Escherichia coli–expressed Ppr was purified (~85% pure) on a HiTrap metal-chelating column (Pharmacia). Ppr was dialyzed and stored at −80°C in 400 mM KCl, 50 mM tris-HCl (pH 7.9), 10 mM $MgCl_2$, 0.1 mM EDTA, and 50% glycerol. *p*-Hydroxycinnamic acid was attached as described [S. Devanathan *et al.*, *Arch. Biochem. Biophys.* **340**, 83 (1997)].

16. T. E. Meyer, G. Tollin, J. H. Hazzard, M. A. Cusanovich, *Biophys. J.* **56**, 559 (1989); T. E. Meyer, G. Tollin, T. P. Causgrove, P. Cheng, R. E. Blankenship, *ibid.* **59**, 988 (1991); L. Ujj *et al.*, *Biophys. J.* **75**, 406 (1998).
17. Samples containing 60 μg of protein were diluted with storage buffer (*15*) to a volume of 288 μl. Half of the reaction mixture was incubated in the dark and the other in >400-nm light (5.5 mW/cm²) for 30 min at room temperature. [γ-^{32}P]ATP mix {1.0 mM adenosine 5′-triphosphate (ATP) and 0.024 mM [γ -^{32}P] ATP (7000 Ci/mmole, ICN)} was added to initiate the reactions. Partitions of 12 μl (2.5 μg of protein) were removed at time intervals, mixed with SDS–loading buffer, and placed on ice. The samples were later fractionated by SDS–polyacrylamide gel electrophoresis. ^{32}P-Labeled protein bands were quantified by PhosphorImager analysis (Molecular Dynamics, Sunnyvale, CA).
18. Z. Jiang and C. E. Bauer, unpublished data.
19. The extinction coefficient of Ppr at 280 nm was 101,820 M^{-1} cm^{-1} and that of *p*-hydroxycinnamate was 45,000 M^{-1} cm^{-1}, so the maximum ratio of absorption of Ppr at 434 nm to that at 280 nm is 0.442. The ratio of absorption of reconstituted Ppr at 434 nm to that at 280 nm was 0.29, giving a ~60% efficiency of attachment (*7*).
20. *ppr* amino acid residues 114 to 750 were replaced with a spectinomycin resistance gene in the suicide vector pGmLacZ. After delivery by conjugation (*21*), double recombinants were selected and confirmed by PCR analysis.
21. Z. Y. Jiang, H. Gest, C. E. Bauer, *J. Bacteriol.* **179**, 5720 (1997).
22. C. R. Martin, *Int. Rev. Cytol.* **147**, 233 (1993).
23. GenBank Accession number AF121274.
24. The *R. centenum chs* promoter was amplified by PCR with primers TAGGTACCGATGAACAGCCAGGCGAG and TAGCATGCCGTGAAAACGGGGGAGAG. The PCR product was cloned into the *lacZ* reporter plasmid pZJD11 (*21*) and maintained with antibiotics.
25. *R. centenum* was cultured in peptone–yeast extract–soytone (*12*) and illuminated with infrared light (>700 nm). White light was provided with a 400-W pressure sodium Lumalux lamp (LU400, Osram Sylvania, Danvers, MA) and blue light (400 to 450 nm) by passing the white light through filters. Cells were assayed for β-Gal activity as described (*21*).
26. Supplemental Web material is available at www.sciencemag.org/feature/data/1039035.shl.
27. J. K. Xiong, K. Inou, C. E. Bauer, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14851 (1998).
28. We thank T. Meyer and R. Hangarter for comments. Supported by NIH grant GM 40941 (C.E.B.) and a NSF postdoctoral fellowship (B.G.R.).

3 February 1999; accepted 14 June 1999

An Allele of *COL9A2* Associated with Intervertebral Disc Disease

Susanna Annunen,[1] Petteri Paassilta,[1] Jaana Lohiniva,[1] Merja Perälä,[1] Tero Pihlajamaa,[1] Jaro Karppinen,[2] Osmo Tervonen,[3] Heikki Kröger,[5] Seppo Lähde,[3] Heikki Vanharanta,[2] Lasse Ryhänen,[4] Harald H. H. Göring,[6] Jürg Ott,[6,7] Darwin J. Prockop,[8] Leena Ala-Kokko[1,8]*

Intervertebral disc disease is one of the most common musculoskeletal disorders. A number of environmental and anthropometric risk factors may contribute to it, and recent reports have suggested the importance of genetic factors as well. The *COL9A2* gene, which codes for one of the polypeptide chains of collagen IX that is expressed in the intervertebral disc, was screened for sequence variations in individuals with intervertebral disc disease. The analysis identified a putative disease-causing sequence variation that converted a codon for glutamine to one for tryptophan in six out of the 157 individuals but in none of 174 controls. The tryptophan allele cosegregated with the disease phenotype in the four families studied, giving a lod score (logarithm of odds ratio) for linkage of 4.5, and subsequent linkage disequilibrium analysis conditional on linkage gave an additional lod score of 7.1.

