a gene (distance, <2 kb). *SOX5* (OR, 0.45) encodes a transcription regulator notably expressed in lymph nodes and lymphoid tissues [27] and is also known to be involved in the transforming growth factor β (TGF-β)/SMAD chondrogenesis signaling pathway [28, 29]. There is no other experimental evidence linking this protein to the pathogenesis of HIV-1 infection. *RXRG* (OR, 3.29) encodes a retinoic acid nuclear receptor mediating the antiproliferative effects of retinoic acid and is known to repress HIV-1 transcription and replication [30, 31]. Several studies have also associated vitamin A deficiency with a bad prognosis in patients with AIDS, but the benefit of vitamin A supplementation in AIDS patients is still controversial [32, 33]. The genetic association for *PRMT6* (OR, 0.24) points toward a direct interaction between the host and the virus. Indeed, *PRMT6* codes for an arginine *N*-methyltransferase previously reported to methylate HIV-1 Tat and Rev, impairing their function [34, 35]. The product of *PRMT6* also methylates the high mobility group protein HMGA1, thereby modifying HMGA1 interactions with DNA [36], which could alter HIV-1 integration into the human genome [37]. The modulation of *PRMT6* transcription could

thus affect HIV-1 integration or replication and prevent rapid progression. Finally, an association with rapid progression was discovered for *TGFBRAP1* (OR, 0.34). *TGFBRAP1* is expressed in most lymphoid tissues and is involved in the TGF-β signaling pathway [38]. TGF-β is a pleiotropic immunosuppressive cytokine involved in immune homeostasis and in the differentiation of and balance between type 17 T helper (Th17) cells (T cells protecting the mucosal barrier integrity [39]) and regulatory T (T_{reg}) cells (T cells essential for immune suppression [40]). Interestingly, recent works have also supported a combined role for TGF-β and retinoic acid for T_{reg} cell differentiation [41, 42]. T_{reg} cells are rapidly induced after simian immunodeficiency virus (SIV) and HIV-1 infection [43, 44] and may have a deleterious effect during the chronic phase of infection [45, 46]. Moreover, an impairment in the balance between Th17 and T_{reg} cells during HIV-1 infection was recently described [40, 47, 48]. Finally, TGF-β can stimulate HIV-1 Tat transcription [49] and, reciprocally, be induced by Tat early during infection [50], supporting a key role for the intermingled effects of TGF-β and Tat in the early development of HIV-1 infection. Overall, our results support the central role played by the TGF-β pathway and the balance between Th17 and T_{reg} cells in AIDS progression. Interestingly, *PRMT6* and *TGFBRAP1* haplotypes involving promoter and exonic SNPs were also associated with disease progression; moreover, some of the haplotype SNPs could be associated with downstream differential mRNA expression (Figure 5).

In conclusion, as for all genetic studies, our results—including SNPs with higher *P* values, which were not discussed—will need to be confirmed by replication in other cohorts and by other investigations, such as fine gene mapping and biological experiments.

ANRS GENOMIC GROUP

The ANRS Genomic Group is composed of Prof Jean-François Delfraissy (Agence Nationale de Recherche sur le SIDA, Paris), Dr Laurence Meyer (Hôpital Kremin-Bicêtre, France), Prof Philippe Broët (Hôpital Kremlin-Bicêtre, France), Dr Cyril Dalmasso (Hôpital Kremlin-Bicêtre, France), Prof Patrice Debré (Hôpital La Salpétrière, Paris), Dr Ioannis Théodorou (Hôpital La Salpétrière, Paris), Prof Christine Rouzioux (Hôpital Necker, Paris), Cédric Coulonges (Conservatoire National des Arts et Métiers, Paris), Sigrid Le Clerc (Conservatoire National des Arts et Métiers, Paris), Sophie Limou (Conservatoire National des Arts et Métiers, Paris), and Prof Jean-François Zagury (Conservatoire National des Arts et Métiers, Paris).

Acknowledgments

We are grateful to all the patients and medical staff who have kindly collaborated with the GRIV project.

References

1. Kingsmore SF, Lindquist IE, Mudge J, Gessler DD, Beavis WD. Genome-wide association studies: progress and potential for drug discovery and development. Nat Rev Drug Discov **2008**; 7:221–30.
2. Dalmasso C, Carpentier W, Meyer L, et al. Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. PLoS ONE **2008**; 3:e3907.
3. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. Science **2007**; 317: 944–7.
4. Limou S, Le Clerc S, Coulonges C, et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by *HLA* genes (ANRS Genomewide Association Study 02). J Infect Dis **2009**; 199:419–26.
5. Flores-Villanueva PO, Hendel H, Caillat-Zucman S, et al. Associations of MHC ancestral haplotypes with resistance/susceptibility to AIDS disease development. J Immunol **2003**; 170:1925–9.
6. Vasilescu A, Terashima Y, Enomoto M, et al. A haplotype of the human CXCR1 gene protective against rapid disease progression in HIV-1+ patients. Proc Natl Acad Sci U S A **2007**; 104:3354–9.
7. Winkler CA, Hendel H, Carrington M, et al. Dominant effects of CCR2-CCR5 haplotypes in HIV-1 disease progression. J Acquir Immune Defic Syndr **2004**; 37:1534–8.
8. Hercberg S, Galan P, Preziosi P, et al. Background and rationale behind the SU.VI.MAX Study, a prevention trial using nutritional doses of a combination of antioxidant vitamins and minerals to reduce cardiovascular diseases and cancers: SUpplementation en VItamines et Mineraux AntioXydants Study. Int J Vitam Nutr Res **1998**; 68:3–20.
9. Balkau B. An epidemiologic survey from a network of French Health Examination Centres, (D.E.S.I.R.): epidemiologic data on the insulin resistance syndrome. Rev Epidemiol Sante Publique **1996**; 44:373–5.
10. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet **2007**; 81:559–75.
11. Delaneau O, Coulonges C, Zagury JF. Shape-IT: new rapid and accurate algorithm for haplotype inference. BMC Bioinformatics **2008**; 9:540.
12. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B **1995**; 57:289–300.
13. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stati Med **1990**; 9:811–8.
14. Perneger TV. What's wrong with Bonferroni adjustments. BMJ **1998**; 316:1236–8.
15. Storey JD, Tibshirani R. Statistical significance for genomewide studies. PNAS **2003**; 100:9440–5.
16. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics **2000**; 155:945–59.
17. Devlin B, Roeder K. Genomic control for association studies. Biometrics **1999**; 55:997–1004.
18. Ge D, Zhang K, Need AC, et al. WGAViewer: software for genomic annotation of whole genome association studies. Genome Res **2008**; 18:640–3.
19. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. Nat Genet **2007**; 39:1202–7.
20. Ammosova T, Washington K, Debebe Z, Brady J, Nekhai S. Dephosphorylation of CDK9 by protein phosphatase 2A and protein phosphatase-1 in Tat-activated HIV-1 transcription. Retrovirology **2005**; 2:47.
21. Brass AL, Dykxhoorn DM, Benita Y, et al. Identification of host proteins required for HIV infection through a functional genomic screen. Science **2008**; 319:921–6.
22. Wilson SA, Brown EC, Kingsman AJ, Kingsman SM. TRIP: a novel double stranded RNA binding protein which interacts with the leucine rich repeat of flightless I. Nucleic Acids Res **1998**; 26:3460–7.
23. Wilson SA, Sieiro-Vazquez C, Edwards NJ, et al. Cloning and characterization of hIF2, a human homologue of bacterial translation initi-

ation factor 2, and its interaction with HIV-1 matrix. Biochem J **1999**; 342:97–103.
24. Exline CM, Feng Z, Stoltzfus CM. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol **2008**; 82:3921–31.
25. Huber C, Pons O, Hendel H, et al. Genomic studies in AIDS: problems and answers—development of a statistical model integrating both longitudinal cohort studies and transversal observations of extreme cases. Biomed Pharmacother **2003**; 57:25–33.
26. Froguel P, Blakemore AI. The power of the extreme in elucidating obesity. N Engl J Med **2008**; 359:891–3.
27. Su AI, Cooke MP, Ching KA, et al. Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A **2002**; 99:4465–70.
28. Furumatsu T, Tsuda M, Taniguchi N, Tajima Y, Asahara H. Smad3 induces chondrogenesis through the activation of SOX9 via CREB-binding protein/p300 recruitment. J Biol Chem **2005**; 280:8343–50.
29. Ikeda T, Kawaguchi H, Kamekura S, et al. Distinct roles of Sox5, Sox6, and Sox9 in different stages of chondrogenic differentiation. J Bone Miner Metab **2005**; 23:337–40.
30. Kiefer HL, Hanley TM, Marcello JE, Karthik AG, Viglianti GA. Retinoic acid inhibition of chromatin remodeling at the human immunodeficiency virus type 1 promoter: uncoupling of histone acetylation and chromatin remodeling. J Biol Chem **2004**; 279:43604–13.
31. Maeda Y, Yamaguchi T, Hijikata Y, et al. All-trans retinoic acid attacks reverse transcriptase resulting in inhibition of HIV-1 replication. Hematology **2007**; 12:263–6.
32. Austin J, Singhal N, Voigt R, et al. A community randomized controlled clinical trial of mixed carotenoids and micronutrient supplementation of patients with acquired immunodeficiency syndrome. Eur J Clin Nutr **2006**; 60:1266–76.
33. Mehta S, Fawzi W. Effects of vitamins, including vitamin A, on HIV/AIDS patients. Vitam Horm **2007**; 75:355–83.
34. Invernizzi CF, Xie B, Richard S, Wainberg MA. PRMT6 diminishes HIV-1 Rev binding to and export of viral RNA. Retrovirology **2006**; 3:93.
35. Xie B, Invernizzi CF, Richard S, Wainberg MA. Arginine methylation of the human immunodeficiency virus type 1 Tat protein by PRMT6 negatively affects Tat interactions with both cyclin T1 and the Tat transactivation region. J Virol **2007**; 81:4226–34.
36. Sgarra R, Lee J, Tessari MA, et al. The AT-hook of the chromatin architectural transcription factor high mobility group A1a is arginine-methylated by protein arginine methyltransferase 6. J Biol Chem **2006**; 281: 3764–72.
37. Li L, Yoder K, Hansen MS, Olvera J, Miller MD, Bushman FD. Retroviral cDNA integration: stimulation by HMG I family proteins. J Virol **2000**; 74:10965–74.
38. Charng MJ, Zhang D, Kinnunen P, Schneider MD. A novel protein distinguishes between quiescent and activated forms of the type I transforming growth factor beta receptor. J Biol Chem **1998**; 273:9365–8.
39. Manel N, Unutmaz D, Littman DR. The differentiation of human T_H-17 cells requires transforming growth factor-beta and induction of the nuclear receptor RORgammat. Nat Immunol **2008**; 9:641–9.
40. de St Groth BF, Landay AL. Regulatory T cells in HIV infection: pathogenic or protective participants in the immune response? AIDS **2008**; 22:671–83.
41. Benson MJ, Pino-Lagos K, Rosemblatt M, Noelle RJ. All-trans retinoic acid mediates enhanced T reg cell growth, differentiation, and gut homing in the face of high levels of co-stimulation. J Exp Med **2007**; 204:1765–74.
42. Mucida D, Park Y, Kim G, et al. Reciprocal TH17 and regulatory T cell differentiation mediated by retinoic acid. Science **2007**; 317:256–60.
43. Estes JD, Wietgrefe S, Schacker T, et al. Simian immunodeficiency virus–induced lymphatic tissue fibrosis is mediated by transforming growth factor beta 1-positive regulatory T cells and begins in early infection. J Infect Dis **2007**; 195:551–61.
44. Kekow J, Wachsman W, McCutchan JA, Cronin M, Carson DA, Lotz M. Transforming growth factor beta and noncytopathic mechanisms of immunodeficiency in human immunodeficiency virus infection. Proc Natl Acad Sci U S A **1990**; 87:8321–5.
45. Kared H, Lelievre JD, Donkova-Petrini V, et al. HIV-specific regulatory T cells are associated with higher CD4 cell counts in primary infection. AIDS **2008**; 22:2451–60.
46. Weiss L, Donkova-Petrini V, Caccavelli L, Balbo M, Carbonneil C, Levy Y. Human immunodeficiency virus-driven expansion of CD4+CD25+ regulatory T cells, which suppress HIV-specific CD4 T-cell responses in HIV-infected patients. Blood **2004**; 104:3249–56.
47. Brenchley JM, Paiardini M, Knox KS, et al. Differential Th17 CD4 T-cell depletion in pathogenic and nonpathogenic lentiviral infections. Blood **2008**; 112:2826–35.
48. Favre D, Lederer S, Kanwar B, et al. Critical loss of the balance between Th17 and T regulatory cell populations in pathogenic SIV infection. PLoS Pathog **2009**; 5:e1000295.
49. Li JM, Shen X, Hu PP, Wang XF. Transforming growth factor beta stimulates the human immunodeficiency virus 1 enhancer and requires NF-kappaB activity. Mol Cell Biol **1998**; 18:110–21.
50. Zocchi MR, Contini P, Alfano M, Poggi A. Pertussis toxin (PTX) B subunit and the nontoxic PTX mutant PT9K/129G inhibit Tat-induced TGF-beta production by NK cells and TGF-beta–mediated NK cell apoptosis. J Immunol **2005**; 174:6054–61.

