Am. J. Hum. Genet. 73:812–822, 2003

Genomewide Linkage and Linkage Disequilibrium Analyses Identify *COL6A1,* on Chromosome 21, as the Locus for Ossification of the Posterior Longitudinal Ligament of the Spine

Toshihiro Tanaka,[1,4,*] Katsunori Ikari,[1,2,*] Kozo Furushima,[1,4] Akihiro Okada,[4] Hiroshi Tanaka,[6] Ken-Ichi Furukawa,[5] Kenichi Yoshida,[1] Toshiyuki Ikeda,[3] Shiro Ikegawa,[3] Steven C. Hunt,[7] Jun Takeda,[8] Satoshi Toh,[4] Seiko Harata,[4] Toshiaki Nakajima,[1] and Ituro Inoue[1]

[1]Division of Genetic Diagnosis, The Institute of Medical Science, University of Tokyo, [2]Department of Orthopedic Surgery, Institute of Rheumatology, Tokyo Women's Medical University, and [3]Laboratory of Bone and Joint Disease, SNP Research Center, The Institute of Physical and Chemical Research (RIKEN), Tokyo; Departments of [4]Orthopaedic Surgery and [5]Pharmacology, School of Medicine, Hirosaki University, Hirosaki, Japan; [6]Department of Orthopaedic Surgery, School of Medicine, Yamaguchi University, Ube, Japan; [7]Cardiovascular Genetics, University of Utah, Salt Lake City; and [8]Department of Cell Regulation, Institute for Molecular and Cellular Regulation, Gunma University, Maebashi, Japan

Ossification of the posterior longitudinal ligament (OPLL) of the spine is a subset of "bone-forming" diseases, characterized by ectopic ossification in the spinal ligaments. OPLL is a common disorder among elderly populations in eastern Asia and is the leading cause of spinal myelopathy in Japan. We performed a genomewide linkage study with 142 affected sib pairs, to identify genetic loci related to OPLL. In multipoint linkage analysis using GENEHUNTER-PLUS, evidence of linkage to OPLL was detected on chromosomes 1p, 6p, 11q, 14q, 16q, and 21q. The best evidence of linkage was detected near D21S1903 on chromosome 21q22.3 (maximum $Z_{lr} = 3.97$); therefore, the linkage region was extensively investigated for linkage disequilibrium with single-nucleotide polymorphisms (SNPs) covering 20 Mb. One hundred fifty positional candidate genes lie in the region, and 600 gene-based SNPs were genotyped. There were positive allelic associations with seven genes ($P < .01$) in 280 patients and 210 controls, and four of the seven genes were clustered within a region of 750 kb, ~1.2 Mb telomeric to D21S1903. Extensive linkage disequilibrium and association studies of the four genes indicated that SNPs in the collagen 6A1 gene (*COL6A1*) were strongly associated with OPLL ($P = .000003$ for the SNP in intron 32 [−29]). Haplotype analysis with three SNPs in *COL6A1* gave a single-point *P* value of .0000007. Identification of the locus of susceptibility to OPLL by genomewide linkage and linkage disequilibrium studies permits us to investigate the pathogenesis of the disease, which may lead to the development of novel therapeutic tools.

Introduction

Ossification of the posterior longitudinal ligament (OPLL [MIM 602475]) of the spine was first reported in Japan and is a common disorder among Japanese and other Asian populations (Matsunaga and Sakou 1997). The incidence of OPLL in Japan is 2%–4% of the general population >30 years of age, with a male predominance of 2:1 (Matsunaga and Sakou 1997). Heterotopic ossification of the spinal ligament is the specific feature of OPLL that causes compression of the spinal cord and leads to various degrees of myelopathy. Typical symptoms of OPLL are sensory and motor disturbance of the upper and lower extremities, abnormal reflexes, hyperresponsive deep reflexes, and bladder-bowel dysfunction. Various degrees of dysfunction, such as precise action and gait disturbance, lead to the restriction of activities involved in daily living and the deterioration of quality of life. Multiple etiologies for OPLL need to be considered because of the late-onset nature of the disease; however, the disease is, to some extent, genetically determined, as is demonstrated by the classic epidemiologic study and by the estimated relative risk of 10 for siblings of affected individuals (Terayama 1989; Sakou et al. 1991). OPLL is a high-bone-mass disease (or systemic hyperostosis), and patients with OPLL, especially women >60 years of age, have increased systemic bone mineral density (Yamauchi et al. 1999). Another ossification disorder, diffuse idiopathic skeletal hyperostosis (DISH), appears to be related to OPLL (Trojan et al. 1992). DISH is a common skeletal disease in middle-aged and elderly patients (with frequencies in individuals >50 years of age of 25% in men and 15% in women) in Western

Received May 12, 2003; accepted for publication July 17, 2003; electronically published September 4, 2003.

Address for correspondence and reprints: Dr. Ituro Inoue, Division of Genetic Diagnosis, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639 Japan. E-mail: ituro@ims.u-tokyo.ac.jp

* These authors contributed equally to this work.

0002-9297/2003/7304-0010$15.00

countries. The disease is characterized by ligamentous ossification of the anterolateral aspect of the spinal column, sometimes leading to bony ankylosis (Resnick et al. 1978). The extent of overlap of OPLL with DISH is uncertain; however, the pathophysiological mechanism of DISH may be similar to that of OPLL (Trojan et al. 1992; Weinfeld et al. 1997).