Intervertebral disc disease is among the most common musculoskeletal disorders. It is a major cause of work disability and an extremely costly health care problem. It is typically associated with sciatica, which has a prevalence of about 5% in Finland (*1*). Sciatica is defined as pain caused by a lesion of the spinal nerve root radiating along the femoral or sciatic nerve from the back into the dermatome of the root. It is usually related to either disc protrusion or herniation, which will cause both chemical and mechanical irritation of the nerve root. Even though there is an association between various environmental and anthropometric risk factors and sciatica, their effects are modest (*2, 3*). A recent twin study suggests that genetic factors may be involved in the pathogenesis of intervertebral disc disease and sciatica (*3*). This is supported by findings of a considerable genetic predisposition to early-onset sciatica and lumbar disc herniation in certain families (*4*). Therefore, intervertebral disc disease appears to be similar to other common complex diseases with multiple genetic forms and a high phenocopy rate, such as breast cancer and Alzheimer's disease.

Intervertebral discs contain an abundant extracellular matrix of proteoglycans and collagens (*5*). The outer layer, the annulus fibrosus, consists mainly of collagen I, and the interior structure of the disc, the nucleus pulposus, is about 50% proteoglycan, mainly aggrecan, and 20% collagen II. Both contain small amounts of collagen IX. Recent results indicate that mutations in collagen IX and aggrecan can cause age-related disc degeneration and herniation in mice (*6, 7*).

Collagen IX is a heterotrimer of three α chains, α1(IX), α2(IX), and α3(IX), encoded by the genes *COL9A1*, *COL9A2*, and *COL9A3*, respectively. It consists of three collagenous (COL1 to COL3) and four noncollagenous (NC1 to NC4) domains (*8*). The COL2 domain is covalently linked to collagen II fibrils (*9*). Collagen IX is thought to serve as a bridge between collagens and noncollagenous proteins in tissues.

To study the role of collagen IX in intervertebral disc disease and associated sciatica, we selected 157 unrelated Finnish individuals (ages 19 to 78; mean = 44, SD = 13) with unilateral pain of duration over 1 month radiating from the back to below the knee (dermatomes L4, L5, and S1). Therefore, the subjects had the most characteristic symptom of herniated intervertebral disc (*1*). After clinical evaluation, 156 of them were examined by magnetic resonance imaging (MRI) and one was evaluated by computerized tomography (CT) (*10*). Radiologically detectable intervertebral disc disease was present at the time of the examination in 73% of the cases. Initially, *COL9A2* was screened for sequence variations by conformation-sensitive gel electrophoresis (CSGE) in 10 patients (*11*), yielding a unique CSGE pattern in the polymerase chain reaction (PCR) products for exon 19 in one case (Fig. 1). Sequencing (*12*) indicated a

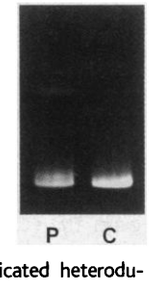

Fig. 1. CSGE analysis of exon 19 of the *COL9A2* gene. The exon and its flanking sequences were amplified by PCR from the proband (P) and a control (C) using primers specific for intron 18 (5′-TGGATCTCAGTTTCCCTACCTG, −92 to −71 in infron 18) and infron 19 (5′-CAAGAGGTGGTGATTGAGCAAGAGC, +99 to +75 in infron 19). The analysis indicated heteroduplexes in the proband's sample.

[1]Collagen Research Unit, Biocenter and Department of Medical Biochemistry; [2]Department of Physical Medicine and Rehabilitation; [3]Diagnostic Radiology; [4]Internal Medicine, University of Oulu, 90220 Oulu, Finland. [5]Department of Surgery, University of Kuopio, 70211 Kuopio, Finland. [6]Department of Genetics and Development, Columbia University, New York, NY 10032, USA. [7]Laboratory of Statistical Genetics, Rockefeller University, New York, NY 10021, USA. [8]Center for Gene Therapy, MCP Hahnemann University, Philadelphia, PA 19102, USA.