A Genome-Wide Association Analysis Identified a Novel Susceptible Locus for Pathological Myopia at 11q24.1

Hideo Nakanishi[1,2], Ryo Yamada[2,3], Norimoto Gotoh[1,2], Hisako Hayashi[1,2], Kenji Yamashiro[1], Noriaki Shimada[4], Kyoko Ohno-Matsui[4], Manabu Mochizuki[4], Masaaki Saito[5], Tomohiro Iida[5], Keitaro Matsuo[6], Kazuo Tajima[7], Nagahisa Yoshimura[1]*, Fumihiko Matsuda[2,8]*

1 Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan, **2** Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **3** Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan, **4** Department of Ophthalmology and Visual Science, Tokyo Medical and Dental University Graduate School of Medicine, Tokyo, Japan, **5** Department of Ophthalmology, Fukushima Medical University, Fukushima, Japan, **6** Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan, **7** Aichi Cancer Center Research Institute, Nagoya, Japan, **8** CEA/Institute de Genomique, Centre National de Génotypage, Evry, France

Abstract

Myopia is one of the most common ocular disorders worldwide. Pathological myopia, also called high myopia, comprises 1% to 5% of the general population and is one of the leading causes of legal blindness in developed countries. To identify genetic determinants associated with pathological myopia in Japanese, we conducted a genome-wide association study, analyzing 411,777 SNPs with 830 cases and 1,911 general population controls in a two-stage design (297 cases and 934 controls in the first stage and 533 cases and 977 controls in the second stage). We selected 22 SNPs that showed *P*-values smaller than 10^{-4} in the first stage and tested them for association in the second stage. The meta-analysis combining the first and second stages identified an SNP, rs577948, at chromosome 11q24.1, which was associated with the disease ($P=2.22\times10^{-7}$ and OR of 1.37 with 95% confidence interval: 1.21–1.54). Two genes, *BLID* and *LOC399959*, were identified within a 200-kb DNA encompassing rs577948. RT–PCR analysis demonstrated that both genes were expressed in human retinal tissue. Our results strongly suggest that the region at 11q24.1 is a novel susceptibility locus for pathological myopia in Japanese.

Citation: Nakanishi H, Yamada R, Gotoh N, Hayashi H, Yamashiro K, et al. (2009) A Genome-Wide Association Analysis Identified a Novel Susceptible Locus for Pathological Myopia at 11q24.1. PLoS Genet 5(9): e1000660. doi:10.1371/journal.pgen.1000660

Editor: Takashi Gojobori, National Institute of Genetics, Japan

Received May 26, 2009; **Accepted** August 24, 2009; **Published** September 25, 2009

Funding: The study was supported in part by grants-in-aid for scientific research (19390442 and 27091294) from the Japan Society for the Promotion of Science, Tokyo, Japan; by Okawa Foundation for Information and Telecommunications; and by the Japanese National Society for the Prevention of Blindness. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: nagaeye@kuhp.kyoto-u.ac.jp (NY); fumi@genome.med.kyoto-u.ac.jp (FM)

Introduction

Myopia is a refractive error (http://en.wikipedia.org/wiki/Refractive_error) of the eye in which parallel rays of light focus in a plane anterior to the retina resulting in blurred vision. Myopia is one of the most common ocular disorders worldwide, and is in much higher prevalence in Asians than in Caucasians. Recent population-based surveys in the elderly reported that the prevalence of myopia was approximately 25% in the Caucasian populations [1] and 40% in the East Asian (Chinese and Japanese) populations [2,3].

Myopia is divided into two distinct subsets, namely, common and pathological myopia. Pathological myopia, also called high myopia, is distinguished from common myopia, also called low/moderate myopia, by excessive increase in axial length of the eyeball, which is the most important contributor to the myopic refraction [4,5]. The axial length of the eyeball in adults is approximately 24 mm, and its elongation by 1 mm without other compensatory changes results in a myopic shift of −2.5 to −3.0 diopters (D). It has been shown that distribution of the axial lengths of the adult myopic population is bimodal [6], and the subgroup with elongated axial length in the bimodal distribution corresponds to pathological myopia. This group comprises 1% to 5% of the population [3,7], and is commonly defined by axial length greater than 26.0 mm which is equivalent to refractive errors greater than −6 D [8].

The excessive elongation of the eyeball causes mechanical strain with subsequent degenerative changes of the retina, choroid, and sclera. The degenerative changes at the posterior pole of the eye such as chorioretinal atrophy or posterior staphyloma are clinically important and unique to pathological myopia [9]. These unique degenerative changes at the posterior pole result in uncorrectable visual impairment due to decreased central vision and make pathological myopia one of the leading causes of legal blindness in developed countries [10–13].

It has been reported that not only environmental factors, such as near work and higher education, but also genetic factors contribute to the development of myopia, in particular, of pathological myopia [14]. Previous twin studies reported that the estimated heritability of refractive error and axial length is up to 0.90 [15,16], although that might be overestimated due to common environmental effects [17]. Multiple family-based whole genome linkage analyses of myopia reported at least 16 susceptible chromosomal loci (MYP1–16 in OMIM database; 10 loci for

Author Summary

Myopia is one of the most common ocular disorders with elongation of axis of the eyeball. Pathological myopia or high myopia, a subset of myopia which is characterized with excessive axial elongation and degenerative changes of the eye, is a leading cause of visual impairment. Since genetic factors play significant roles in its development, identification of genetic determinants is an urgent and important issue. Although family-based linkage analyses have isolated at least 16 susceptible chromosomal loci for pathological or common myopia, no gene responsible for the disease has been identified. We conducted the first genome-wide case/control association study of pathological myopia in a two-stage design using 411,777 markers with 830 Japanese patients and 1,911 Japanese controls. We identified a region strongly suggestive for the disease susceptibility at chromosome 11q24.1 containing *BLID* and *LOC399959*. Their expression was confirmed in human retina with RT–PCR. *BLID* encodes an inducer of apoptotic cell death, and apoptosis is known to play an important functional role in pathological myopia. We believe that our study contributes to further dissect the molecular events underlying the development and progression of pathological myopia.

pathological myopia [18–27] and 6 for common myopia [28–30]). Among them, at least 8 chromosomal loci, such as 12q21–23 (MYP3), 22q12 (MYP6) and 2q37.1 (MYP12) were successfully validated by at least two independent studies [31,32]. However, no genes responsible for the disease have been identified.

The genome-wide association (GWA) study using single nucleotide polymorphisms (SNPs) as markers is an alternative approach to identify genetic risk factors of common diseases. This approach has been successfully applied to identify genetic risk factors for multigenetic diseases including ophthalmic diseases such as age-related macular degeneration [33,34] and exfoliation syndrome [35]. To identify the genetic risk factors of pathological myopia, we conducted a two-stage GWA-based case/control association analysis using 411,777 markers with 830 Japanese patients and 1,911 Japanese controls (297 cases and 934 controls in the first stage, and 533 cases and 977 controls in the second stage).

Results

Characterization of the patients with pathological myopia

A total of 839 pathological myopic patients with axial length greater than 26.0 mm in both eyes were enrolled in the current study. In order to maximize the detection power, patients with axial length greater than 28.0 mm in both eyes were enrolled in the first stage of genome scan. No other clinical features were accounted for the assignment of patients to either stage. 824 out of 839 patients (98.2%) had degenerative changes specific to pathological myopia. Other features of cases and controls who passed quality control procedures of genotyping results (see Materials and Methods) were summarized in Table 1.

Genome-wide association analysis

For the first stage, we scanned the genome of 302 cases using the Illumina HumanHap550 BeadChip, which launches 561,466 relatively frequent SNPs (minor allele frequency >0.05) distributed across the human genome at an average interval of 6.5 kilobases (kb). Five cases and 149,689 SNPs were excluded due to quality control criteria (see details in Materials and Methods) and genotyping results of 411,777 SNPs in autosomes for 297 cases were used for the statistical analysis. They were compared with 934 controls from the JSNP database [36] for association with phenotype using χ^2 test for trend. Genomic Control (GC) method [37] revealed only a slight inflation of the test statistics (GC parameter $\lambda = 1.068$). We identified 29 SNPs in 22 chromosomal regions with *P*-value adjusted by GC being smaller than 10^{-4} (Figure 1 and Table S1). Among them, seven SNPs at chromosome 8p12 were in strong linkage disequilibrium (LD) and likewise two SNPs at chromosome 10q22.2 (pair-wise $D' > 0.95$ and $r^2 > 0.9$). Thus, we selected one representing SNP from each region and tested 22 SNPs in the second stage.

Table 1. Characteristics of the study population used in the study[a].

Cases/Controls	Category	Subcategory	First stage	Second stage
Cases: Patients with pathological myopia	Number		297	533
	Age (years)		58.8±13.2	59.0±14.3
	Gender	Male	93	171
		Female	204	362
	Axial length (mm)	Right eyes	29.97±1.36	29.04±1.97
		Left eyes	29.84±1.37	28.91±1.89
	Refraction of the phakic eyes (Diopter)[b]	Right eyes	−14.94±4.04	−12.40±4.48
		Left eyes	−14.64±3.98	−12.07±4.72
Controls: general Japanese population	Number		934	977
	Age (years)		NA	48.3±16.3
	Gender	Male	NA	497
		Female	NA	480

The ± sign is a standard deviation.

[a]The study population after quality control procedures.

[b]For the calculations of refraction, 177 eyes (29.8%) in the first stage and 303 eyes (28.4%) in the second stage that had undergone cataract surgery or corneal refractive surgery were excluded.

doi:10.1371/journal.pgen.1000660.t001

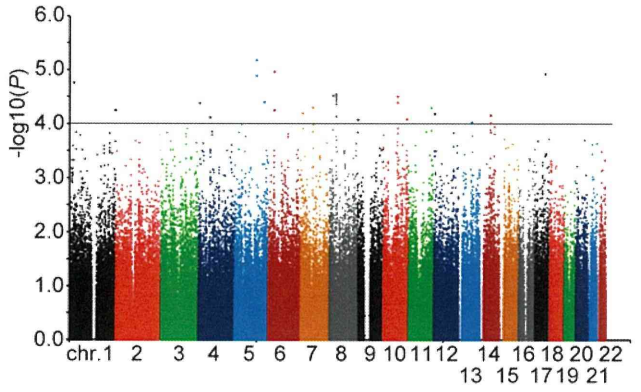

Figure 1. Manhattan plot of the first stage results for pathological myopia. Adjusted *P*-values obtained by the trend χ^2 test for 411,777 SNPs on autosomes in 297 pathological myopic cases and 934 general population controls are plotted in $-\log_{10}$ scale according to their chromosome location.
doi:10.1371/journal.pgen.1000660.g001

For the second stage analysis, 537 cases and 980 population controls were genotyped by Taqman method. Among them, four cases and three controls were excluded due to low call rates (<90%). Genotyping success rates of the 22 SNP markers in the remaining 1,510 samples were greater than 96.8%. The genotype counts of the first and second stages were combined for meta-analysis. One SNP, rs577948, showed a strongly suggestive association ($P=2.22\times10^{-7}$) (Table 2) in the meta-analysis whereas the remaining 21 SNPs were not significant ($P>10^{-5}$) (Table S1).

Evaluation of the region with rs577948

The SNP rs577948 which showed $P=2.22\times10^{-7}$ by meta-analysis with OR of 1.37 (95% confidence interval (CI): 1.21–1.54) for the risk allele (nominal $P=2.80\times10^{-5}$ and $P=1.42\times10^{-3}$ in the first and second stages, respectively) (Table 2) was located at chromosome 11q24.1 (Figure 2A). Using the results of the first stage, an LD block which extended a 55-kb region containing rs577948 was generated. Six additional SNP markers within the block were included in the genome scan chip (Figure 2B). Among them, we selected three markers with adjusted *P*-value smaller than 0.01 in the first stage for further genotyping by Taqman method with DNAs used for the second stage. Weaker associations than that of rs577948 were obtained for these three markers by meta-analysis (Table 2). As shown in Figure 2B, two genes were located in a 200-kb region containing rs577948. *BLID* is a cell death inducer containing BH3-like motif [38], which is located approximately 44-kb upstream of rs577948. The other gene, *LOC399959*, is a hypothetical non-coding RNA [39] which encompassed 114-kb DNA in the region, and rs577948 is located in its second intron.

Expression of the *BLID* and *LOC399959*

BLID is known as a cell-death inducer expressed in cytoplasm, in mitochondria at lower abundance, and in various human cancer cells from different tissues [38]. *LOC399959* was reported as a hypothetical non-coding RNA with a relatively ubiquitous expression pattern. We assessed the expression of the genes by RT-PCR using cDNAs of human retina and brain and those of HeLa cells as positive control. Expressions of both genes were detected in human retinal tissue as well as in human brain and HeLa cells (Figure 3).

Table 2. Association of SNP markers within the linkage disequilibrium block on chromosome 11q24.1 with pathological myopia in Japanese population.