Because genetic factors appear to have a crucial role in OPLL, the use of molecular genetic studies is important to the understanding of the molecular etiologies of OPLL and will lead to the development of new therapeutics. Elsewhere, we reported linkage evidence in the HLA region of chromosome 6 in 91 affected sib pairs and identified the collagen 11A2 gene (*COL11A2* [MIM 120290]) as a possible candidate (Koga et al. 1998). Allelic association studies between OPLL and molecular variants in *COL11A2* demonstrated that a nucleotide substitution at intron 6 (−4), a T→A substitution, is significantly associated with OPLL, and the functional variant results in altered splicing, which is protective in the pathogenesis of OPLL (Maeda et al. 2001). Nakamura et al. (1999) reported that the nucleotide pyrophosphatase gene (*NPPS* [MIM 173335]), which has been identified as causal in a mouse model of OPLL (Okawa et al. 1998), is also associated with OPLL.

To understand the whole picture of OPLL susceptibility, we conducted a genomewide linkage study in 70 Japanese nuclear families, comprising 169 subjects and 142 affected sib pairs. At least six potential loci were linked to OPLL, with evidence for linkage being particularly strong on chromosome 21. Linkage analysis alone cannot provide the necessary resolution to identify the underlying genee, especially in complex diseases. Fine mapping of the linked regions can be attempted using linkage disequilibrium analysis with SNPs. Because a recent study of asthma susceptibility found that the 1-LOD decrease in the support interval of the linkage likely contains the susceptibility locus (Van Eerdewegh et al. 2002), the interval of the 1-LOD decrease on chromosome 21, ~20 cM, was extensively studied by use of linkage disequilibrium mapping to pinpoint the locus of OPLL. When 600 SNPs of 150 genes were screened, significant associations were revealed between OPLL and molecular variants in collagen 6A1 (*COL6A1* [MIM 120220]). The disease-related linkage disequilibrium was also assessed by haplotype-based association study.

Subjects and Methods

Disease Criteria and Subjects

The ethics committees of Hirosaki University and Kagoshima University approved the study, and all patients gave written informed consent. OPLL was diagnosed by the observation of ectopic bone formation in the posterior longitudinal ligament during x-ray or computed tomography examinations of the cervical, thoracic, or lumbar portions of the spine. The disease status of OPLL in radiograms was classified into four types, according to criteria of the Investigation Committee on the Ossification of the Spinal Ligaments, the Ministry of Health, Labor, and Welfare, Japan: (i) segmental, (ii) continuous, (iii) mixed, and (iv) localized, circumscribed, or bridged. All four types of OPLL were examined in the present study.

The siblings of patients with OPLL were screened for excessive spinal bone formation by x-ray examination, and affected siblings, including asymptomatic individuals, were recruited (Koga et al. 1998; Furushima et al. 2002). Samples for DNA analysis were collected from a total 142 affected sib pairs from 70 Japanese families; 98 pairs were from Kagoshima, 40 pairs from Hirosaki, 2 pairs from Asahikawa, and 1 pair each from Wakayama and Okinawa. The family structure was as follows: 51 affected pairs, 12 affected trios, 5 affected quartets, 1 affected quintet, and 1 affected sextet. The affected siblings included 76 women and 93 men. For the allelic association study, 342 unrelated patients with OPLL (73 familial cases, including 70 probands used for linkage study and 269 sporadic cases) and 298 unrelated subjects without OPLL were recruited in Kagoshima and Hirosaki. All the subjects without OPLL were >60 years of age and had no signs of spinal ossification when examined by standard x-ray methods, thereby excluding excluding most unmanifested diseases.

Microsatellite Genotyping

Multiplex fluorescent genotyping was performed using ABI PRISM linkage mapping set version 2 (Applied Biosystems). Because several markers were not polymorphic in the Japanese population (Ikari et al. 2001), a set of 47 markers obtained from the Genome Database were added to the original set to fill in the gap (Onda et al. 2001). An extra sequence was attached to the 5′ end of the reverse primer to promote nontemplated addition of adenine for accurate genotyping (Brownstein et al. 1996). Four markers (D6S446, D9S1678, D17S938, and DXS8067) generated incomplete data on the first round of genotyping and were excluded from further analysis. Marker positions (in Kosambi centimorgans) were obtained from the Marshfield Medical Research Foundation (Broman et al. 1998). For chromosome 21, which demonstrated significant linkage when the framework marker set was used, markers were added for dense mapping that covered <5 cM in the region (markers D21S1904, D21S1884, D21S1258, D21S262, D21S1900, D21S1893, D21S1411, and D21S1903). Microsatellite genotyping was performed as described