*To whom correspondence should be addressed (in Finland). E-mail: leena.ala-kokko@oulu.fi

heterozygous substitution of Trp for either Gln^{326} or Arg^{326} in the COL2 domain (Fig. 2). This finding was surprising, because Trp is rarely found in collagenous domains and there are no Trp residues in the collagenous domains of collagen IX in humans or in the mouse (*13*). The remaining 147 patients with sciatica were then analyzed, and the Trp allele was found in five of them (Table 1).

Exon 19 was also analyzed in 174 unrelated Finnish controls. All patients and controls were from the same region of Finland. The control group consisted of 101 asymptomatic subjects (ages 21 to 73; mean = 37, SD = 10), 54 with osteoarthritis, and 19 with various chondrodysplasias but no history of sciatica. Not a single Trp allele was found among the 348 chromosomes analyzed (Table 1).

Coinheritance of the Trp allele and the phenotype was studied in the families of four original patients. The two other families were not available for the study. The families were evaluated clinically and by CT (one family) or MRI (three families) (*10*) and analyzed for the presence of the Trp allele. All family members who had inherited the allele had intervertebral disc disease (Fig. 3).

To evaluate the statistical evidence for a connection between the Trp allele in the *COL9A2* gene and intervertebral disc disease, we conducted linkage and linkage disequilibrium analysis. Artificial pedigrees were created from the case-control data (*14, 15*), which made it possible to perform the linkage and linkage disequilibrium analysis on the pedigrees and the singletons jointly, using standard analysis software (*16*). This is more satisfactory than splitting the data into subsets for different analyses, as is generally done, because joint analysis of linkage and linkage disequilibrium can give more information than the sum of the parts (*15, 17*).

A dominant inheritance model with full penetrance and a high phenocopy rate [disease allele (D) frequency 0.0024; penetrances $f_{DD} = 1, f_{D+} = 1, f_{++} = 0.0434$] was chosen for the analysis. The model predicts a disease prevalence of 4.8%, which is the estimated frequency of sciatica in Finland (*1*). A high phenocopy rate was chosen, as the etiology of intervertebral disc disease is believed to be multifactorial (*2*). The modeled disease locus explains only 10% of all cases. The pedigree members were assigned affected status according to the clinical and radiological findings (*18*). All singleton subjects without intervertebral disc disease were used as controls, as no statistically significant differences in allele frequencies at the polymorphism were found among the three groups (Table 1).

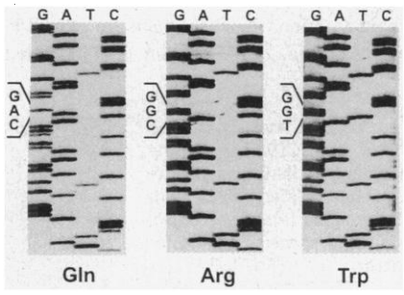

Fig. 2. Sequences for exon 19 of the *COL9A2* gene. The PCR products were cloned and sequenced, indicating three alleles, containing a CAG codon for Gln, a CGG codon for Arg, and a TGG codon for Trp.

Table 2 summarizes the linkage and linkage disequilibrium results. Linkage analysis gave a lod score of $Z_1 = 4.5$ ($P < 10^{-5}$) at a recombination fraction of 0.12 (*19, 20*). After linkage had been demonstrated, linkage disequilibrium analysis given the presence of linkage was performed (*21*). An additional lod score for linkage disequilibrium of $Z_2 = 7.1$ ($P < 10^{-7}$) was obtained (*22*). The joint lod score is therefore $Z_1 + Z_2 = 11.6$ ($P < 10^{-10}$).

To investigate the extent to which the results were dependent on the assumed disease model, we varied individual model parameters while keeping the predicted population prevalence unchanged at 4.8%. The linkage and linkage disequilibrium lod scores remained high for a dominant model when the proportion of cases explained by the disease locus was varied and when high but incomplete penetrance was used (Table 2).