				Meta-analysis[c]		First stage (N = 1,231)				Second stage (N = 1,510)			
SNP ID	Position[a]	Ref.[b]	Var.[b]	*P*-value	OR (95%CI)[d]	Ref. allele freq.		Nominal *P*	OR (95%CI)[d]	Ref. allele freq.		Nominal *P*	OR (95%CI)[d]
						Case (N = 297)	Control (N = 934)			Case (N = 533)	Control (N = 977)		
rs577948	121535400	A	G*	2.22×10^{-7}	1.37 (1.21–1.54)	0.40	0.50	2.80×10^{-5}	1.50 (1.24–1.81)	0.42	0.48	1.42×10^{-3}	1.29 (1.11–1.50)
rs11218544	121544262	T*	G	5.48×10^{-6}	1.33 (1.18–1.51)	0.70	0.61	7.90×10^{-5}	1.50 (1.23–1.83)	0.66	0.61	8.94×10^{-3}	1.24 (1.06–1.44)
rs10892819	121579254	T	G*	0.04	1.15 (1.01–1.31)	0.69	0.75	2.98×10^{-3}	1.36 (1.11–1.67)	0.72	0.73	0.74	1.03 (0.87–1.22)
rs11218553	121590345	A	G*	8.28×10^{-3}	1.18 (1.04–1.34)	0.60	0.67	1.77×10^{-3}	1.36 (1.12–1.65)	0.66	0.68	0.39	1.07 (0.91–1.26)

[a]The position of markers on chromosome 11 refers to NCBI Build 36.1.
[b]Ref. and Var. are the reference and variant nucleotides, respectively, that are defined on the reference sequence of NCBI Build 36.1.
[c]Statistical results using the Mantel-Haenzel method as a fixed-effect model were shown.
[d]Odds ratios (ORs) were calculated for the causative allele (indicated with an asterisk).
doi:10.1371/journal.pgen.1000660.t002

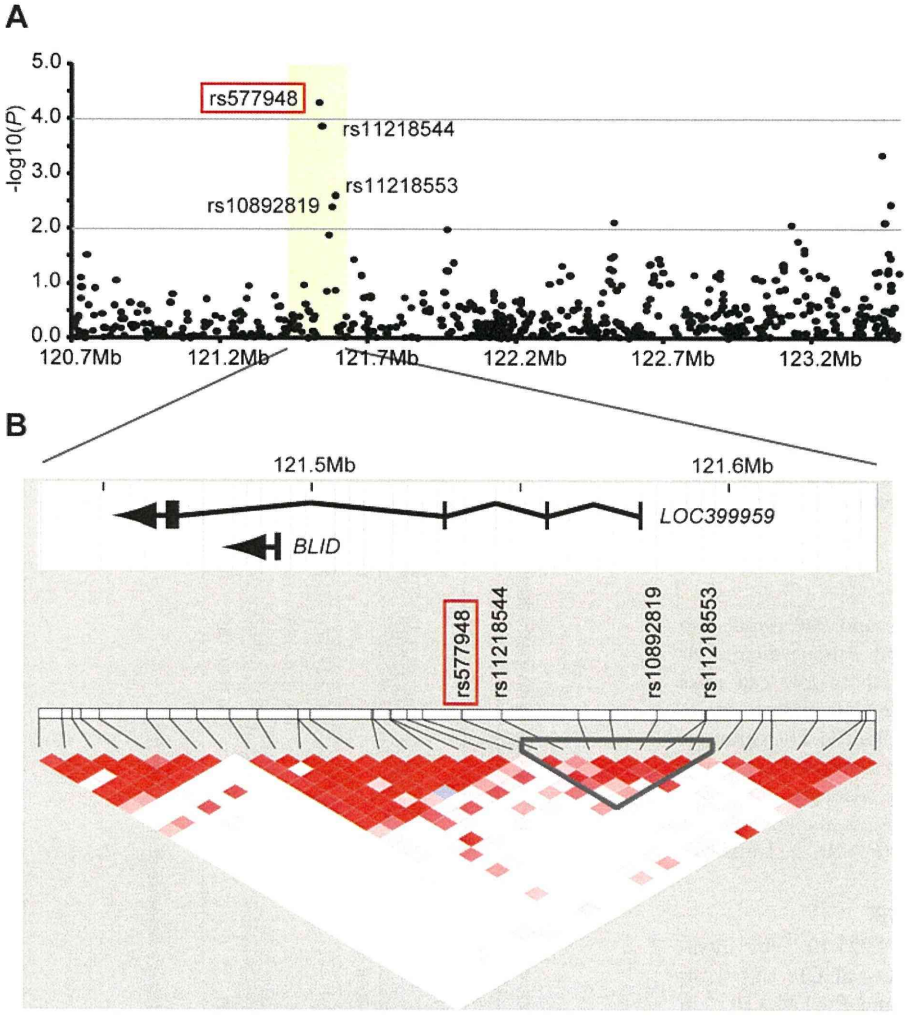

Figure 2. Results of genome scan at 11q24.1 locus containing the *BLID* and *LOC399959* genes. (A) Adjusted *P*-values on $-\log_{10}$ scale for SNPs examined for their association by the trend χ^2 test. (B) Structures, orientations and locations of the *BLID* and *LOC399959* genes on NCBI Reference Sequence Build 36.1, together with pair-wise LD estimates of the SNP markers located within a 200-kb region encompassing the rs577948 marker (red box). Three additional SNP markers (rs11218544, rs11218553, and rs10892819), that showed adjusted *P*-value$<10^{-2}$ in the first stage, are also indicated.
doi:10.1371/journal.pgen.1000660.g002

Discussion

Myopic refraction and axial length are reported to be a complex trait under polygenic control in which contribution of each gene is relatively small [40]. In the current study, two-stage GWA analysis identified a region at chromosome 11q24.1, in which rs577948 showed strongly suggestive $P=2.22\times10^{-7}$ with OR of 1.37 (95% CI: 1.21–1.54) for the allele G.

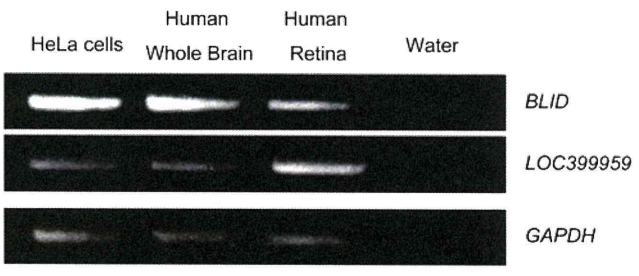

Figure 3. Expression of the *BLID* and *LOC399959* genes in the human retina. RT–PCR analyses of *BLID* and *LOC399959* expression in HeLa cells, the human Brain and the human retina. *GAPDH* was used as an internal control for cDNA quantification.
doi:10.1371/journal.pgen.1000660.g003

Our GWA study identified only one strongly-suggestive locus. This may principally be due to the sample size of our study not being adequate. Recent genetic studies of complex traits with higher prevalence enroll much larger number of samples. In contrast, recruitment of patients with pathological myopia is difficult due to its lower prevalence, particularly those with degenerative changes (namely degenerative myopia). In order to improve insufficient detection power, we assigned pathological myopia patients with longer axis (greater than 28.0 mm) to the first stage. This strategy might be the reason we were successful in identifying the candidate region with relatively small number of cases.

Insufficiency of detection power due to a limitation in sample number may be a reason for difference between the findings of preceding linkage studies and ours. OMIM database lists 10 MYP regions (MYP1–5, 11–13, 15 and 16) for pathological myopia [18–27]

and 6 MYP regions (MYP6–10 and 14) for common myopia [28–30]. None of these 16 MYPs are on chromosome 11q. Stambolian and colleagues reported heterogeneity LOD score of 1.24 at 11q23 in their linkage study for common myopia in Ashkenazi Jewish descent, which is the closest locus to our region reported to date [29]. Because the linkage signal was not strong and the band 11q23 (chr11, position 110,000 kb to 120,700 kb in the NCBI database) is more than 800 kilobases apart from our LD block in 11q24.1 (chr11, position 121,535 kb to 121,590 kb), whether or not they overlap each other is inconclusive. On the other hand, our study did not identify the associated SNPs in any of MYPs.

Although the insufficiency of detection power may be a reason for difference between our study and the linkage studies, there are other possible reasons. In general, any difference in the study designs could cause heterogeneous results. Firstly, there are two definitions of pathological myopia based on two distinct criteria, namely, the axial length and refractive error. In the current study, we enrolled pathological myopic patients based on the axial length (greater than 26.0 mm in both eyes), and not on the refractive error commonly used in the previous studies (refractive errors greater than −6 D). We focused on patients with vision-threatening degenerative changes [9] and the axial length fits better than refractive error for our purpose. The mean refraction in our myopic patients was −13.14±4.57 D (eyes that had undergone cataract surgery or corneal refractive surgery were excluded from this calculation) which indeed correspond to pathological myopic group in the previous linkage studies. On the other hand, it is not clear whether the patients enrolled in the linkage studies fulfill our criteria because the distribution of axial length and degenerative phenotypes in the cases are unknown. The difference in definition of pathological myopia may result in different susceptibility loci between studies.

Secondly, the methodology used is different between studies, namely, linkage analysis and association analysis using linkage disequilibrium mapping. The results of linkage and association studies of complex genetic traits are often different. Family-based linkage analysis is much more suitable for identifying rare genetic variants with large effects whereas SNP-based GWA analysis is more powerful in detection of relatively common variants with smaller effects in complex diseases [41].

Finally, the difference can also be due to the ethnicities of the samples enrolled. In the current study, all cases and controls were Japanese. Only one genome-wide linkage study has previously been published for pathological myopia in Japanese [42] and the others were for non-Japanese populations.

It would be interesting and important to examine the association of our locus in other ethnicities. Ethnic variations in disease susceptibility genes have been reported in various genetic traits including ophthalmological disorders. One such example is an SNP in the complement factor H gene (rs1061170) which has a large effect size with age-related macular degeneration in Caucasians [33,43,44] but much smaller in East Asian populations due to a remarkably lower risk allele frequency (~35% in Caucasians and ~5% in East Asians) [45]. Another example is exfoliation syndrome and *LOXL1* where the risk allele of rs1048661 is inverted between Icelandic (allele G) and Japanese (allele T) populations [35,46]. Because of a large variation in prevalence of myopia among ethnic groups, a future trans-ethnic investigation of myopia risk genes will be important to dissect genetic backgrounds underlying the etiology of myopia.

Although the susceptibility locus contains *BLID* and *LOC399959*, it seems premature to discuss the involvement of *LOC399959* in myopia since it is a hypothetical non-coding gene. *BLID* plays a proapoptotic role involving the BH3-like domain by inducing a caspase-dependent mitochondrial cell death pathway [38]. Indeed, several animal and pathological studies suggested the functional role of apoptosis in pathological myopia [47,48]. Moreover, a recent genome-wide linkage study followed by a fine-scale association mapping identified a myopia susceptibility gene locus containing the *PARL* gene which inhibits the mitochondrial pathway of apoptosis by interaction with *OPA1* [49]. In this context, *BLID* seems functionally relevant with the pathogenesis of pathological myopia. However, the true functional origin of association in this region has yet to be determined by further detailed investigation along with replication studies to validate our findings.

Materials and Methods

Study subjects

All procedures used in this study conformed to the tenets of the Declaration of Helsinki. The Institutional Review Board and the Ethics Committee of each institution approved the protocols used. All the participants were fully informed of the purpose and procedures, and a written consent was obtained from each.

Japanese pathological myopic cases were recruited at the Center for Macular Diseases of Kyoto University Hospital, the High Myopia Clinic of Tokyo Medical and Dental University, and Fukushima Medical University Hospital. All subjects underwent comprehensive ophthalmologic examinations, including dilated indirect and contact lens slit-lamp biomicroscopy, automatic objective refraction evaluation, and measurement of the axial length by applanation A-scan ultrasonography (UD-6000, Tomey, Nagoya, Japan) or partial coherence interferometry (IOLMaster, Carl Zeiss Meditec, Dublin, CA).

As a general population control of the first stage, genotype count data of 934 healthy Japanese subjects were obtained from the JSNP database [36]. For the second stage, 980 healthy Japanese individuals were recruited at Aichi Cancer Center Research Institute. Genomic DNAs were extracted from peripheral blood leukocytes with QuickGene-610L DNA extraction kit (FUJIFILM Co., Tokyo, Japan).

Genome-wide association analysis

We designed to scan the genome in two stages. A total of 839 patients and 1,914 controls were separated into two groups; 302 cases and 934 controls for the first stage, and 537 cases and 980 controls for the second stage. In order to increase the detection power, patients with longer axis of the eyeball (greater than 28.0 mm) were principally assigned to the first stage.

For the first stage analysis, 561,466 SNPs were genotyped in 302 patients of pathological myopia using Illumina Human-Hap550 chips (Illumina Inc., San Diego, CA). This chip covers approximately 87% of the common genetic variations in the Asian population [50]. Cluster definition for each SNP was performed using Illumina BeadStudio Genotyping Module. A systematic quality control procedure of the genome scan results was applied as follows. Samples were evaluated for data quality first and markers were subsequently excluded. Genetic proximity of sample pairs was evaluated with pi-hat in PLINK [51] and four samples with indication of kinship or sample duplication were excluded. Genotypes in X chromosome were used for checking the precision of the phenotype record, and only one sample was removed due to mismatch in gender. The final sample size of pathological myopia was 297. As a population-based control, genotype count data by the genome scanning of 934 healthy Japanese subjects using the same chip were obtained from the JSNP database [36]. The chip contained 515,154 markers in

autosomes that are common in the cases and controls. We excluded 78 SNPs due to low successful call rate (<95%) in the cases, 1,760 SNPs due to the distortion of Hardy-Weinberg Equilibrium (HWE) in the controls ($P<10^{-3}$ by HWE exact test) and 46,722 monomorphic SNPs. 54,817 SNPs with minor allele frequency less than 0.05 in both cases and controls were also excluded. After these quality control procedures, a total of 411,777 SNPs were used for the statistical analysis. The genotyping call rate was greater than 97.43% (median call rate 99.99%) for DNA sample and 98.21% (median call rate 100%) for SNP marker.

Association between genotypic distribution of each SNP and the disease was examined using a χ^2 test for trend. The OR and the 95% CI were estimated using Woolf's method [52]. Inflation in the test statistics was assessed using the genomic-control method [37]. Haploview [53] software was used to infer the LD in the targeted regions. SNPs with *P*-value adjusted by genomic control being smaller than 10^{-4} were selected as candidates for second stage. Among the candidate SNPs, LD indices (D′ and r^2) were calculated with Haploview and when multiple SNPs were in strong LD (D′>0.95 and r^2>0.9), one representative SNP was chosen to be genotyped in the second stage.