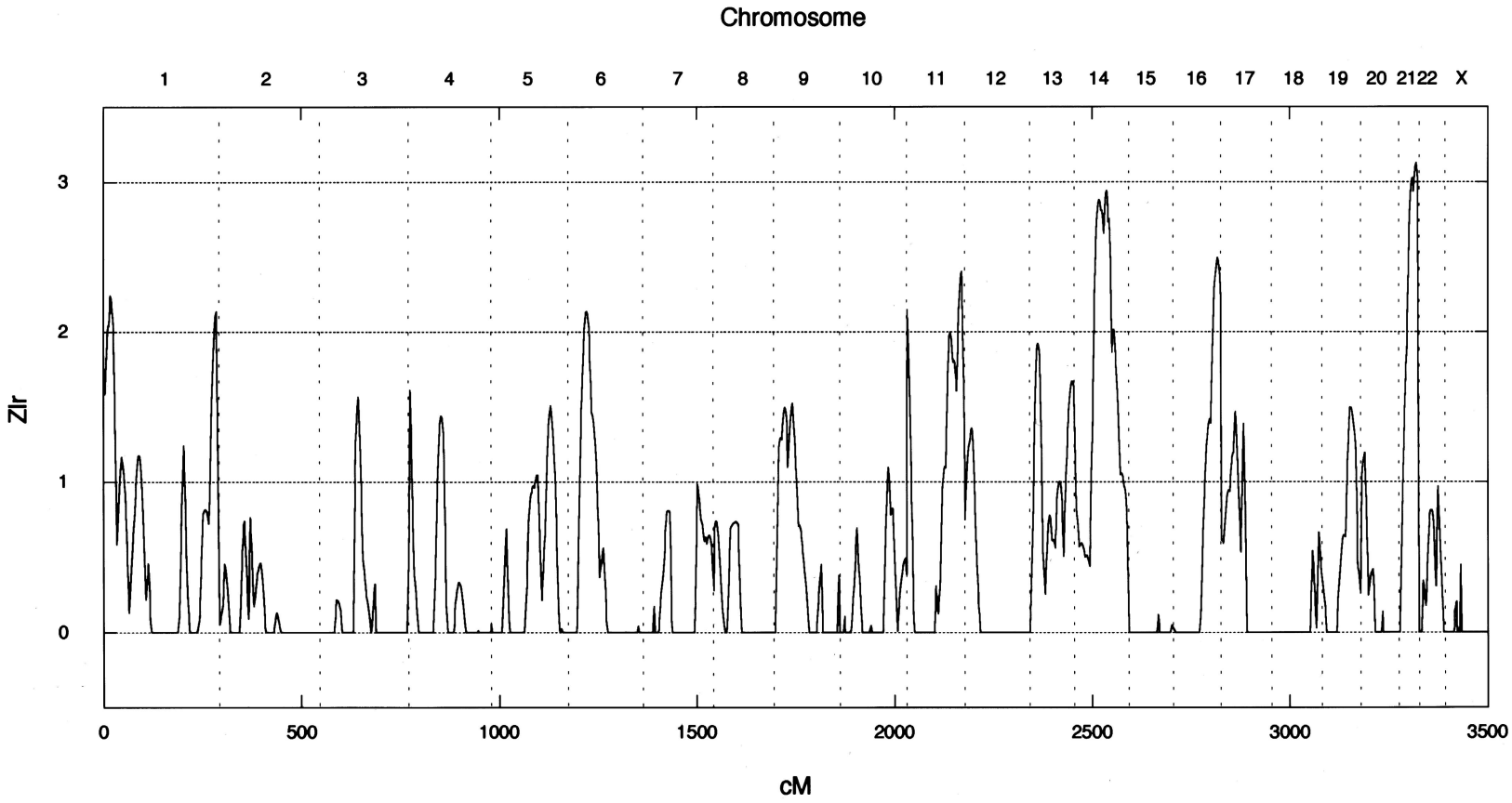

Figure 1 Genomewide linkage analysis of OPLL among 142 affected sib pairs. Summary of the framework genome scan for OPLL susceptibility loci showing the Z_{lr} scores that were calculated by GENEHUNTER-PLUS for all chromosomes. Dotted lines indicate chromosome boundaries, and the distance from the p terminus of chromosome 1 is shown on the *X*-axis.