Because we speculated that the Trp allele may actually itself be the disease-predisposing allele, the analyses were repeated in a way that addressed this question more directly. Under this hypothesis, recombination between the disease locus and the polymorphism is clearly impossible, as they are the same locus. Furthermore, under this hypothesis, the frequency of the Trp allele is necessarily equal to the frequency of the disease-causing allele, as they are really the same, and the disease allele and the Trp allele are in complete linkage disequilibrium. The previous analyses were therefore repeated under these constraints (*23*). The results are shown in Table 3. Although these lod scores are much lower than those in Table 2, this does not necessarily contradict the hypothesis of causality: The reason is that the results obtained in "parametric" linkage or linkage disequilibrium analysis depend on the chosen disease model. If the disease model is incorrect—and in practice it is impossible to specify the model with 100% accuracy, in particular for common diseases such as intervertebral disc disease—the obtained lod scores may be greatly reduced and the recombination fraction overestimated (*24, 25*). This could easily explain why the lod scores are lower in Table 3 than in Table 2. The same phenomenon explains the high rate of false negative results in multipoint "parametric" linkage analysis (*24, 25*). On the other hand, an incorrect disease model does not systematically lead to false positive evidence of linkage (*26*).

The statistical analyses provided strong evidence that the Trp allele is associated with the phenotype. All 26 individuals carrying the Trp allele had intervertebral disc disease. The results do not, however, prove a direct causal role for the Trp allele in the etiology of the disease, and other interpretations remain open. It is conceivable, for instance, that the true disease locus may lie in close physical proximity to *COL9A2*, with the disease-predisposing allele in linkage disequilibrium with the Trp allele. The fact that the families had some individuals who were symptomatic but did not have the Trp allele does not contradict the hypothesis that the Trp allele causes the disease. Indeed, this finding is not even surprising, as sciatica and intervertebral disc disease are very common and have a

Table 1. Allele counts and frequencies of the sequence variations at α2-326 in collagen IX.

Subject group	Allele counts (allele frequencies)			Total
	Trp	Gln	Arg	
Disc disease	6 (0.019)	228 (0.726)	80 (0.255)	314 (1.000)
Unaffected	0 (0.000)	149 (0.738)	53 (0.262)	202 (1.000)
Osteoarthritis	0 (0.000)	83 (0.769)	25 (0.231)	108 (1.000)
Chondrodysplasias	0 (0.000)	26 (0.684)	12 (0.316)	38 (1.000)

Table 2. Summary of lod scores obtained in linkage and linkage disequilibrium analyses.

Proportion of cases explained by modeled disease locus	Disease allele frequency	Penetrances for disease locus genotypes (DD, D+, ++)	Linkage (Z_1) (*20*) ($\hat{\theta}$)†	Linkage disequilibrium given linkage (Z_2) (*22*) ($\hat{\theta}$)†
0.5	0.0121	1.00, 1.00, 0.025	4.519 (0.12)	4.328 (0.11)
0.2	0.0048	1.00, 1.00, 0.039	4.514 (0.12)	5.902 (0.11)
0.1*	0.0024	1.00, 1.00, 0.043	4.508 (0.12)	7.072 (0.11)
0.1	0.0027	0.90, 0.90, 0.043	3.826 (0.12)	6.949 (0.11)
0.1	0.0032	0.75, 0.75, 0.043	2.958 (0.13)	6.986 (0.12)

*Main disease model. †Estimated recombination fraction.

Table 3. Summary of lod scores obtained in linkage and linkage disequilibrium analyses when testing whether the Trp allele is itself a disease-causing allele.

Proportion of cases explained by modeled disease locus	Disease allele frequency	Penetrances for disease locus genotypes (DD, D+, ++)	Tight linkage (Z_3) (*24, 25*)	Complete linkage disequilibrium given tight linkage (Z_4) (*24, 25*)
0.5	0.0121	1.00, 1.00, 0.025	2.208	−34.105
0.2	0.0048	1.00, 1.00, 0.039	2.358	−2.278
0.1*	0.0024	1.00, 1.00, 0.043	1.283	6.395
0.1	0.0027	0.90, 0.90, 0.043	0.921	6.154
0.1	0.0032	0.75, 0.75, 0.043	0.601	5.733

*Main disease model.

prevalence of about 5% in the Finnish population (*1*). Any one locus is therefore likely to be responsible for only a small proportion of all affected individuals, and different disease-predisposing alleles will often segregate within one pedigree. In the disease model chosen here, the locus accounted for 10% of the disease prevalence, as stated above.