In the second stage, 537 cases and 980 controls were genotyped with the Taqman SNP assay using the ABI PRISM 7700 system (Applied Biosystems, Foster City, CA). The 302 pathological myopic cases in the first stage were also genotyped to validate the concordance between Illumina Infinium assay and Taqman assay. Samples with low successful call rate (<90%) were excluded from the study. Subsequently four cases and three controls were excluded and data of 533 cases and 977 controls were used for the analysis. The concordance rate ranged between 98.68% and 100% for the 22 SNPs. The genotype counts of the first and second stages were combined for meta-analysis using the Mantel-Haenzel method [54] as a fixed-effect model. The OR heterogeneity between the first stage and the second stage was evaluated using Cochran's Q-statistic *P*-value. The data from the second stage were also evaluated for association independently from the first stage.

Screening for *BLID* and *LOC399959* expression

Human retina cDNAs were obtained from Takara Bio Inc. (Kyoto, Japan). Total RNA of HeLa cells and human whole brain were also obtained from the same manufacturer and cDNAs were synthesized using the First-Strand cDNA Synthesis Kit (GE Healthcare Life Sciences, Piscataway, NJ). Two pairs of oligonucleotides were synthesized for RT-PCR; 5′-TTGGGTTCCAACAAAGAACC-3′ and 5′-CTTTTACAGGGCCTCAGCAG-3′ for *BLID*, and 5′-GGCGACATCAGACAGACAGA-3′ and 5′-AGGACCAGCTGAAAGGAACA-3′ for *LOC399959*. Expression of glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) was tested for cDNA quantification using 5′-GACAACAGCCTCAAGATCATCA-3′ and 5′-GGTCCACCACTGACACGTTG-3′. PCR reactions were performed under the following condition: initial denaturation at 96°C or 2 minutes, followed by 35 cycles (for *BLID* and *LOC399959*) or 18 cycles (for *GAPDH*) at 96°C for 20 seconds, 60°C for 40 seconds, and polymerization at 72°C for 40 seconds.

Supporting Information

Table S1 Summary results for the 29 SNPs significant at the $P<10^{-4}$ level after population stratification adjustment in the first stage of the genome-wide association analysis.
Found at: doi:10.1371/journal.pgen.1000660.s001 (0.06 MB XLS)

Acknowledgments

We would like to thank Dr. Mineo Ozaki, Dr. Yasuo Kurimoto, Dr. Akio Oishi, Dr. Shoji Kuriyama, and Dr. Shin Yoshitake for their assistance in recruiting patients. We also thank Dr. Mikita Suyama for his expertise in bioinformatics and Mr. Victor Renault and Ms. Akiko Yoshizumi for their technical assistance in data handling.

Author Contributions

Conceived and designed the experiments: HN NG. Performed the experiments: HN NG HH. Analyzed the data: HN RY NY FM. Contributed reagents/materials/analysis tools: NG HH KY NS KOM MM MS TI KM KT. Wrote the paper: HN RY NY FM. Obtained the funding: FM, NY. Performed statistical analysis: RY.

References

1. Kempen JH, Mitchell P, Lee KE, Tielsch JM, Broman AT, et al. (2004) The prevalence of refractive errors among adults in the United States, Western Europe, and Australia. Arch Ophthalmol 122: 495–505.
2. Wong TY, Foster PJ, Hee J, Ng TP, Tielsch JM, et al. (2000) Prevalence and risk factors for refractive errors in adult Chinese in Singapore. Invest Ophthalmol Vis Sci 41: 2486–2494.
3. Sawada A, Tomidokoro A, Araie M, Iwase A, Yamamoto T (2008) Refractive errors in an elderly Japanese population: the Tajimi study. Ophthalmology 115: 363–370 e3.
4. Wong TY, Foster PJ, Ng TP, Tielsch JM, Johnson GJ, et al. (2001) Variations in ocular biometry in an adult Chinese population in Singapore: the Tanjong Pagar Survey. Invest Ophthalmol Vis Sci 42: 73–80.
5. Shufelt C, Fraser-Bell S, Ying-Lai M, Torres M, Varma R (2005) Refractive error, ocular biometry, and lens opalescence in an adult population: the Los Angeles Latino Eye Study. Invest Ophthalmol Vis Sci 46: 4450–4460.
6. Tron EJ (1940) The optical elements of the refractive power of the eye. In: Ridley F, Sprsby A, eds. Modern Trends in Ophthalmology. London: Butterworth. pp 245–255.
7. Katz J, Tielsch JM, Sommer A (1997) Prevalence and risk factors for refractive errors in an adult inner city population. Invest Ophthalmol Vis Sci 38: 334–340.
8. Jacobi FK, Zrenner E, Broghammer M, Pusch CM (2005) A genetic perspective on myopia. Cell Mol Life Sci 62: 800–808.
9. Curtin BJ, Karlin DB (1970) Axial length measurements and fundus changes of the myopic eye. I. The posterior fundus. Trans Am Ophthalmol Soc 68: 312–334.
10. Klaver CC, Wolfs RC, Vingerling JR, Hofman A, de Jong PT (1998) Age-specific prevalence and causes of blindness and visual impairment in an older population: the Rotterdam Study. Arch Ophthalmol 116: 653–658.
11. Buch H, Vinding T, La Cour M, Appleyard M, Jensen GB, et al. (2004) Prevalence and causes of visual impairment and blindness among 9980 Scandinavian adults: the Copenhagen City Eye Study. Ophthalmology 111: 53–61.
12. Iwase A, Araie M, Tomidokoro A, Yamamoto T, Shimizu H, et al. (2006) Prevalence and causes of low vision and blindness in a Japanese adult population: the Tajimi Study. Ophthalmology 113: 1354–1362.
13. Xu L, Wang Y, Li Y, Wang Y, Cui T, et al. (2006) Causes of blindness and visual impairment in urban and rural areas in Beijing: the Beijing Eye Study. Ophthalmology 113: 1134 e1–11.
14. Young TL, Metlapally R, Shay AE (2007) Complex trait genetics of refractive error. Arch Ophthalmol 125: 38–48.
15. Hammond CJ, Snieder H, Gilbert CE, Spector TD (2001) Genes and environment in refractive error: the twin eye study. Invest Ophthalmol Vis Sci 42: 1232–1236.
16. Dirani M, Chamberlain M, Shekar SN, Islam AF, Garoufalis P, et al. (2006) Heritability of refractive error and ocular biometrics: the Genes in Myopia (GEM) twin study. Invest Ophthalmol Vis Sci 47: 4756–4761.
17. Paget S, Vitezica ZG, Malecaze F, Calvas P (2008) Heritability of refractive value and ocular biometrics. Exp Eye Res 86: 290–295.
18. Schwartz M, Haim M, Skarsholm D (1990) X-linked myopia: Bornholm eye disease. Linkage to DNA markers on the distal part of Xq. Clin Genet 38: 281–286.
19. Young TL, Ronan SM, Alvear AB, Wildenberg SC, Oetting WS, et al. (1998) A second locus for familial high myopia maps to chromosome 12q. Am J Hum Genet 63: 1419–1424.
20. Young TL, Ronan SM, Drahozal LA, Wildenberg SC, Alvear AB, et al. (1998) Evidence that a locus for familial high myopia maps to chromosome 18p. Am J Hum Genet 63: 109–119.
21. Paluru P, Ronan SM, Heon E, Devoto M, Wildenberg SC, et al. (2003) New locus for autosomal dominant high myopia maps to the long arm of chromosome 17. Invest Ophthalmol Vis Sci 44: 1830–1836.

22. Paluru PC, Nallasamy S, Devoto M, Rappaport EF, Young TL (2005) Identification of a novel locus on 2q for autosomal dominant high-grade myopia. Invest Ophthalmol Vis Sci 46: 2300–2307.
23. Zhang Q, Guo X, Xiao X, Jia X, Li S, et al. (2005) A new locus for autosomal dominant high myopia maps to 4q22–q27 between D4S1578 and D4S1612. Mol Vis 11: 554–560.
24. Zhang Q, Guo X, Xiao X, Jia X, Li S, et al. (2006) Novel locus for X linked recessive high myopia maps to Xq23–q25 but outside MYP1. J Med Genet 43: e20.
25. Nallasamy S, Paluru PC, Devoto M, Wasserman NF, Zhou J, et al. (2007) Genetic linkage study of high-grade myopia in a Hutterite population from South Dakota. Mol Vis 13: 229–236.
26. Lam CY, Tam PO, Fan DS, Fan BJ, Wang DY, et al. (2008) A genome-wide scan maps a novel high myopia locus to 5p15. Invest Ophthalmol Vis Sci 49: 3768–3778.
27. Paget S, Julia S, Vitezica ZG, Soler V, Malecaze F, et al. (2008) Linkage analysis of high myopia susceptibility locus in 26 families. Mol Vis 14: 2566–2574.
28. Hammond CJ, Andrew T, Mak YT, Spector TD (2004) A susceptibility locus for myopia in the normal population is linked to the PAX6 gene region on chromosome 11: a genomewide scan of dizygotic twins. Am J Hum Genet 75: 294–304.
29. Stambolian D, Ibay G, Reider L, Dana D, Moy C, et al. (2004) Genomewide linkage scan for myopia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 22q12. Am J Hum Genet 75: 448–459.
30. Wojciechowski R, Moy C, Ciner E, Ibay G, Reider L, et al. (2006) Genomewide scan in Ashkenazi Jewish families demonstrates evidence of linkage of ocular refraction to a QTL on chromosome 1p36. Hum Genet 119: 389–399.
31. Tang WC, Yap MK, Yip SP (2008) A review of current approaches to identifying human genes involved in myopia. Clin Exp Optom 91: 4–22.
32. Li YJ, Guggenheim JA, Bulusu A, Metlapally R, Abbott D, et al. (2009) An International Collaborative Family-based Whole Genome Linkage Scan for High-grade Myopia. Invest Ophthalmol Vis Sci. in press.
33. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, et al. (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308: 385–389.
34. Dewan A, Liu M, Hartman S, Zhang SS, Liu DT, et al. (2006) HTRA1 promoter polymorphism in wet age-related macular degeneration. Science 314: 989–992.
35. Thorleifsson G, Magnusson KP, Sulem P, Walters GB, Gudbjartsson DF, et al. (2007) Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. Science 317: 1397–1400.
36. Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, et al. (2002) JSNP: a database of common gene variations in the Japanese population. Nucleic Acids Res 30: 158–162.
37. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997–1004.
38. Broustas CG, Gokhale PC, Rahman A, Dritschilo A, Ahmad I, et al. (2004) BRCC2, a novel BH3-like domain-containing protein, induces apoptosis in a caspase-dependent manner. J Biol Chem 279: 26780–26788.
39. Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, et al. (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl Acad Sci U S A 99: 16899–16903.
40. Klein AP, Duggal P, Lee KE, Klein R, Bailey-Wilson JE, et al. (2005) Support for polygenic influences on ocular refractive error. Invest Ophthalmol Vis Sci 46: 442–446.
41. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273: 1516–1517.
42. Nishizaki R, Ota M, Inoko H, Meguro A, Shiota T, et al. (2009) New susceptibility locus for high myopia is linked to the uromodulin-like 1 (UMODL1) gene region on chromosome 21q22.3. Eye 23: 222–229.
43. Edwards AO, Ritter R, 3rd, Abel KJ, Manning A, Panhuysen C, et al. (2005) Complement factor H polymorphism and age-related macular degeneration. Science 308: 421–424.
44. Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, et al. (2005) Complement factor H variant increases the risk of age-related macular degeneration. Science 308: 419–421.
45. Gotoh N, Yamada R, Hiratani H, Renault V, Kuroiwa S, et al. (2006) No association between complement factor H gene polymorphism and exudative age-related macular degeneration in Japanese. Hum Genet 120: 139–143.
46. Hayashi H, Gotoh N, Ueda Y, Nakanishi H, Yoshimura N (2008) Lysyl oxidase-like 1 polymorphisms and exfoliation syndrome in the Japanese population. Am J Ophthalmol 145: 582–585.
47. Xu GZ, Li WW, Tso MO (1996) Apoptosis in human retinal degenerations. Trans Am Ophthalmol Soc 94: 411–430; discussion 430–431.
48. Mao J, Liu S, Wen D, Tan X, Fu C (2006) Basic fibroblast growth factor suppresses retinal neuronal apoptosis in form-deprivation myopia in chicks. Curr Eye Res 31: 983–987.
49. Andrew T, Maniatis N, Carbonaro F, Liew SH, Lau W, et al. (2008) Identification and replication of three novel myopia common susceptibility gene loci on chromosome 3q26 using linkage and linkage disequilibrium mapping. PLoS Genet 4: e1000220. doi:10.1371/journal.pgen.1000220.
50. Eberle MA, Ng PC, Kuhn K, Zhou L, Peiffer DA, et al. (2007) Power to detect risk alleles using genome-wide tag SNP panels. PLoS Genet 3: e170. doi:10.1371/journal.pgen.0030170.
51. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
52. Woolf B (1955) On estimating the relation between blood group and disease. Ann Hum Genet 19: 251–253.
53. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265.
54. Woolson RF, Bean JA (1982) Mantel-Haenszel statistics and direct standardization. Stat Med 1: 37–39.

nature genetics

A regulatory variant in *CCR6* is associated with rheumatoid arthritis susceptibility

Yuta Kochi[1,12], Yukinori Okada[2,3,12], Akari Suzuki[1], Katsunori Ikari[4], Chikashi Terao[5], Atsushi Takahashi[2], Keiko Yamazaki[6], Naoya Hosono[6], Keiko Myouzen[1], Tatsuhiko Tsunoda[7], Naoyuki Kamatani[2], Tatsuya Furuichi[8], Shiro Ikegawa[8], Koichiro Ohmura[5], Tsuneyo Mimori[5], Fumihiko Matsuda[9], Takuji Iwamoto[4], Shigeki Momohara[4], Hisashi Yamanaka[4], Ryo Yamada[1,9], Michiaki Kubo[6], Yusuke Nakamura[10,11] & Kazuhiko Yamamoto[1,3]

Rheumatoid arthritis is a common autoimmune disease with a complex genetic etiology. Here, through a genome-wide association study of rheumatoid arthritis, we identified a polymorphism in *CCR6*, the gene encoding chemokine (C-C motif) receptor 6 (a surface marker for Th17 cells) at 6q27, that was associated with rheumatoid arthritis susceptibility and was validated in two independent replication cohorts from Japan (rs3093024, a total of 7,069 individuals with rheumatoid arthritis (cases) and 20,727 controls, overall odds ratio = 1.19, $P = 7.7 \times 10^{-19}$). We identified a triallelic dinucleotide polymorphism of *CCR6* (CCR6DNP) in strong linkage disequilibrium with rs3093024 that showed effects on gene transcription. The CCR6DNP genotype was correlated with the expression level of *CCR6* and was associated with the presence of interleukin-17 (IL-17) in the sera of subjects with rheumatoid arthritis. Moreover, CCR6DNP was associated with susceptibility to Graves' and Crohn's diseases. These results suggest that *CCR6* is critically involved in IL-17–driven autoimmunity in human diseases.