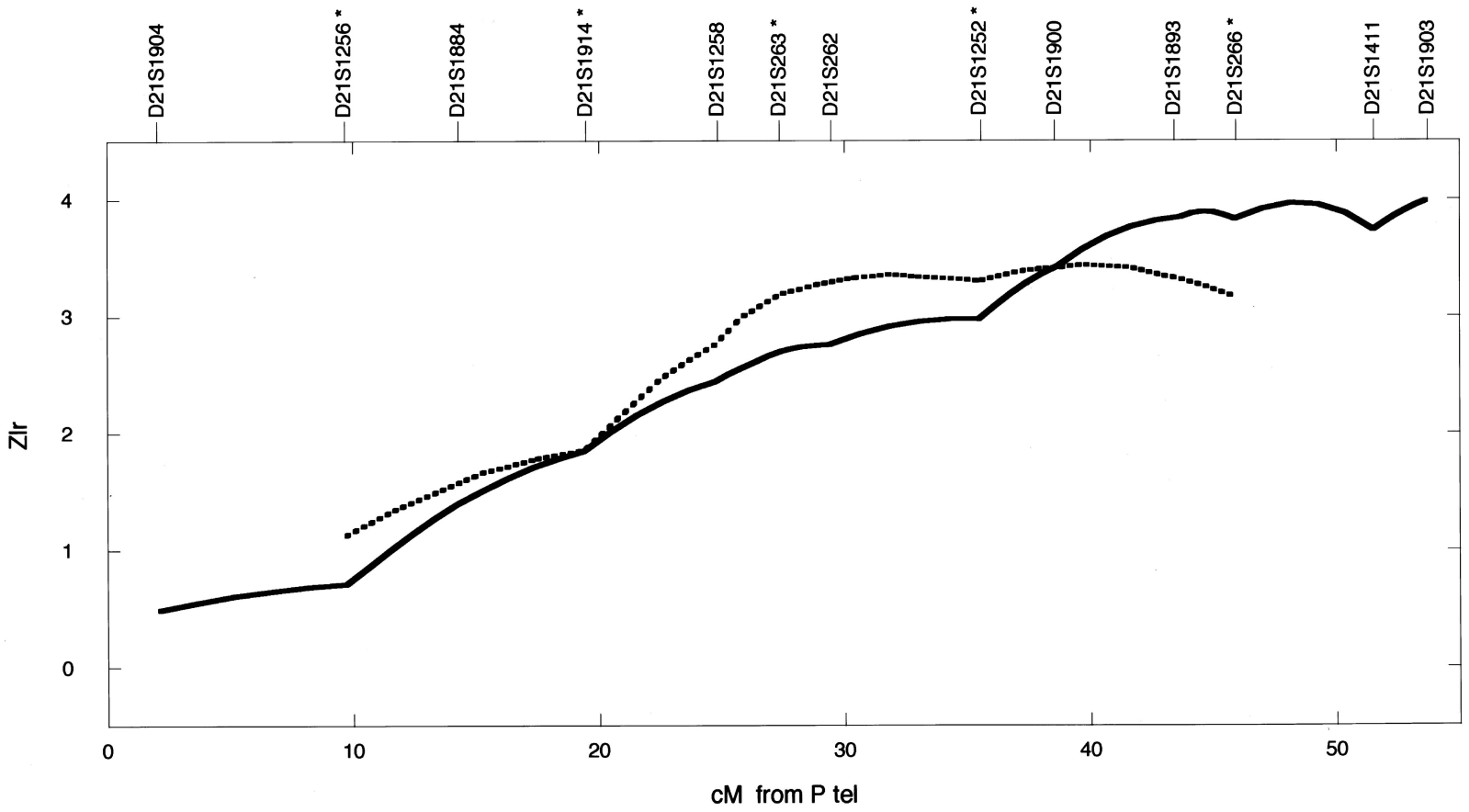

Figure 2 High-resolution mapping of chromosome 21. The solid line indicates the result of high-resolution linkage with 13 markers. The dotted line indicates the framework linkage result; asterisks (*) indicate the markers in the framework marker set. A Z_{lr} score of 3.97 is detected within the peak region close to the q terminus by fine mapping. The *Y*-axis indicates Z_{lr} score, and the *X*-axis indicates distance from the p terminus of chromosome 21.

Table 1

Genes Showing Positive Associations ($P < .01$) with OPLL on Chromosome 21

Gene and dbSNP ID	Allele Frequency in Subjects: With OPLL ($n = 280$)	Allele Frequency in Subjects: Without OPLL ($n = 210$)	χ^2 (1 df)	P
PRKCBP2:				
rs762178	.184	.123	6.73	.0095
DSCR1:				
JST022042	.041	.095	8.05	.0046
CBR1:				
rs20572	.220	.137	10.85	.0010
COL18A1:				
JST016803	.345	.454	11.31	.0007
JST016805	.329	.441	12.25	.0005
JST016806	.336	.442	10.64	.0011
PCBP3:				
rs2839036	.411	.312	12.75	.0004
rs2839037	.106	.043	15.16	.0001
rs395418	.239	.167	9.19	.0020
COL6A1:				
JST003751	.402	.310	11.23	.0008
JST016807	.124	.063	13.44	.0002
JST016808	.440	.329	16.27	.00005
JST106809	.439	.327	15.95	.00006
COL6A2:				
rs914246	.341	.261	8.18	.004

elsewhere (Furushima et al. 2002). All marker genotypes were checked for Mendelian inheritance, using the Checkfam program (Saito et al. 2002), and genotyping inconsistencies were corrected.

Affected-Sib-Pair Linkage Analysis

GENEHUNTER-PLUS (Kong and Cox 1997) and GENEHUNTER 2.1 (Kruglyak et al. 1996) were used to analyze the data on affected sib pairs. Multipoint analysis of the data from the genomewide scan and from the fine mapping was performed by weighting each family equally with GENEHUNTER-PLUS, a modified version of GENEHUNTER. GENEHUNTER-PLUS assumes a linear model for risk and thus provides more-accurate calculations of the variance than does the original program. GENEHUNTER 2.1 was used for estimates of the mean proportion of alleles shared identical by descent (IBD) and for exclusion mapping. The linkage exclusion was set at a LOD score of −2. GENEHUNTER 2.1 was also used to calculate information content. Allele frequencies of microsatellite markers were estimated in 64 unrelated Japanese subjects (Ikari et al. 2001). For dense mapping on chromosome 21, allele frequencies of the markers were estimated with 70 probands, to perform more stringent linkage analysis.