To exclude the possibility that other sequence variations in the gene might cause the disease, we used CSGE to analyze all the exons and exon boundaries of *COL9A2* in patients with the Trp substitution (*11*). The analysis did not identify any other possible causes of disease in the coding sequences or RNA splice sites, but it indicated two additional, presumably neutral polymorphisms: an A to G change affecting the third nucleotide in the codon for proline (CCA to CCG) at nucleotide 9 in exon 21, and a G to A change at nucleotide +17 in intron 30. These were analyzed in 90 and 286 individuals, respectively, and were found in all six original patients with the Trp allele but in none of the other individuals analyzed. The coexistence of the Trp allele and the two other rare sequence variants indicates that the individuals with the Trp allele have inherited the same, relatively rare, ancestral haplotype.

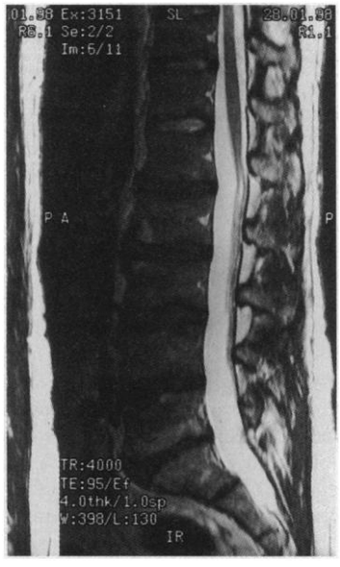

Fig. 3. MRI findings of a 45-year-old male who has the Trp allele, showing endplate degeneration in the L2-L3, L4-L5, and L5-S1 interspaces, and protrusions in all interspaces except L2-L3.

There are several possible mechanisms by which Trp substitution could contribute to the disease. As the most hydrophobic amino acid, Trp may disrupt the collagen triple helix, and it is also possible that it could interfere with the interaction between collagens IX and II or prevent the action of lysyl oxidase, which catalyzes cross-link formation (*27*), because the Trp in the α2(IX) chain of collagen IX is located only three amino acid residues NH_2-terminal of the covalent lysine-derived cross-link between the α3(IX) chain and collagen II. In addition, the role of collagen IX in intervertebral disc disease is supported by the finding of intervertebral disc degeneration and herniation in a long-term follow-up of transgenic mice expressing a *Col9a1* gene with a large in-frame deletion (*6*).

Thus far, only two collagen IX mutations, one in the *COL9A2* and one in the *COL9A3* gene, have been reported in humans (*28*). Both mutations cause a skipping of exon 3 and result in multiple epiphyseal dysplasia. They lead to similar deletions of 12 amino acids in the COL3 domain of the molecule, which suggests the importance of this domain in the pathogenesis of the dysplasia.

There are now a large number of examples that illustrate the difficulty of relating genotypes to phenotypes caused by mutated genes, including mutated collagen genes (*27, 29*). For example, different mutations in collagen II cause phenotypes ranging from only ocular manifestations or osteoarthritis to various chondrodysplasias with or without ocular symptoms (*29*). Thus, it is not surprising that Trp for Gln substitution in the COL2 domain and splicing mutation in the COL3 domain of the collagen IX molecule can cause different phenotypes.