Rheumatoid arthritis manifests as inflammation of synovial tissue and joint destruction. The T-cell–driven autoimmune response and the inflammatory cytokine cascade are two important arms of rheumatoid arthritis pathogenesis, involving both environmental and genetic factors[1]. The etiological role of $CD4^+$ helper T cells is supported by evidence that the specific alleles of the *HLA-DRB1* gene encoding the major histocompatibility complex class II proteins are associated with disease risk[2]. Among the subsets of helper T cells, Th1 cells that characteristically produce interferon-γ are considered to have a prominent role in rheumatoid arthritis, as they are highly prevalent in the synovial tissues of subjects with the disease[3,4]. Moreover, a newly identified subset of helper T cells that produces IL-17, hence termed Th17 cells[5], are now thought to have an important role in rheumatoid arthritis, as has been demonstrated in several mouse arthritis models[6–8]. However, less is known about the role of Th17 cells in human rheumatoid arthritis. In fact, among the loci identified in recent genome-wide association studies (GWAS) of rheumatoid arthritis, few have been identified that may affect the axis of Th17 cells specifically, with the exception of the genetic markers near *IL21*[9–12].

To identify rheumatoid arthritis susceptibility loci in the Japanese population, we performed a GWAS by genotyping over 550,000 SNP markers. We applied stringent quality-control criteria and tested 2,303 rheumatoid arthritis cases (of whom 79.2% were positive for anti-cyclic citrullinated peptide (anti-CCP) antibodies; **Supplementary Table 1**) and 3,380 controls for 393,217 autosomal SNPs with minor allele frequency of >0.05. The results of principal-component analysis (PCA)[13] did not demonstrate substantial population stratification in our study population (**Supplementary Fig. 1** and **Supplementary Note**), as has not been previously anticipated for the Japanese population[14]. The inflation of test statistics, $\lambda_{\text{genomic control}}$ (λ_{GC})[15], was 1.097. We identified significant associations in two chromosomal regions (6p21 and 6q27) that satisfied a genome-wide significance threshold of $P < 5 \times 10^{-8}$ (**Fig. 1a** and **Supplementary Fig. 2**). The peak of association in 6p21 was observed at a SNP near *HLA-DRB1* at the major rheumatoid arthritis susceptibility locus (rs13192471, $P = 1.9 \times 10^{-58}$, OR = 1.97, 95% CI 1.82–2.14; **Supplementary Table 2**). Among the markers in 6q27, the smallest P value was observed at a SNP in the region containing *CCR6* (rs3093024, $P = 4.5 \times 10^{-10}$, OR = 1.27, 95% CI 1.18–1.37; **Fig. 1b**, **Table 1** and **Supplementary Tables 2** and **3**). The association of rs3093024 was still significant after PCA correction[13] ($P = 5.5 \times 10^{-9}$). We observed suggestive associations ($5 \times 10^{-8} \leq P < 1 \times 10^{-5}$) in five

[1]Laboratory for Autoimmune Diseases, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan. [2]Laboratory for Statistical Analysis, CGM, RIKEN, Yokohama, Japan. [3]Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. [4]Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan. [5]Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto, Japan. [6]Laboratory for Genotyping Development, CGM, RIKEN, Yokohama, Japan. [7]Laboratory for Medical Informatics, CGM, RIKEN, Yokohama, Japan. [8]Laboratory for Bone and Joint Diseases, CGM, RIKEN, Yokohama, Japan. [9]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [10]Laboratory for International Alliance, CGM, RIKEN, Yokohama, Japan. [11]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [12]These authors contributed equally to this work. Correspondence should be addressed to Y.K. (ykochi@src.riken.jp).

Received 19 January; accepted 6 April; published online 9 May 2010; doi:10.1038/ng.583

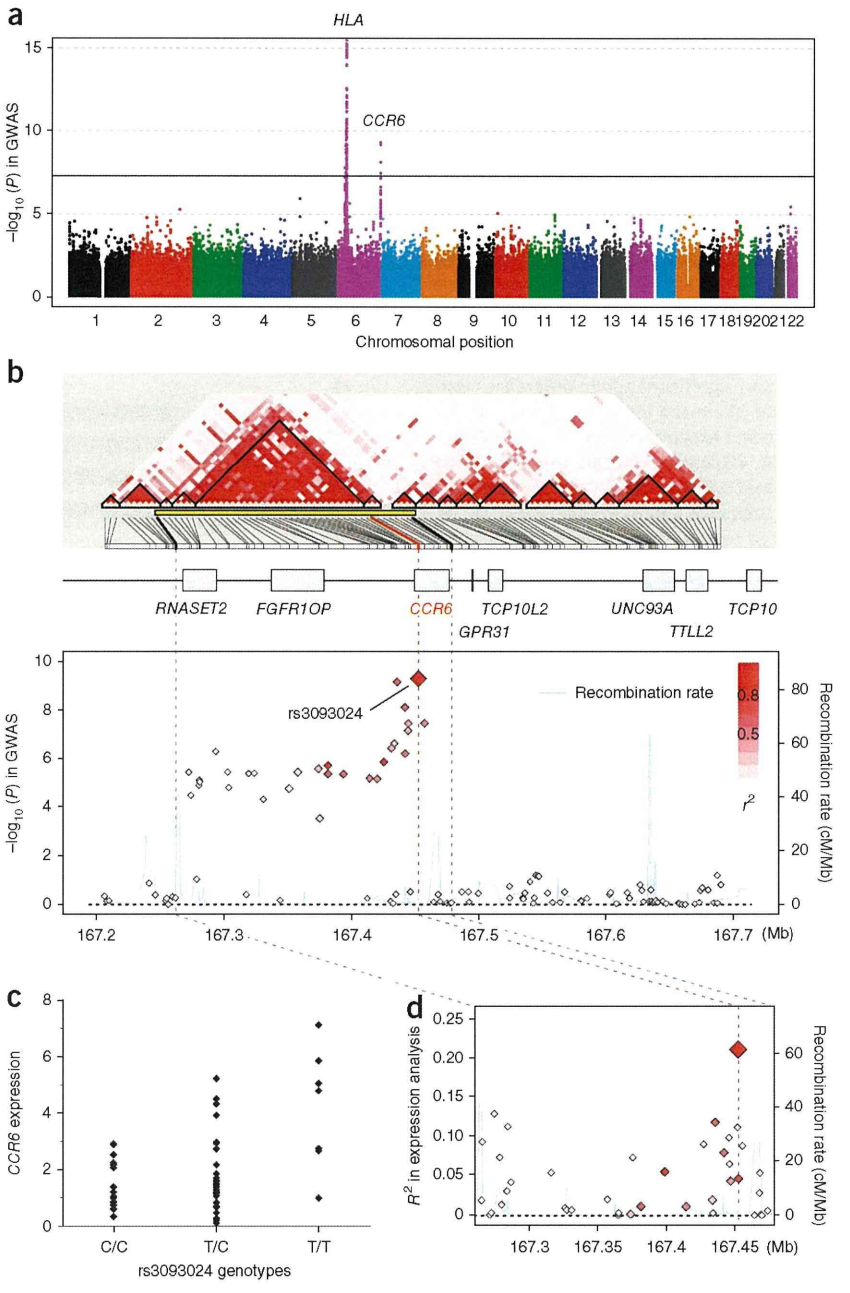

Figure 1 Results of GWAS and expression analysis around *CCR6*. (**a**) A Manhattan plot showing the $-\log_{10}$ (*P* value) of SNPs in the GWAS for 2,303 Japanese rheumatoid arthritis cases and 3,380 controls. Some SNPs in the HLA region are not included because they exceeded the upper limit of the plot. (**b**) LD map (upper), genomic structure (middle) and $-\log_{10}$ (*P* value) of SNPs in the GWAS (lower) around *CCR6*. The *D'*-based LD map (upper) is drawn based on the genotype data of rheumatoid arthritis cases and controls enrolled in the GWAS using Haploview software version 4.1. LD blocks involved in the selection of tag SNPs and the expression analysis are highlighted with a yellow bar. The diamond dots (lower) represent respective SNPs. Their densities in red represent LD indices (R^2), with rs3093024 indicated as the largest dot. (**c**) Correlation between the genotypes of rs3093024 and the transcript levels of CCR6-a (NM_031409) in EBV-transfected cell lines ($n = 57$) stimulated with PMA. (**d**) The coefficients of determination, r^2, of SNPs in the correlation analysis between the genotypes and the expression levels of CCR6-a transcripts. The correlation peak was observed at rs3093024. The red or gray dotted lines represent the correspondence of the chromosomal positions in **b** and **d** (the red line represents rs3093024). The plots of **b** and **d** were drawn by using SNAP version 2.1.

chromosomal regions, which included the region containing the previously associated *STAT4* locus[16–18] (**Supplementary Table 2**). The association results for other previously reported loci[9–12,16,19] are summarized in **Supplementary Table 4**, including replications for the *PADI4* and *TNFAIP3* loci (both $P < 0.01$).

To validate the association with *CCR6*, we performed a replication study using two independent Japanese rheumatoid arthritis case-control cohorts (cohort 1 consisted of 3,662 cases and 15,873 controls, and cohort 2 consisted of 1,106 cases and 1,486 controls). The association of rs3093024 was replicated in both sets ($P = 6.3 \times 10^{-9}$ and $P = 3.1 \times 10^{-3}$ in the two individual cohorts, respectively; combined-analysis $P = 7.7 \times 10^{-19}$, OR = 1.19, 95% CI 1.15–1.24; **Table 1**). A concurrent study[20] also identified an association at the *CCR6* region with rheumatoid arthritis in European populations by performing a meta-analysis of GWAS from a collection of 5,539 autoantibody-positive rheumatoid arthritis cases and 20,169 controls, in which the strongest signal was observed at a SNP in *CCR6* (rs3093023, $P = 3.3 \times 10^{-7}$, OR = 1.13). The landmark SNP in the Japanese population

Table 1 Association analysis of rs3093024 in *CCR6* with rheumatoid arthritis

		Number of subjects		Allele T frequency			
Population	Study set	Cases	Controls	Cases	Controls	OR (95%CI)	*P* value
Japanese	GWAS	2,301	3,368	0.52	0.46	1.27 (1.18–1.37)	4.9×10^{-10}
	Replication study 1[a]	3,662	15,873	0.50	0.46	1.16 (1.11–1.22)	6.3×10^{-9}
	Replication study 2[a]	1,106	1,486	0.51	0.47	1.18 (1.06–1.32)	0.0031
	Combined analysis[b]	7,069	20,727	0.51	0.46	1.19 (1.15–1.24)	7.7×10^{-19}
European	GWAS meta-analysis[c]	5,539	20,169	0.47	0.43	1.13 (1.08–1.18)	3.6×10^{-7}

[a]Cochran-Armitage trend test. [b]Meta-analysis of GWAS and replication studies 1 and 2 (Mantel-Haenszel method). [c]Additive model logistic regression (Wald test) in six GWAS collections and meta-analysis across collections by inverse variance-weighted *z*-scores. Details of the European GWAS meta-analysis are described elsewhere[20]. GWAS, genome-wide association study; OR: odds ratio.