SNP Genotyping

SNPs were obtained from the two public databases: National Center for Biotechnology Information dbSNP and Institute of Medical Science–Japan Science and Technology Corporation database of Japanese SNPs. Gene-based SNPs were selected at 3- to 10-kb intervals to cover the gene. Twelve subjects with OPLL were genotyped for each SNP, to confirm frequencies of minor alleles, and SNPs with minor allele frequencies >0.08 were subjected to further analyses. SNPs in the coding region or regulatory region that may directly affect gene function were given priority, regardless of the minor allele frequencies. PCR analysis was performed with a standard protocol, except that a biotin-labeled primer was used when the pyrosequencing method was applied. SNPs were genotyped using either the pyrosequencing method (a real-time pyrophosphate DNA sequencing, on a PSQ96 Instrument [Pyrosequencing]) or direct sequencing (using BigDye terminator cycle sequencing on an ABI PRISM 3700 DNA analyzer [Applied Biosystems]).

Other Statistical Analyses

Allelic frequencies of polymorphisms in cases and controls were compared using a contingency χ^2 test. Haplotype frequencies for multiple loci were estimated using the expectation-maximization method with the Arlequin program, which is available at the Arlequin Web site or SNPAlyze program (Dynacom). Pairwise LD was estimated as $D = x_{11} - p_1 q_1$, where x_{11} is the frequency of haplotype A_1B_1, and p_1 and q_1 are the frequencies of alleles A_1 and B_1 at loci A and B, respectively. A standardized LD coefficient, r, is given by $D/(p_1 p_2 q_1 q_2)^{1/2}$, where p_2 and q_2 are the frequencies of the other alleles at loci A and B, respectively (Hill and Robertson 1968). Lewontin's coefficient, D', is given by $D' = D/D_{max}$, where $D_{max} = \min(p_1 q_2, p_2 q_1)$ when $D < 0$ or $D_{max} = \min(p_1 q_1, p_2 q_2)$ when $D > 0$ (Lewontin 1964).

Results

Affected-Sib-Pair Linkage Analysis

A total of 70 Japanese families comprising 169 affected subjects and 142 sib pairs were included in the genome-scan data set. Multipoint Z_{lr} scores for all chromosomes (except the Y chromosome) are displayed in figure 1. The region with the most prominent evidence of linkage was on chromosome 21q22, with a maximum Z_{lr} score of 3.09 near marker D21S266. In addition to chromosome 21, five others—on chromosomes 1p, 6p, 11q, 14q, and 16q—showed evidence of linkage (defined as $Z_{lr} > 2.2$). Because the best evidence of linkage was observed with markers on chromosome 21q, high-resolution mapping was performed by adding eight mark-