References and Notes

1. M. Heliövaara *et al.*, *J. Epidemiol. Community Health* **41**, 251 (1987).
2. M. Heliövaara, *Ann. Med.* **21**, 257 (1989).
3. M. C. Battié *et al.*, *Spine* **20**, 2601 (1995).
4. C. L. Nelson, C. J. Janecki, P. L. Gildenberg, G. Sava, *Clin. Orthop.* **88**, 142 (1972); G. P. Varlotta, M. D. Brown, J. L. Kelsey, A. L. Golden, *J. Bone Jt. Surg. Am. Vol.* **73**, 124 (1991); H. Matsui, N. Terahata, H. Tsuji, N. Hirano, Y. Naruse, *Spine* **17**, 1323 (1992); R. Scapinelli, *Acta Orthop. Belgica* **59**, 371 (1993).
5. D. R. Eyre and H. Muir, *Biochim. Biophys. Acta* **492**, 29 (1977); J. P. G. Urban, in *Musculoskeletal Soft Tissue Aging: Impact on Mobility*, J. A. Buckwalter, V. M. Goldberg, S. L.-Y. Woo, Eds. (American Academy of Orthopaedic Surgeons, Rosemont, IL, 1993), pp. 391–412; J. A. Buckwalter, *Spine* **20**, 1307 (1995).
6. T. Kimura *et al.*, *Int. Orthop.* **20**, 177 (1996).
7. H. Watanabe, K. Nakata, K. Kimata, I. Nakanishi, Y. Yamada, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6943 (1997).
8. M. van der Rest and R. Mayne, in *Structure and Function of Collagen Types*, R. Mayne and S. E. Burgeson, Eds. (Academic Press, Orlando, FL, 1987), pp. 195–221; L. M. Shaw and B. R. Olsen, *Trends Biochem. Sci.* **16**, 191 (1991); M. van der Rest and R. Mayne, *J. Biol. Chem.* **263**, 1615 (1988); L. Vaughan *et al.*, *J. Cell. Biol.* **106**, 991 (1988).
9. D. R. Eyre, S. Apon, J. J. Wu, L. H. Ericsson, K. A. Walsh, *FEBS Lett.* **220**, 337 (1987); J. J. Wu, P. E. Woods, D. R. Eyre, *J. Biol. Chem.* **267**, 23007 (1992); M. Diab, J. J. Wu, D. R. Eyre, *Biochem. J.* **314**, 327 (1996).
10. MRI scans obtained with a 1.5-T imaging system (Signa, General Electric) consisted of sagittal images with TR/TE of 4000/95 ms and axial images with TR/TE of 640/14 ms. CT images (Hi Speed Advantage, GE Medical Systems) consisted of scans through the L2-L3 interspace to the L5-S1 interspace. All MRI and CT images were read by two experienced neuroradiologists blinded to the results of the mutation analysis and clinical history and physical status of the patient. Clinical assessment was also performed blinded to the results of the mutation analysis. The following radiological findings were considered as indications of intervertebral disc disease: (i) disc extrusion; (ii) any herniation at two or more levels; (iii) endplate degeneration at one or more levels in patients under 30 years old, at two or more levels in patients 30 to 50 years old, or at four or more levels in patients over 50 years old; or (iv) bulging or protrusion (or both) at four or more levels [D. Weishaupt, M. Zanetti, J. Hodler, N. Boos, *Radiology* **209**, 661 (1998); M. C. Jensen *et al.*, *N. Engl. J. Med.* **331**, 69 (1994)].
11. Genomic DNA was extracted from blood leukocytes and used to screen for sequence variations in the *COL9A2* gene by CSGE. PCR primers to amplify the exonic sequences and the flanking sequences were designed on the basis of the genomic sequences [T. Pihlajamaa *et al.*, *Matrix Biol.* **17**, 237 (1998)], and PCR and CSGE analysis were performed as described [J. Körkkö, S. Annunen, T. Pihlajamaa, D. J. Prockop, L. Ala-Kokko, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1681 (1998)].
12. PCR products that contained heteroduplexes on the CSGE analysis were sequenced (T7 Sequenase PCR Product Sequencing Kit, USB, or ABI Prism 377 and dRhodamine Terminator Cycle Sequencing Ready Reaction Kit, Perkin Elmer). Some of the PCR products were cloned and sequenced (T7 Sequencing Kit, Pharmacia).
13. Y. Muragaki, T. Kimura, Y. Ninomiya, B. R. Olsen, *Eur. J. Biochem.* **192**, 703 (1990); M. Perälä, M. Hänninen, J. Hästbacka, K. Elima, E. Vuorio, *FEBS Lett.* **319**, 177 (1993); R. G. Brewton *et al.*, *Genomics* **30**, 329 (1995); M. Perälä *et al.*, *J. Biol. Chem.* **269**, 5064 (1994); I. Rokos, Y. Muragaki, M. Warman, B. R. Olsen, *Matrix Biol.* **14**, 1 (1994).
14. Artificial pedigrees were created by coding two case or control individuals as the parents of a hypothetical offspring of unknown phenotype and genotype (*15*). Creating such pedigrees (that is, triads) does not create false, artificial information. Such pedigrees do not provide any linkage information whatsoever, but they do give information about linkage disequilibrium, just as if the unrelated case and control individuals were analyzed separately.
15. K. Kainulainen *et al.*, *Hypertension* **33**, 844 (1999).
16. Linkage and linkage disequilibrium analyses were performed using the ILINK program of the FASTLINK software package [G. M. Lathrop, J.-M. Lalouel, C. Julier, J. Ott, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3443 (1984); A. Becker, D. Geiger, A. A. Schaffer, *Hum. Hered.* **48**, 49 (1998)].
17. E. Hellsten *et al.*, *Genomics* **16**, 720 (1993); P. J. Tienari, J. D. Terwilliger, J. Ott, J. Palo, L. Peltonen,

ibid. **19**, 320 (1994); H. H. H. Göring and J. D. Terwilliger, in preparation.