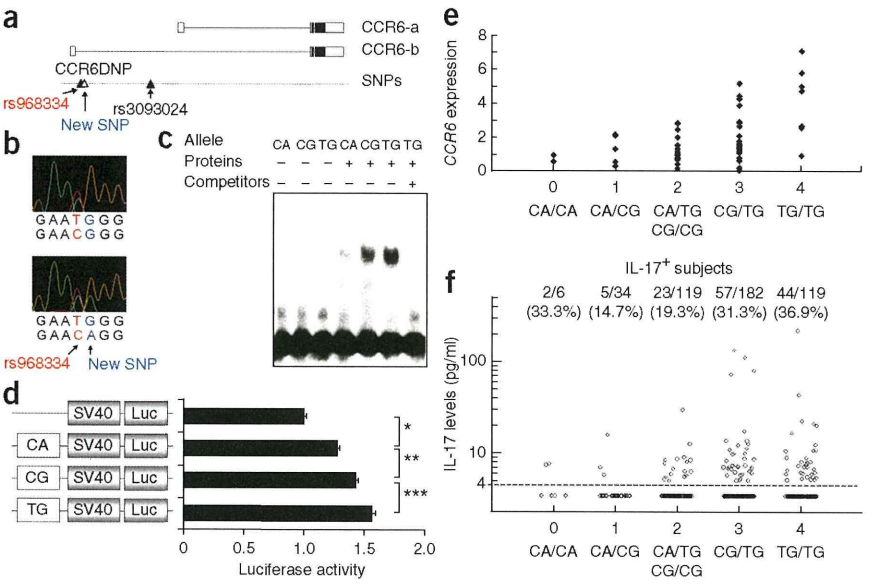

Figure 2 CCR6DNP genotype is correlated with the expression level of *CCR6* and IL-17 status of rheumatoid arthritis cases. (**a**) Genomic position of disease-associated polymorphisms in *CCR6* (CCR6-a and CCR6-b correspond to transcripts NM_031409 and NM_004367 in GenBank, respectively). (**b**) A newly identified SNP in the 3′ flanking sequence of rs968334. (**c**) Binding of nuclear factors from PSC cells to the 31-bp sequences around each allele of CCR6DNP was evaluated by EMSA. Unlabeled probes in 400-fold excess as compared to the labeled probes were used for the competition experiment. The densities of bands were quantified and normalized to that of the TG allele, and significant allelic differences were observed (the mean intensities from four independent experiments were 0.26, 0.76 and 1 for CA, CG and TG, respectively; $P < 0.0001$ by Student's *t*-test). (**d**) Enhanced activity of the 31-bp sequence region around CCR6DNP as evaluated by luciferase assay. SV40, promoter sequence of SV40; luc, luciferase). Data represent mean ± s.e.m. Representative data from three experiments performed in octuplicate are shown. $*P = 1.3 \times 10^{-7}$, $**P = 7.4 \times 10^{-5}$ and $***P = 0.0024$ by Student's *t*-test. (**e**) Correlation between CCR6DNP genotype and expression of *CCR6* as measured by quantitative TaqMan PCR of PSC cells ($n = 57$). (**f**) Correlation between CCR6DNP genotype and IL-17 status in the sera of rheumatoid arthritis cases. The ratios of cases that showed detectable levels of IL-17 (>4 pg/ml) are shown above (IL-17 active cases) and are significantly associated with the CCR6DNP genotype ($P = 1.2 \times 10^{-3}$ by trend test).

(rs3093024) was in almost absolute linkage disequilibrium (LD) with rs3093023 in the European population ($r^2 > 0.99$ and $D' > 0.99$) and showed a similar level of association ($P = 3.6 \times 10^{-7}$, OR = 1.13). Together, these results provide strong evidence for association of the *CCR6* locus with risk of rheumatoid arthritis. This locus encompasses several genes, three of them in the region showing strong LD with rs3093024 (**Fig. 1b**) and six located within 0.5 Mb of either side of this LD region. Although the disease-associated polymorphisms may affect the function of these flanking genes, there is no bibliographical evidence so far showing their functional relevance in rheumatoid arthritis pathogenesis other than that previously shown for *CCR6*.

Recent studies have shown that CCR6 is a specific marker for Th17 cells, distinguishing them from other helper T-cells[8,21], although CCR6 is also expressed in IL-17–producing γδ T cells[22] as well as in subsets of B cells and dendritic cells[23]. Considering the roles of these cells in the pathogenesis of rheumatoid arthritis, *CCR6* is a strong candidate gene in the locus. In a search for causal variants, we first sequenced the *CCR6* coding region in 24 rheumatoid arthritis cases but found no variation that alters its amino acid sequence. We then evaluated whether the *CCR6* polymorphisms could affect gene expression. Epstein-Barr virus (EBV)-transfected cell lines established from Japanese individuals (Pharma SNP Consortium (PSC) cells, $n = 57$) were examined for the expression levels of two major *CCR6* transcripts (NM_031409 and NM_004367 in GenBank, termed here CCR6-a and CCR6-b, respectively). Correlation analysis between the SNP genotype and gene expression level showed a significant correlation between rs3093024 and CCR6-a, not present in resting cells but rather seen in cells stimulated with phorbol myristate acetate (PMA) ($R^2 = 0.21$, $P = 3.2 \times$; 10^{-4} **Fig. 1c**). The expression levels increased with the number of risk (T) alleles. A similar trend was observed for CCR6-b, but this trend was not statistically significant at a level of $\alpha = 0.05$. These observations suggested the presence of regulatory variants in this region that could affect expression levels, especially those of CCR6-a.

To search for regulatory variants, we expanded the region for analysis to approximately 210 kb, covering the entire *CCR6* gene and the surrounding LD blocks containing disease-associated SNPs with $P < 10^{-3}$ (**Fig. 1b**). In addition to the SNPs registered in the Phase II HapMap database for Japanese samples[24], we identified 19 newly discovered SNPs by resequencing the 12-kb region that comprises the 5′ flanking region of *CCR6* in 24 rheumatoid arthritis cases (**Supplementary Fig. 3a**). Of these 312 SNPs, 40 tag SNPs were selected for further correlation analysis with gene expression using a threshold of $r^2 > 0.8$. Correlation analysis revealed a peak of correlation at rs3093024 with the expression level of CCR6-a in cells stimulated with PMA (**Fig. 1d**).

This concordance of peaks in rs3093024 that correlated disease susceptibility with *CCR6* expression made this SNP, or variants in LD with it, a strong candidate as a causal variant. We thus examined all identified SNPs in strong LD ($r^2 > 0.8$) with rs3093024 for their potential to affect gene expression (**Supplementary Fig. 3a**). Allelic differences in binding of nuclear proteins were first evaluated for the sequences surrounding each SNP by electrophoretic mobility shift assays (EMSAs) using nuclear extracts from PSC cells. Of the six SNPs examined, only rs968334 (located 6.7 kb upstream of rs3093024 and with $r^2 = 0.89$) showed a substantial difference in binding of nuclear proteins between the two alleles (**Supplementary Fig. 3b**). Because the nucleotide immediately 3′ of rs968334 was also found to be polymorphic through resequencing this region (**Fig. 2a,b** and **Supplementary Fig. 3a**), we analyzed the haplotype frequency of these two SNPs in 376 rheumatoid arthritis cases by direct sequencing and haplotype phasing. These two SNPs comprised three haplotypes (CA, CG and TG). We termed this 2-base-pair polymorphism *CCR6* dinucleotide polymorphism (CCR6DNP) and thereafter analyzed it as a single, tri-allelic variant. When the binding of nuclear proteins was examined for these three alleles, higher binding was seen in the order of CA < CG < TG (**Fig. 2c**). Although no significant shift was detected in supershift assays using antibodies for nuclear proteins predicted by *in silico* procedures, including YY1 and LYF1 (data not

Table 2 Association analysis of CCR6DNP and autoimmune diseases

		Genotype count						Allele frequency			
		CA/CA	CG/CA	CG/CG	TG/CA	TG/CG	TG/TG	CA	CG	TG	
Disease	Number of subjects	0	1	2	2	3	4	0	1	2	*P* value[a]
Rheumatoid arthritis	2,297	29	161	398	208	897	604	0.09	0.40	0.50	2.4×10^{-9}
Graves' disease	1,783	25	133	334	150	715	426	0.09	0.43	0.48	2.4×10^{-5}
Crohn's disease	483	5	40	81	52	195	110	0.11	0.41	0.48	0.020
Controls	1,808	28	189	379	168	683	361	0.11	0.45	0.44	–

[a]*P* values to determine trends for genotype using the χ^2 test.

shown), this differential binding of unknown nuclear proteins should affect the transcriptional activity of each allele. We next performed reporter assays using oligonucleotides of these three alleles cloned into a plasmid vector containing the viral SV40 promoter. In parallel with the results from the gel shift assays, different enhancing activities were detected in the order of CA < CG < TG (**Fig. 2d**). Taken together, these results suggest that CCR6DNP is a strong candidate for a disease-causing variation that affects gene expression.

To analyze the effect of CCR6DNP on disease susceptibility and on phenotypes within individuals, we established a genotyping assay for CCR6DNP (**Supplementary Fig. 4**). We scored the alleles as 0, 1 and 2 for CA, CG and TG, respectively, according to the enhanced transcription observed in the *in vitro* assays. The diplotypes of individuals were in turn scored as 0, 1, 2, 3 and 4 by summing each individual's allele scores. When the expression of *CCR6* in PSC cells ($n = 57$) was regressed with this diplotype score for CCR6DNP, a significant correlation was observed ($R^2 = 0.23$, $P = 1.7 \times 10^{-4}$; **Fig. 2e**), supporting the regulatory effect of CCR6DNP. When the distribution of the diplotype scores for CCR6DNP was compared between 2,303 rheumatoid arthritis cases and 1,820 controls, a significant association was also observed ($P = 2.4 \times 10^{-9}$ by trend test; **Table 2** and **Supplementary Table 1**). Compared with the CA allele, the ORs of the other alleles demonstrated an identical trend order (CA, 1.00; CG, 1.14; and TG, 1.25) with those indicated in the *in vitro* assays, which supports the hypothesis that the triallelic variant is causal. Conditional regression analysis of the newly identified SNP with rs968334 did not indicate a significant independent association at a level of $\alpha = 0.05$, presumably because of the strong LD between the SNPs ($D' = 1$) or because of inadequate statistical power due to the low minor-allele frequency of the newly identified SNP (minor allele frequency = 0.10). We also genotyped individuals with Graves' disease ($n = 1,783$) and Crohn's disease ($n = 483$); we found significant association of CCR6DNP with both Graves' disease ($P = 2.4 \times 10^{-5}$) and Crohn's disease ($P = 0.020$; **Table 2**).

To seek further association between the CCR6DNP genotype and phenotypes within individuals, we analyzed the serum level of IL-17 in rheumatoid arthritis cases. Among 451 cases examined, 28.5% showed detectable levels of IL-17 (here termed 'IL-17 active cases'). There was a significant association between the CCR6DNP genotype and the proportion of IL-17 active cases ($P = 1.2 \times 10^{-3}$ by trend test; **Fig. 2f**). This association was still significant after adjustment for age and sex using a logistic regression model ($P = 1.4 \times 10^{-3}$). No significant correlation was observed between CCR6DNP and the titer of IL-17 at $\alpha = 0.05$. We also analyzed the association between the CCR6DNP genotype and autoantibody positivity (anti-CCP antibodies and rheumatoid factors) in the same rheumatoid arthritis cases, but did not observe any significant association (**Supplementary Fig. 5**). These results suggest that CCR6DNP has a more direct link to the activation of Th17 cells and their secretion of IL-17 than in autoantibody response in rheumatoid arthritis.

The role of Th17 cells in autoimmune arthritis has been well demonstrated in SKG mice, an animal model for rheumatoid arthritis which results from a mutation in *Zap70* (ref. 8). In this model, $Ccr6^+$ Th17 cells are recruited into the inflamed joints by Ccl20 (a known ligand for Ccr6) that is produced by synoviocytes and Th17 cells themselves. The arthritogenic role of Th17 cells in SKG mice is supported by the observation that the administration of blocking anti-Ccr6 antibodies substantially inhibits arthritis[8]. In humans, the majority of circulating Th17 cells express CCR6 (ref. 21), and its ligand, CCL20, is also detected in inflamed synovial tissues[25]. Preferential expression of *CCR6* in synovial tissues from rheumatoid arthritis cases compared to those from osteoarthritis cases has been previously described[25] and was confirmed in our samples (**Supplementary Fig. 6**). Our observation that CCR6DNP is associated with rheumatoid arthritis susceptibility, as well as our detection of enhanced expression of *CCR6* and the status of IL-17 in the serum of rheumatoid arthritis cases, implies that the same mechanism found in the mouse model might also underlie the pathogenesis of human rheumatoid arthritis.

The CCR6-CCL20–mediated migration of Th17 cells to inflamed tissues may also be important in other autoimmune diseases, as is implicated by the association of Graves' disease and Crohn's disease susceptibility to CCR6DNP observed in the present study. In Crohn's disease, a genetic marker in *FGFR1OP* (adjacent to *CCR6*) was also implicated in a meta-analysis of GWAS[26], supporting the association of *CCR6* to disease. As several other genes implicated in Th17 function, such as *IL23R* and *STAT3*, have been associated with Crohn's disease susceptibility, Th17 cells are considered to play a substantial role in Crohn's disease[27,28]. In contrast, the lack of confirmatory evidence for an association of *IL23R* and *STAT3* in rheumatoid arthritis[9–12] implies that this pathway might be less important in rheumatoid arthritis as compared to Crohn's disease. Therefore, it is possible that the *CCR6* polymorphism could be involved in rheumatoid arthritis pathogenesis by influencing *CCR6*-expressing cell types other than Th17 cells. As enhanced expression of *CCR6* may differentially contribute to pathogenesis between different autoimmune diseases, studies of other autoimmune diseases are needed to clarify the etiological role of the *CCR6* polymorphism in autoimmunity.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession numbers. RefSeq, NM_031409 and NM_004367.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank K. Shimada, Y. Hayashi, K. Kobayashi, M. Kitazato and other members of the Laboratory for Autoimmune Diseases, RIKEN and Y. Katagiri at Tokyo Women's Medical University, for their technical assistance. We also

thank the members of BioBank Japan and the Rotary Club of Osaka-Midosuji District 2660 Rotary International for supporting our study. This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Sciences and Technology of the Japanese government. The replication study of rheumatoid arthritis was performed under the support of Genetics and Allied research in Rheumatoid Arthritis Networking (GARNET) consortium.