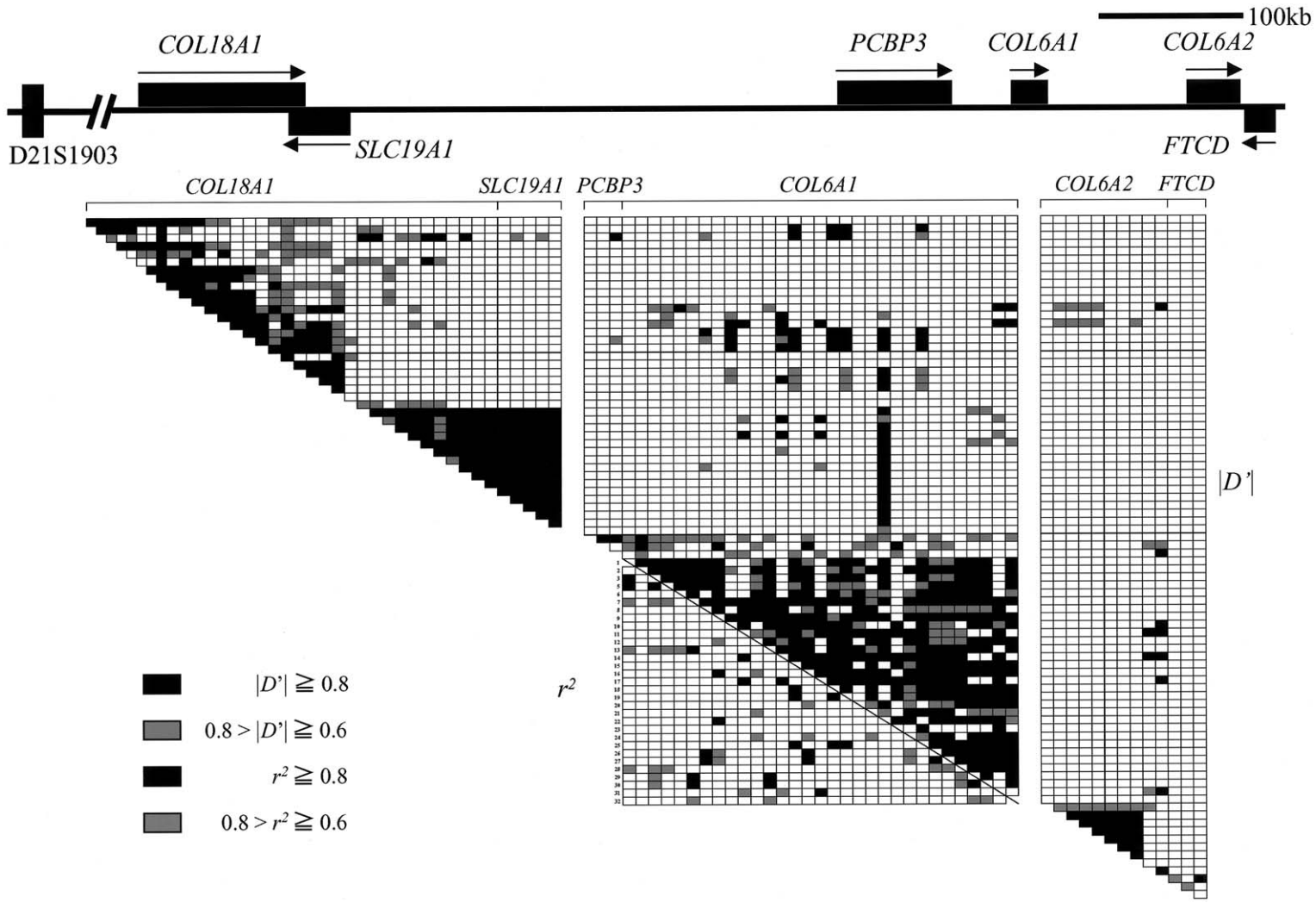

Figure 3 Pairwise linkage disequilibrium analysis on the 750-kb region where candidate genes are clustered. The upper figure shows six genes, including four associated genes (shown in table 1) within 750 kb of the linkage region near the marker D21S1903. The lower figure shows pairwise linkage disequilibrium blocks in the region. Pairwise linkage disequilibrium between all pairs was evaluated by the D' statistic, and *COL6A1* linkage disequilibrium was analyzed with the r^2 statistic. SNPs with allele frequencies <.05 were excluded for the linkage disequilibrium estimate.

Table 2

Polymorphisms in *COL6A1* and Their Association with OPLL

Number	Location[a]	NCBI dbSNP ID	Nucleotide Substitution	Amino Acid Change	Allele Frequency in Subjects (No. of Subjects)		χ^2 (df = 1)	*P*
					With OPLL	Without OPLL		
1	Exon1 (−5)	rs7671	G→C	…	.323 (314)	.262 (284)	5.33	.020
2	Intron 2 (+651)	rs2072699	G→A	…	.217 (92)	.215 (86)	.003	.958
3	Intron 2 (+758)	Non-dbSNP	T→C	…	.352 (91)	.238 (86)	5.44	.019
4	Intron 3 (+1326)	rs760437	C→T	…	.049 (92)	.068 (88)	.61	.435
5	Intron 4 (+13)	rs754507	A→C	…	.320 (86)	.227 (77)	3.48	.062
6	Intron 4 (+20)	Non-dbSNP	G→C	…	.176 (85)	.132 (76)	1.23	.267
7	Intron 4 (+37)	Non-dbSNP	A→G	…	.203 (64)	.089 (62)	6.58	.010
8	Intron 8 (−208)[b]	rs2839076	G→C	…	.123 (312)	.087 (281)	4.08	.043
9	Intron 9 (+62)	Non-dbSNP	T→C	…	.423 (300)	.346 (272)	7.27	.007
10	Intron 9 (+188)	rs2277813	C→G	…	.174 (290)	.133 (267)	3.61	.057
11	Intron 9 (−19)[b]	rs2277814	G→A	…	.051 (89)	.084 (89)	1.61	.205
12	Exon 15 (+39)	rs1980982	T→C	Gly/Gly	.402 (331)	.310 (287)	11.23	.0008
13	Intron 15 (+39)	Non-dbSNP	C→T	…	.264 (331)	.200 (285)	7.06	.008
14	Intron 18 (−138)[b]	rs760439	G→A	…	.031 (81)	.085 (59)	3.91	.048
15	Intron 19 (+27)	rs2850173	C→A	…	.137 (84)	.072 (83)	3.72	.053
16	Intron 19 (+116)	rs2075893	T→C	…	.455 (77)	.327 (78)	5.30	.021
17	Intron 19 (+361)	rs2742071	T→C	…	.043 (92)	.074 (88)	1.51	.219
18	Intron 19 (−152)[b]	rs2850174	T→G	…	.132 (87)	.091 (77)	1.39	.238
19	Intron 20 (+745)	rs2850175	A→C	…	.133 (338)	.077 (297)	10.26	.001
20	Intron 20 (−32)[b]	rs2839077	C→T	…	.417 (333)	.340 (296)	8.06	.005
21	Intron 21 (+18)	rs2276254	A→C	…	.340 (316)	.236 (277)	15.37	.00009
22	Intron 21 (−36)[b]	rs2276255	A→G	…	.076 (79)	.026 (76)	3.90	.048
23	Intron 22 (+19)	rs2276256	G→C	…	.234 (92)	.264 (89)	.45	.504
24	Intron 32 (−29)[b]	Non-dbSNP	T→C	…	.335 (336)	.216 (298)	21.99	.000003
25	Intron 33 (+15)	rs2236485	G→A	…	.124 (342)	.063 (292)	13.44	.0002
26	Intron 33 (+20)	rs2236486	A→G	…	.440 (342)	.329 (293)	16.27	.00005
27	Intron 33 (+55)	rs2236487	A→G	…	.439 (330)	.327 (278)	15.95	.00006
28	Intron 33 (+88)	rs2236488	C→T	…	.311 (323)	.252 (270)	5.08	.02
29	Exon 35 (+85)	rs1053312	G→A	Arg/His	.248 (322)	.199 (271)	4.97	.044
30	Exon 35 (+203)	rs1053315	G→A	Ala/Ala	.249 (321)	.199 (269)	4.24	.040
31	Exon 35 (+205)	Non-dbSNP	C→T	Ser/Leu	.059 (321)	.065 (269)	.17	.677
32	Exon 35 (+332)	rs1053320	C→T	Ser/Ser	.235 (321)	.180 (269)	5.32	.021