18. Pedigree members who tested positive for the disease in either the clinical or the radiological test were coded as affected, and those testing negative in both tests were coded as unaffected. Individuals who were only examined by one method (clinical or radiological) and were found to be normal were coded as having unknown phenotypic status. Unexamined individuals, including married-in spouses and deceased individuals from earlier generations, were likewise coded as unknown in their phenotype.
19. $Z_1 = \log[L(\hat{\vartheta})/L(\vartheta = 0.5)]$ is a test for linkage. [$L(\vartheta)$ denotes the likelihood as a function of the recombination fraction ϑ.] The allele frequencies of the observed polymorphism were treated as nuisance parameters using "profile" likelihoods [R. Royall, *Statistical Evidence: A Likelihood Paradigm* (Chapman & Hall, London, 1997), pp. 158–161]; that is, the likelihood was maximized over these allele frequencies independently under the null hypothesis of no linkage and under the alternative hypothesis of linkage [(*25*), pp. 186–187; H. H. H. Göring and J. D. Terwilliger, in preparation]. A lod score of 3 is generally considered to be significant [N. E. Morton, *Am. J. Hum. Genet.* **7**, 277 (1955)] and asymptotically corresponds to a pointwise *P* value of 0.0001 [J. Chotai, *Ann. Hum. Genet.* **48**, 359 (1984)].
20. Affecteds-only analysis provides only little evidence for linkage (lod score 0.5), whereas the linkage disequilibrium results conditional on linkage remain significant (lod score 6.3) (data not shown). It is not that surprising, however, that the linkage lod score is reduced, because much linkage information is provided by unaffected individuals when the disease is dominant with high penetrance. To avoid spurious results, we only coded pedigree members as unaffected when they were found to be normal in two different diagnostic schemes, as mentioned above.
21. $Z_2 = \log[L(\hat{\vartheta}, \hat{\delta})/L(\hat{\vartheta}, \delta = 0)]$ is a test for linkage disequilibrium conditional on linkage. δ here denotes linkage disequilibrium. In the denominator [identical to the numerator of Z_1 (*20*), only written differently], the likelihood is maximized over the marker allele frequencies using ILINK, as before. In the numerator, the likelihood is maximized over the disease-marker haplotype frequencies, keeping the marginal frequency of the disease-predisposing allele constant. This maximization was performed with the EH program [(*25*), pp. 200–203] using only unrelated individuals, as ILINK would have changed the marginal disease-allele frequency also. Although the haplotype frequencies estimated with the EH program may not be the maximum likelihood estimates, this procedure is conservative in that the true maximum likelihood estimates are used in the denominator. This lod score has two degrees of freedom. (There are five free parameters in the numerator: ϑ and four haplotype frequencies, the frequencies of the two remaining haplotypes being constrained by the marginal allele frequencies of the disease locus. There are three free parameters in the denominator: ϑ and two allele frequencies, with the frequency of the remaining allele being constrained because the allele frequencies need to sum to 1.) For interpretation, the *P* value for the observed lod score is given in the text.
22. The evidence for linkage disequilibrium conditional on linkage using the whole data set (families and unrelated singletons) is much higher than if the case and control data had been analyzed separately. Using Fisher's "exact" test to test for equality of Trp allele frequencies among cases (6 of 314 chromosomes) and controls (0 of 348 chromosomes) only gives a *P* value of about 0.01. The reason for this apparent discrepancy is that, in joint linkage and linkage disequilibrium analysis, the linkage disequilibrium provides not only phase information for pedigree founders, but also genotype information for untyped founders. For this reason, joint analysis of linkage and linkage disequilibrium is often much more than the sum of the parts, sometimes leading to much higher lod scores (*15*, *17*).
23. $Z_3 = \log[L(\vartheta = 0)/L(\vartheta = 0.5)]$ is a test of complete linkage. In the denominator, the likelihood was maximized over the allele frequencies of the polymorphism. In the numerator, however, the frequency of the Trp allele was fixed to be equal to the frequency of the disease-predisposing allele (because we wanted to test whether the Trp allele itself is a disease-causing allele), and the likelihood was maximized over the remaining allele frequencies of the polymorphism. $Z_4 = \log[L(\vartheta = 0, \delta = \text{complete})/L(\vartheta = 0, \delta = 0)]$ is a test of complete linkage disequilibrium conditional on complete linkage. In the denominator (identical to the numerator in Z_3, only written differently), the frequency of the Trp allele was set to be equal to the frequency of the disease allele and the likelihood maximized over the remaining marker allele frequencies. In the numerator, complete linkage disequilibrium was assumed. In other words, all Trp chromosomes carry the disease allele and vice versa, and all non-Trp chromosomes carry the "normal" allele at the disease locus and vice versa. The marginal frequency of the haplotype with the disease-predisposing allele was kept constant, and the likelihood was maximized over the remaining haplotype frequencies using ILINK (*16*).
24. F. Clerget-Darpoux, C. Bonaiti-Pellie, J. Hochoz, *Biometrics* **42**, 393 (1986); N. Risch and L. Giuffra, *Hum. Hered.* **42**, 77 (1992); H. H. H. Göring and J. D. Terwilliger, in preparation.
25. J. D. Terwilliger and J. Ott, *Handbook of Human Genetic Linkage* (Johns Hopkins Univ. Press, Baltimore, MD, 1994).
26. J. A. Williamson and C. I. Amos, *Genet. Epidemiol.* **7**, 309 (1990).
27. C. M. Kielty, I. Hopkinson, M. E. Grant, in *Connective Tissue and Its Heritable Disorders*, P. M. Royce and B. Steinmann, Eds. (Wiley-Liss, New York, 1993), pp. 103–147; N. Nagan and H. M. Kagan, *J. Biol. Chem.* **269**, 22366 (1994).
28. Y. Muragaki *et al.*, *Nature Genet.* **12**, 103 (1996); P. Paassilta *et al.*, *Am. J. Hum. Genet.*, **64**, 1036 (1999).
29. H. Kuivaniemi, G. Tromp, D. J. Prockop, *Hum. Mutat.* **9**, 300 (1997).
30. We thank J. D. Terwilliger for helpful comments and A. Härju for her expert technical assistance. Supported by grants from the Academy of Finland (L.A.-K. and M.P.) and the Arthritis Foundation (D.J.P.) and by NIH grants AR39740 (D.J.P.) and HG00008 (J.O.).