AUTHOR CONTRIBUTIONS

Y.K., Y.O. and K. Yamamoto. designed the study and drafted the manuscript. Y.O., A.T., T.T. and R.Y. analyzed the GWAS data. N.H. and M.K. performed the genotyping for the GWAS. Y.K. performed expression analysis of *CCR6* and functional analysis of *CCR6* polymorphisms. Y.K. and K.M. established the genotyping method for CCR6DNP. K.I., S.M. and H.Y. analyzed data for the first replication cohort of rheumatoid arthritis. C.T., K.O., T.M., R.Y. and F.M. analyzed the data for the second replication cohort of rheumatoid arthritis. Y.K. analyzed the data for the Graves' disease cohort. K. Yamazaki analyzed the data for the Crohn's disease cohort. T.F. and S.I. analyzed the data for the fourth control cohort. T.I. and K.I. analyzed *CCR6* expression in the synovial tissues. Y.K. and A.S. analyzed the sera of subjects with rheumatoid arthritis. M.K., N.K. and Y.N. contributed to overall GWAS study design.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Firestein, G.S. Evolving concepts of rheumatoid arthritis. *Nature* **423**, 356–361 (2003).
2. Newton, J.L., Harney, S.M., Wordsworth, B.P. & Brown, M.A. A review of the MHC genetics of rheumatoid arthritis. *Genes Immun.* **5**, 151–157 (2004).
3. Qin, S. *et al.* The chemokine receptors CXCR3 and CCR5 mark subsets of T cells associated with certain inflammatory reactions. *J. Clin. Invest.* **101**, 746–754 (1998).
4. Yamada, H. *et al.* Th1 but not Th17 cells predominate in the joints of patients with rheumatoid arthritis. *Ann. Rheum. Dis.* **67**, 1299–1304 (2008).
5. Harrington, L.E. *et al.* Interleukin 17-producing $CD4^+$ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nat. Immunol.* **6**, 1123–1132 (2005).
6. Murphy, C.A. *et al.* Divergent pro- and antiinflammatory roles for IL-23 and IL-12 in joint autoimmune inflammation. *J. Exp. Med.* **198**, 1951–1957 (2003).
7. Pernis, A.B. Th17 cells in rheumatoid arthritis and systemic lupus erythematosus. *J. Intern. Med.* **265**, 644–652 (2009).
8. Hirota, K. *et al.* Preferential recruitment of CCR6-expressing Th17 cells to inflamed joints via CCL20 in rheumatoid arthritis and its animal model. *J. Exp. Med.* **204**, 2803–2812 (2007).
9. Plenge, R.M. *et al. TRAF1–C5* as a risk locus for rheumatoid arthritis—a genomewide study. *N. Engl. J. Med.* **357**, 1199–1209 (2007).
10. Plenge, R.M. *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat. Genet.* **39**, 1477–1482 (2007).
11. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
12. Gregersen, P.K. *et al. REL*, encoding a member of the NF-κB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.* **41**, 820–823 (2009).
13. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
14. Yamaguchi-Kabata, Y. *et al.* Japanese population structure, based on SNP genotypes from 7,003 individuals compared to other ethnic groups: effects on population-based association studies. *Am. J. Hum. Genet.* **83**, 445–456 (2008).
15. Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor. Popul. Biol.* **60**, 155–166 (2001).
16. Remmers, E.F. *et al. STAT4* and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.* **357**, 977–986 (2007).
17. Kobayashi, S. *et al.* Association of *STAT4* with susceptibility to rheumatoid arthritis and systemic lupus erythematosus in the Japanese population. *Arthritis Rheum.* **58**, 1940–1946 (2008).
18. Lee, H.S. *et al.* Association of *STAT4* with rheumatoid arthritis in the Korean population. *Mol. Med.* **13**, 455–460 (2007).
19. Suzuki, A. *et al.* Functional haplotypes of *PADI4*, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **34**, 395–402 (2003).
20. Stahl, E.A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* advance online publication, doi:10.1038/ng.582 (9 May 2010).
21. Annunziato, F. *et al.* Phenotypic and functional features of human Th17 cells. *J. Exp. Med.* **204**, 1849–1861 (2007).
22. Haas, J.D. *et al.* CCR6 and NK1.1 distinguish between IL-17A and IFN-gamma-producing γδ effector T cells. *Eur. J. Immunol.* **39**, 3488–3497 (2009).
23. Schutyser, E., Struyf, S. & Van Damme, J. The CC chemokine CCL20 and its receptor CCR6. *Cytokine Growth Factor Rev.* **14**, 409–426 (2003).
24. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
25. Matsui, T. *et al.* Selective recruitment of CCR6-expressing cells by increased production of MIP-3 α in rheumatoid arthritis. *Clin. Exp. Immunol.* **125**, 155–161 (2001).
26. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
27. Brand, S. Crohn's disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease. *Gut* **58**, 1152–1167 (2009).
28. Wang, K. *et al.* Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. *Am. J. Hum. Genet.* **84**, 399–405 (2009).

ONLINE METHODS

Samples enrolled in the study. We enrolled 7,090 rheumatoid arthritis cases, 1,783 Graves' disease cases, 483 Crohn's disease cases and 21,712 unaffected controls through several medical institutes in Japan under the support of the BioBank Japan Project[29], the University of Tokyo, Institute of Rheumatology Rheumatoid Arthritis (IORRA), Kyoto University and the GARNET consortium. All subjects were of Japanese origin and provided written informed consent. Details of the samples are summarized in **Supplementary Table 1** and **Supplementary Note**. This research project was approved by the ethical committees of the BioBank Japan Project, RIKEN, the University of Tokyo, Tokyo Women's Medical University and Kyoto University.

Genotyping and quality control. In the GWAS, 2,322 rheumatoid arthritis cases and 3,473 controls were genotyped using the Illumina Human610-Quad and Illumina Human 550v3 Genotyping BeadChips (Illumina), respectively. After the exclusion of samples with low call rates (<98%), SNPs with low call rates (<99%), non-autosomal SNPs and those SNPs not shared among cases and controls were also excluded.

We excluded closely related samples using identity by descent estimated by PLINK[30] version 1.06. For sample pairs in a first degree of kinship, we excluded the individual who was a control, or in cases where that had not uniquely been determined, the individual had lower call rate than the other. We next excluded samples who were outliers in terms of ancestries. We performed PCA for the genotype data of the samples along with European (CEU), African (YRI) and East-Asian (Japanese and Han Chinese, denoted JPT + CHB) individuals obtained from Phase II HapMap database (release 22)[24] using EIGENSTRAT[13] version 2.0. A PCA plot clearly separated the samples into three clusters, as previously indicated[14] (**Supplementary Fig. 1**). The concordance of the cluster between JPT + CHB individuals and our samples suggested that our samples had homogeneous ancestries. We visually excluded the samples who were estimated to be outliers.

We excluded SNPs with minor allele frequency less than 0.05 in either rheumatoid arthritis cases or controls, or with Hardy-Weinberg equilibrium P value[31] $>10^{-6}$ in controls. Cluster plots of SNPs were checked by visual inspection, and SNPs with ambiguous calls were excluded. Finally, 393,217 SNPs for 2,303 rheumatoid arthritis cases and 3,380 controls were obtained. The genotyping methods for the replication studies are described in **Supplementary Table 1**.

Association analysis. The associations of the SNPs were tested with the Cochran-Armitage trend test. Combined analysis was performed with the Mantel-Haenszel method. Correction for potential population stratification was performed by the genomic control method[15] and by using the results of PCA[13] (see **Supplementary Note** for details).

Replication study for previously reported rheumatoid arthritis susceptibility loci. We evaluated the associations in previously reported rheumatoid arthritis susceptibility loci[9–12,16–19,32–39] using the samples in our GWAS (2,303 rheumatoid arthritis cases and 3,380 controls) and the controls in the first replication study (15,873 controls). We imputed the SNPs that were not genotyped using MACH 1.0 (see URLs), with Phase II HapMap JPT + CHB individuals as references[24]. All the imputed SNPs demonstrated *Rsq* values more than 0.95.

Genotyping method for CCR6DNP. The triallelic dinucleotide polymorphism in *CCR6* (CCR6DNP) was genotyped using two sets of TaqMan MGB probes (**Supplementary Table 5**). Approximately 4 ng DNA was amplified in a 5 μl reaction mixture comprising 200 nM each of probe and 900 nM each of primer in TaqMan genotyping master mix (Applied Biosystems). After denaturation at 95 °C for 10 min, 50 cycles of denaturation at 95 °C for 15 s and annealing and extension at 58 °C for 1 min were performed. The results for the reactions of two probe sets were analyzed together and the genotype of each sample was determined by an algorithm as shown in **Supplementary Figure 4**. The genotypes of 376 individuals completely matched (100%) those obtained by direct sequencing.

DNA resequencing. Unknown genetic variants were revealed by directly sequencing of the DNA of 24 individuals affected with rheumatoid arthritis. DNA fragments were amplified for sequencing with appropriate primers and were purified using a MultiScreen PCR filter plate (Millipore). The amplified DNAs were sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and signals were detected using an Applied Biosystems ABI 3700 DNA Analyzer.

Quantification of gene expression. EBV-transformed lymphoblastoid cell lines (n = 57) were established by the Pharma SNP Consortium (PSC, Tokyo, Japan) using peripheral blood lymphocytes of Japanese healthy individuals. Cells were incubated for 2 h in medium alone (RPMI 1640 medium containing 10% FBS albumin, 1% penicillin and 1% streptomycin) or with 100 ng/ml PMA. Conditions for cell stimulation were optimized before the experiment as previously described[40]. Cells were then harvested and total RNA was isolated using an RNeasy Mini Kit (Qiagen) with DNase treatment. Total RNA from synovial tissues was isolated from rheumatoid arthritis cases (n = 74) and osteoarthritis cases (n = 8) who underwent total knee replacement surgery. Total RNA (1 μg) was reverse transcribed using TaqMan Gold RT-PCR reagents with random hexamers (Applied Biosystems). Real-time quantitative PCR was performed in triplicate using an ABI PRISM 7900 and TaqMan gene expression assays (Applied Biosystems). Specific probes for each transcript of *CCR6* were used: Hs00171121_m1 was used for CCR6-a (NM_031409) and a custom-made probe set was used for CCR6-b (NM_004367); both the custom probe and Hs00171121_ml were comprised of a minor groove binder probe (**Supplementary Table 5**). Possible contamination of DNA was excluded by performing PCR with total RNA without the reverse transcription step. The data were normalized to *GAPDH* levels. *GUS* and 18S rRNA levels were also evaluated for internal control, and similar results were obtained among the three probes.

Correlation analysis between genotype and gene expression. DNA was extracted from PSC cells and genotyped for the tag SNPs using TaqMan genotyping assays (Applied Biosystems). The levels of each transcript of *CCR6* were regressed with the genotype in a linear model, assuming an additive effect of the allele. The correlation coefficients R^2 were used for fine mapping of the correlation between SNP genotypes and *CCR6* expression in this region. The significance of regression was tested by an F test.

Electrophoretic mobility shift assay (EMSA). EMSA and preparation of nuclear extract from PSC cells were performed as previously described[41]. Following stimulation with 50 ng/ml PMA for 2 h, cells were collected and lysed in Buffer A (20 mM HEPES at pH 7.6, 20% glycerol, 10 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA at pH 8.0, 1 mM DTT, 0.1% NP-40 detergent and a protease inhibitor cocktail). The cells were then incubated on ice for 10 min, centrifuged, and the pellets were resuspended in Buffer B (which contains Buffer A with 500 mM NaCl). Following incubation on ice for 30 min and centrifugation, the supernatant fraction containing nuclear proteins was collected. Oligonucleotides (31 base pairs) were designed that corresponded to genomic sequences surrounding the SNPs of interest. Single-stranded oligonucleotide probes were labeled using a Biotin 3′ End DNA Labeling Kit (Pierce Biotechnology), and sense and antisense oligonucleotides were then annealed. DNA-protein interactions were detected by using a LightShift Chemiluminescent EMSA kit (Pierce Biotechnology) according to the manufacturer's instructions. The DNA-protein complexes were separated on a nondenaturing 5% polyacrylamide gel in 1× TBE (Tris-borate-EDTA) running buffer for 70 min at 150 V. The gel was transferred to a nitrocellulose membrane, and signals were detected using a LAS-3000 lumino image analyzer (Fujifilm).

***In silico* prediction of transcription factor binding site.** DNA sequences around each allele of the CCR6DNP were evaluated for the potential of transcriptional factor binding using Match software version 1.0 (see URLs). This software uses a library of mononucleotide weight matrices from TRANSFAC 6.0 (see URLs).

Luciferase assay. Oligonucleotides (31 base pairs) were generated using the allelic sequences of nucleotides surrounding CCR6DNP (**Supplementary Table 6**). A single copy of the oligonucleotide was cloned into the pGL3-Promoter Vector (Promega) at the MulI and BglII restriction enzyme binding sites. Each construct was then transformed into *Escherichia coli* DH5α (Toyobo). These plasmids were purified using a HiSpeed Plasmid Midi Kit (Qiagen)

after confirmation of the sequence. HEK293A cells (Invitrogen) were grown in Dulbecco's modified Eagle medium (Sigma) supplemented with 10% FBS and antibiotics. Cells ($n = 2 \times 10^5$) were transfected with 0.8 μg constructs and 0.2 μg pRL-TK vector (an internal control for transfection efficiency) using the TransFast Transfection Reagent (Promega). After 20 h, 100 ng/ml PMA was added to the medium. Following an additional incubation for 4 h, cells were collected and luciferase activity was measured using a Dual-Luciferase Reporter Assay System (Promega) and a Lumat LB 9507 Ultra Sensitive Tube Luminometer (Berthold Technologies). Each experiment was independently repeated three times and octuplicate samples were assayed each time.