[a] Numbers in parentheses indicate the nucleotide position from the start of exon or intron with two exceptions: exon 1 (−5), in which the number refers to its position from the start codon, and numbers defined in footnote b.

[b] Numbers in parentheses indicate the distance from the last intronic nucleotide.

ers. The fine mapping on chromosome 21 (fig. 2) indicated that the linkage peak ($Z_{lr} = 3.97$) was close to the q terminus near marker D21S1903. The highest two-point Z_{lr} score (3.69) was detected at D21S1903 on the q terminus, and the highest IBD sharing of 71% ($Z_0 = 0.04$; $Z_1 = 0.50$; $Z_2 = 0.46$) was observed at D21S262 on 21q22.

Screening of Genes on the Linkage Region of Chromosome 21

The best evidence of linkage was detected with markers on chromosome 21q21; however, the linkage region ($Z_{lr} > 2.2$) spans ~30 cM, and the 1-LOD decrease in the support interval was ~20 cM and contained 200 genes. To pinpoint the locus for OPLL, linkage disequilibrium analysis was performed using gene-based SNPs in the region. We first genotyped 600 SNPs of 150 genes in the linkage region from 96 selected patients and 96 controls. The selected patients consisted of 73 unrelated patients with family history and 23 patients with age at onset <45 years. The remaining 50 genes were not investigated, because there were no SNPs in the database or SNPs were not polymorphic in Japanese populations. Allelic association was assessed by χ^2 analysis with a contingency table. Seventy-four SNPs of 24 genes exhibited significant allelic associations ($P < .05$; data not shown). For the second screening, we genotyped the 74 SNPs in an expanded number of patients with OPLL ($n = 280$) and in control subjects without OPLL ($n = 210$). These patients included the initial screening sets. Fourteen SNPs of seven genes had allelic associations ($P < .01$) with OPLL (table 1). The most significant association with OPLL was observed with SNPs of *COL6A1,* which is located 1.2 Mb from the peak linkage marker, D21S1903.

Table 3

Estimated Haplotype Frequency in Subjects with and without OPLL

	Frequency in Subjects				
Haplotype[a]	With OPLL (n = 306)	Without OPLL (n = 270)	χ^2 (df = 1)	P	Odds Ratio (95% CI)
H1 (+/+/+)	.548	.647	11.64	.0006	.66 (.59–.84)
H2 (−/−/−)	.297	.172	24.73	.0000007	2.03 (1.76–2.70)
H3 (+/+/−)	.067	.070	.05	.82	.95 (.75–1.50)
H4 (+/−/−)	.046	.045	.01	.92	1.03 (.78–1.80)
Others[b]	.042	.066	2.69	.101	.61 (.48–1.04)

[a] + = major allele; − = minor allele; the marker order, from left to right, is as follows: intron 21 (+18), intron 32 (−29), and intron 33 (+20).
[b] Haplotypes with frequencies <3%.

Because the four genes (*COL18A1* [MIM 120328], *PCBP3, COL6A1,* and *COL6A2* [MIM 120240]) that showed positive associations with OPLL are clustered within a 750-kb region in the vicinity of marker D21S1903, extensive SNP and pairwise linkage disequilibrium analyses were performed in the region. A total of 92 SNPs were analyzed for linkage disequilibrium structures in the region (fig. 3). Highly structured linkage disequilibrium blocks, as estimated with the D' statistic, were observed; for example, *COL18A1* was separated in two blocks, and the second block included *SLC19A1*. Because we genotyped gene-specific SNPs rather than SNPs in intergenic regions, we could not detect a broad scale of linkage disequilibrium blocks. The structures of pairwise linkage disequilibrium blocks were almost identical between patients and controls. SNPs of *COL6A1* that had the strongest association with OPLL were in a gene-specific linkage disequilibrium block; therefore, *COL6A1* was examined in more detail as the most likely candidate gene for OPLL. A total of 32 distinct SNPs—including the 4 SNPs used for the first screening, 25 additional dbSNPs, and 7 non-dbSNPs identified by direct sequencing—were genotyped in all the patients (n = 342) and controls (n = 298) (table 2). Among 7 SNPs identified in the present study, SNP31 at exon 35 led to a nonsynonymous change (serine to leucine), and other changes occurred in introns (table 2). Of 32 SNPs, 21 were statistically associated with OPLL ($P < .05$); of these, 4 had a particularly strong association ($P < .0001$ with OPLL; table 2). The most significant allelic association with OPLL was observed with a T→C substitution at intron 32 (−29) (χ^2 = 21.99; df = 1; P = .000003; odds ratio = 1.82 [95% CI = 1.60–2.35]). A T→C substitution at exon 15 (+39), which is a silent substitution, had the most significant association with OPLL among six coding SNPs (χ^2 = 11.23; df = 1; P = .0008; odds ratio = 1.49 [95% CI = 1.33–1.89]). For an explanation of numbers in parentheses that follow indications of location, see footnotes to table 2.