24 February 1999; accepted 4 June 1999

Prevention of Graft Versus Host Disease by Inactivation of Host Antigen-Presenting Cells

Warren D. Shlomchik,[1] Matthew S. Couzens,[1] Cheng Bi Tang,[1] Jennifer McNiff,[3] Marie E. Robert,[4] Jinli Liu,[5,6] Mark J. Shlomchik,[5,6]* Stephen G. Emerson[1,2]*†

Graft versus host disease, an alloimmune attack on host tissues mounted by donor T cells, is the most important toxicity of allogeneic bone marrow transplantation. The mechanism by which allogeneic T cells are initially stimulated is unknown. In a murine allogeneic bone marrow transplantation model it was found that, despite the presence of numerous donor antigen-presenting cells, only host-derived antigen-presenting cells initiated graft versus host disease. Thus, strategies for preventing graft versus host disease could be developed that are based on inactivating host antigen-presenting cells. Such strategies could expand the safety and application of allogeneic bone marrow transplantation in treatment of common genetic and neoplastic diseases.

Allogeneic bone marrow transplantation (alloBMT) has revolutionized the treatment of hematopoietic malignancies, inherited hematopoietic disorders, and aplastic anemia. Unfortunately, graft versus host disease (GVHD) remains a major toxicity that greatly limits the application and efficacy of alloBMT (*1*). Most patients who undergo alloBMT receive stem cells from major histocompatibility complex (MHC)–identical donors. In these patients, GVHD is initiated by donor T cells that recognize a subset of host peptides, called minor histocompatibility antigens (miHAs), which are derived from the expression of polymorphic genes that distinguish host from donor (*2*). Presently, therapy for GVHD is limited to immunosuppression directed largely against T cells (*3*).

T cell responses are initiated on antigen-presenting cells (APCs). In alloBMT the unusual situation arises in which both host- and donor-derived APCs are present. We therefore examined the roles of host- and donor-derived APCs in initiating GVHD with the goal of identifying another target for GVHD prevention (*4*–*6*). Because recipient hematopoiesis is ablated by cytotoxic therapies before the transplant, and more intensive radiation augments GVHD, donor APCs might be crucial (*7*). On the other hand, peptides presented to $CD8^+$ T cells by MHC class I molecules are derived primarily from endog-

[1]Department of Medicine, Hematology and Oncology Division, [2]Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA. [3]Department of Dermatology, [4]Department of Pathology, [5]Department of Laboratory Medicine, [6]Department of Immunobiology, Yale University School of Medicine, New Haven, CT 06510, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: emersons@mail.med.upenn.edu