Measurement of autoantibodies and IL-17. Anti-citrullinated peptide antibodies were measured using a MESACUP CCP test (Medical & Biological Laboratories) according to the manufacturer's instructions. Rheumatoid factors were measured by ELISA as previously described[42]. Human IL-17 was measured by ELISA, with a detection limit of 4 pg/ml (eBioscience).

URLs. MACH, http://www.sph.umich.edu/csg/abecasis/mach/; Match software, http://www.gene-regulation.com/; TRANSFAC, http://www.biobase-international.com/.

29. Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696–697 (2007).
30. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
31. Wigginton, J.E., Cutler, D.J. & Abecasis, G.R. A note on exact tests of Hardy-Weinberg equilibrium. *Am. J. Hum. Genet.* **76**, 887–893 (2005).
32. Kochi, Y. *et al.* A functional variant in *FCRL3*, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. *Nat. Genet.* **37**, 478–485 (2005).
33. Tokuhiro, S. *et al.* An intronic SNP in a RUNX1 binding site of *SLC22A4*, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat. Genet.* **35**, 341–348 (2003).
34. Suzuki, A. *et al.* Functional SNPs in *CD244* increase the risk of rheumatoid arthritis in a Japanese population. *Nat. Genet.* **40**, 1224–1229 (2008).
35. Raychaudhuri, S. *et al.* Common variants at *CD40* and other loci confer risk of rheumatoid arthritis. *Nat. Genet.* **40**, 1216–1223 (2008).
36. Thomson, W. *et al.* Rheumatoid arthritis association at 6q23. *Nat. Genet.* **39**, 1431–1433 (2007).
37. Shimane, K. *et al.* The association of a nonsynonymous single-nucleotide polymorphism in *TNFAIP3* with systemic lupus erythematosus and rheumatoid arthritis in the Japanese population. *Arthritis Rheum.* **62**, 574–579 (2010).
38. Raychaudhuri, S. *et al.* Genetic variants at *CD28*, *PRDM1* and *CD2/CD58* are associated with rheumatoid arthritis risk. *Nat. Genet.* **41**, 1313–1318 (2009).
39. Okada, Y. *et al. SLC22A4* polymorphism and rheumatoid arthritis susceptibility: a replication study in a Japanese population and a meta-analysis. *J. Rheumatol.* **35**, 1723–1728 (2008).
40. Nishimoto, K. *et al.* Association study of *TRAF1–C5* polymorphisms with susceptibility to rheumatoid arthritis and systemic lupus erythematosus in Japanese. *Ann. Rheum. Dis.* **69**, 368–373 (2010).
41. Aikawa, Y., Yamamoto, M., Yamamoto, T., Morimoto, K. & Tanaka, K. An anti-rheumatic agent T-614 inhibits NF-κB activation in LPS- and TNF-α-stimulated THP-1 cells without interfering with IκBα degradation. *Inflamm. Res.* **51**, 188–194 (2002).
42. Shimane, K. *et al.* A single nucleotide polymorphism in the *IRF5* promoter region is associated with susceptibility to rheumatoid arthritis in the Japanese population. *Ann. Rheum. Dis.* **68**, 377–383 (2009).

Interferon γ receptor 2 gene variants are associated with liver fibrosis in patients with chronic hepatitis C infection

Bertrand Nalpas,[1] Roubila Lavialle-Meziani,[2] Sabine Plancoulaine,[3,4] Emmanuelle Jouanguy,[3,5] Antoine Nalpas,[1] Mona Munteanu,[6] Frederic Charlotte,[6] Brigitte Ranque,[3,4] Etienne Patin,[3,4] Simon Heath,[2] Hélène Fontaine,[1,4] Anaïs Vallet-Pichard,[1,4] Dominique Pontoire,[7] Marc Bourlière,[8] Jean-Laurent Casanova,[3,4,9,10] Mark Lathrop,[2] Christian Bréchot,[11] Thierry Poynard,[6] Fumihiko Matsuda,[2,12] Stanislas Pol,[1,4] Laurent Abel[3,4,9]

► Supplementary tables are published online only. To view these files please visit the journal online (http://gut.bmj.com).

For numbered affiliations see end of article.

Correspondence to
Dr Laurent Abel, Laboratoire de Génétique Humaine des Maladies Infectieuses, Université Paris Descartes-INSERM U550, Faculté de Médecine Necker, 156 rue de Vaugirard, 75015 Paris, France;
laurent.abel@inserm.fr

Accepted 19 March 2010
Published Online First 29 June 2010

ABSTRACT

Background Only a minority of patients with chronic hepatitis C virus (HCV) infection develops severe liver fibrosis, a process that may be controlled by human genetic factors.

Objective To investigate the role of 384 single nucleotide polymorphisms (SNPs) located in 36 candidate genes related to the fibrogenesis/fibrolysis process.

Methods Patients with chronic HCV infection were gathered from two French cohorts (prospectively and retrospectively). The overall sample consisted of 393 HCV-infected subjects without known risk factors for fibrosis progression, including 134 patients with severe liver fibrosis and 259 without severe fibrosis.

Results Only two SNPs in strong linkage disequilibrium (LD) in the interferon γ receptor 2 gene (*IFNGR2*) were significantly associated with liver fibrosis in both the prospective and the retrospective samples. The strongest association ($p=8\times10^{-5}$) was observed with the G/A SNP rs9976971 with an OR of severe fibrosis for AA versus AG or GG subjects at 2.95 (95% CI 1.70 to 5.11). This effect was higher ($p=9\times10^{-7}$) when taking into account the time of follow-up, and the hazard ratio of progression towards severe fibrosis for AA patients was 2.62 (1.76 to 3.91). Refined sequencing and analysis of the *IFNGR2* region identified two additional variants in strong LD with rs9976971. No haplotypes derived from this cluster of four variants provided stronger evidence for association than rs9976971 alone.

Conclusions This identification of a cluster of four *IFNGR2* variants strongly associated with fibrosis progression in chronic HCV infection underlines the role of IFNγ in the development of liver fibrosis that may pave the way for new treatments.

Significance of this study

What is already known about this subject?

- Only a minority of patients with chronic HCV infection develops severe liver fibrosis.
- Viral and non-genetic host factors cannot account for the variability in the rate of progression towards liver fibrosis in patients with chronic HCV infection.
- There is accumulating evidence for the role of host genetic factors in the development of liver fibrosis, although these factors are as yet largely elusive.

What are the new findings?

- We investigated the role of 384 single nucleotide polymorphisms (SNPs) located in 36 candidate genes related to the fibrogenesis/fibrolysis process.
- We identified a single cluster of variants in the interferon γ receptor 2 gene (*IFNGR2*) strongly associated with progression to severe fibrosis.
- This association was replicated in an independent population sample, and was stronger when taking into account the time of follow-up from contamination to liver biopsy with a hazard ratio of developing severe fibrosis estimated as 2.62 (1.76 to 3.91) for the subjects homozygous for the predisposing allele.

How might it impact on clinical practice in the foreseeable future?

- As IFNγ is an anti-fibrogenic cytokine available for clinical purposes, our study may open new therapeutic avenues for the prevention of cirrhosis in HCV-infected patients.

INTRODUCTION

Hepatitis C virus (HCV) infection is a major public health concern world wide with an estimated 170 million people infected.[1 2] The natural history of patients with HCV chronic infection is characterised by a highly variable disease progression.[1–5] Most subjects never develop cirrhosis during their lifetime while the remaining patients are considered 'rapid fibrosers' and may develop severe fibrosis in <20 years.[1 5–7] While viral factors such as HCV genotypes or viral load do not seem to influence progression, several host factors such as gender (male), age at infection (>40 years old), alcohol consumption (>50 g/day), obesity and its related metabolic disorders, co-infections (in particular by HIV) are associated with the development of fibrosis.[4 5 8 9] However, these factors can account for only a minority of the variability in the rate of

progression, and a number of studies have investigated the role of genetic host factors.[5 10] Most of these studies have tested a single or a few candidate genes, and have not produced conclusive results as they were not clearly replicated.[10 11] A recent study investigated a large number (>24 000) of putative functional single nucleotide polymorphisms (SNPs) and identified two variants associated with severe fibrosis.[12] In a subsequent study which used the same SNPs and focused on Caucasian patients with well-characterised liver histology, a different panel of seven SNPs was found to predict the risk of developing cirrhosis, and this panel needs to be validated in prospective studies.[13]

Liver fibrosis is the consequence of a generalised wound-healing response of hepatic tissue against repeated injury, which results in the formation of scar tissue instead of normal parenchyma.[14] This process is characterised by an imbalance between matrix synthesis (ie, fibrogenesis) and matrix degradation (ie, fibrolysis), and leads to an accumulation of a large variety of matrix proteins, including collagens, proteoglycans and glycoproteins.[15 16] Activated hepatic stellate cells, portal fibroblasts and myofibroblasts of bone marrow origin are the major collagen-producing cells in the injured liver.[14] A number of molecules and regulatory pathways are involved in this complex process of fibrogenesis/fibrolysis.[14–16] Therefore, we hypothesised that variations in human genes involved in fibrogenesis/fibrolysis might account for interindividual variability in development of liver fibrosis among patients with chronic HCV infection. In this work we focused on the role of polymorphisms located in a panel of 36 genes (table 1) encoding either enzymes involved in extracellular matrix turnover (matrix metalloproteinases and their inhibitors) or some cytokines known to exhibit profibrogenic (transforming growth factor β (TGFβ) and related molecules) or anti-fibrogenic (interferon γ (IFNγ) and its receptors) activity. In addition to classic case–control analysis, in this study we analysed the influence of SNPs directly on the time of progression by restricting our sample to patients with a known presumed date of infection and without known risk factors of fibrosis progression such as chronic alcohol intake, associated infections or metabolic syndrome.

Table 1 List of the genes investigated in the association study

Gene name	Abbreviation
Alpha-2-macroglobulin	A2M
Angiotensinogen	AGT
Interferon gamma	IFNG
Interferon gamma receptor 1	IFNGR1
Interferon gamma receptor 2	IFNGR2
Keratin 8	KRT8
Latent transforming growth factor beta binding protein 1	LTBP1
Latent transforming growth factor beta binding protein 2	LTBP2
Latent transforming growth factor beta binding protein 3	LTBP3
Latent transforming growth factor beta binding protein 4	LTBP4
Matrix metalloproteinase 1	MMP1
Matrix metalloproteinase 2	MMP2
Matrix metalloproteinase 3	MMP3
Matrix metalloproteinase 7	MMP7
Matrix metalloproteinase 8	MMP8
Matrix metalloproteinase 9	MMP9
Matrix metalloproteinase 10	MMP10
Matrix metalloproteinase 11	MMP11
Matrix metalloproteinase 12	MMP12
Matrix metalloproteinase 13	MMP13
Matrix metalloproteinase 14	MMP14
Matrix metalloproteinase 15	MMP15
Matrix metalloproteinase 16	MMP16
Matrix metalloproteinase 17	MMP17
Matrix metalloproteinase 24	MMP24
Matrix metalloproteinase 25	MMP25
Transforming growth factor beta 1	TGFB1
Transforming growth factor beta 2	TGFB2
Transforming growth factor beta 3	TGFB3
Transforming growth factor beta receptor I	TGFBR1
Transforming growth factor beta receptor II	TGFBR2
Transforming growth factor beta receptor III	TGFBR3
Tissue inhibitor of metalloproteinase 1	TIMP1
Tissue inhibitor of metalloproteinase 2	TIMP2
Tissue inhibitor of metalloproteinase 3	TIMP3
Tissue inhibitor of metalloproteinase 4	TIMP4

PATIENTS AND METHODS

Patients

We recruited adult Caucasian patients (>18 years of age) with chronic HCV infection defined as the presence of circulating HCV RNA tested by reverse transcriptase PCR. The patients were gathered in two steps. First, we conducted a prospective enrolment of patients from the hepatology units of Necker Hospital in Paris and St Joseph Hospital in Marseille (sample A). The criteria for the inclusion of patients were (*a*) an available liver biopsy before any treatment; (*b*) a known presumed date of HCV acquisition (date of the first exposure to blood products, or of beginning of intravenous drug (IVD) use); (*c*) a low alcohol consumption (less than three or less than two standard drinks a day for men or women, respectively); (*d*) absence of co-infection with HIV or hepatitis B virus; (*e*) absence of any coexisting chronic liver disease or hepatocellular carcinoma. Clinical risk factors, history of HCV acquisition and of alcohol consumption (assessed using time-line follow-back interview) were recorded through face-to-face interviews conducted by doctors trained in addiction problems. In a second step, we gathered additional patients from an existing cohort from the hepatology unit of Pitié-Salpêtrière Hospital in Paris (sample B). The inclusion criteria were the same as for sample A except that the presumed date of infection was not known for all patients. The study was approved by the appropriate institutional review boards, and written informed consent was obtained from all patients.

Most of the enrolled patients had been followed up for a number of years in the corresponding clinics and had had a liver biopsy at the time of their first evaluation. The stage of fibrosis was assessed from liver biopsy samples using METAVIR units, and graded on a five-point scale from 0 to 4.[17] For this study, in order to optimise the phenotype definition, we excluded patients with grade 2, and retained only patients with grades 0 or 1 (F0–1 patients) referred to as having no fibrosis, and patients with grades 3 or 4 (F3–4) referred to as having severe fibrosis. For patients who had had several biopsies, we used either the first biopsy specimen showing severe fibrosis (F3–4 patients) or the last biopsy showing no fibrosis without any treatment (for F0–1 patients). The duration of infection was estimated from the presumed year of HCV acquisition (eg, first exposure to blood transfusion, or to IVD use) to the year of the relevant biopsy.

Genotyping and sequencing methods

In this study we focused on the role of polymorphisms located in a panel of genes encoding either enzymes involved in