Haplotype Analysis of COL6A1

Because multiple SNPs may act in combination to increase the risk of OPLL, haplotypes were constructed on the basis of maximum likelihood. Three SNPs in *COL6A1* (intron 21 [+18], intron 32 [−29], and intron 33 [+20]) that had highly significant associations with OPLL and were not in complete linkage disequilibrium with each other, as estimated using the r measure ($r^2 < 0.7$; fig. 4), were selected for haplotype analysis (table 3). Estimated haplotype frequencies of the four common haplotypes (H1–H4) of subjects with OPLL were compared with those of control subjects without OPLL. There was a significant difference in haplotype frequency between cases and controls: haplotype H2, which comprises all the minor alleles at each site, occurred at a higher frequency in patients with OPLL than in control subjects (χ^2 = 24.73, df = 1, P = .0000007, odds ratio = 2.03 [95% CI = 1.76–2.70]); and haplotype H1, which comprises all the common alleles at each site, had a higher frequency in subjects without OPLL than in patients with OPLL (χ^2 = 11.64; df = 1; P = .0006; odds ratio = 0.66 [95% CI = 0.59–0.84]). The global test also gave a statistically significant result (χ^2 = 26.73; df = 4; P = .000023).

Discussion

The linkage results of the genomewide scan with 142 affected sib pairs revealed that six loci, on chromosomes 1p, 6p, 11q, 14q, 16q, and 21q, confer susceptibility to OPLL ($Z_{lr} > 2.2$), and the most promising locus was on chromosome 21q22 (Z_{lr} = 3.09) (fig. 1). Dense mapping of chromosome 21 revealed increased evidence of linkage on 21q22 near D21S1903 (Z_{lr} = 3.97) that would exceed the threshold of significant linkage (fig. 2) (Lander and Kruglyak 1995). The linkage region was studied extensively through linkage disequilibrium map-

ping and haplotype analysis, to narrow the susceptibility locus of OPLL.

Among five other susceptibility loci, chromosome 6 was of particular interest because previous studies reported that two candidate genes for OPLL (*COL11A2* and *NPPS*) are on that chromosome (Koga et al. 1998; Nakamura et al. 1999). In the multipoint linkage study of chromosome 6, the highest Z_{lr} score (2.22) and the best IBD sharing (59%; $Z_0 = 0.16$; $Z_1 = 0.50$; $Z_2 = 0.34$) was on 6p21, where *COL11A2* is located. On the other hand, the *NPPS* locus did not show any evidence of linkage ($Z_{lr} = 0.00$; mean IBD sharing 50.5%), confirming previous results of a single-point linkage analysis (Furushima et al. 2002).

The second most significant linkage ($Z_{lr} = 2.94$) was observed at chromosome 14, indicating that the chromosome 14q region may harbor another responsible gene. Notably, the linkage region at 14q is the widest linkage peak throughout the genome (43 cM; $Z_{lr} \geqslant 2.2$). Terwilliger et al. (1997) suggest that IBD sharing would be expected to decline gradually over adjacent markers when true linkage is present, resulting in a broad Z_{lr} score distribution. Because the best evidence of linkage was on chromosome 21, the 14q region was not pursued in the present study.

The genomewide linkage results indicated the best evidence of linkage to OPLL on chromosome 21; therefore, we proceeded to narrow the locus through a linkage disequilibrium study. We performed an association study with ~600 SNPs of 150 genes covering the 20-Mb linkage region of chromosome 21q. *COL6A1* is the gene most significantly associated with OPLL in the case-control comparison ($P = .04$–$.000003$) and the haplotype analysis ($P = .0000007$) (tables 2 and 3). In addition, pairwise linkage disequilibrium analysis revealed that *COL6A1* is in the gene-specific linkage disequilibrium block (fig. 3). Thus, susceptibility to OPLL could be accurately pinpointed to the *COL6A1* locus. Mutations in the genes that encode collagen VI subunits (*COL6A1, COL6A2,* and *COL6A3*) are causal of Bethlem myopathy (MIM 158810), a dominant form of myopathy, and Ullrich syndrome (MIM 254090), a recessive form of muscular dystrophy (Jobsis et al. 1996; Lamande et al. 1998; Pape et al. 1999; Camacho Vanegas et al. 2001; Scacheri et al. 2002) characterized by early childhood onset of proximal muscle weakness and contractures of multiple joints (Bethlem and Wijngaarden 1976; Mohire et al. 1988). In the present study, we did not detect any molecular variants in the key glycine residues of the triple-helix domain that result in Bethlem myopathy. *COL6A1* encodes the α_1 chain of type VI collagen, which is an extracellular matrix protein consisting of a short central triple helix flanked by two large globular domains. Because the SNP (the T→C substitution at intron 32 [−29]) that is most strongly associated with OPLL is near the branch site of the intron, that SNP may affect the lariat-shaped structure, thereby causing aberrant splicing. Our previous linkage and association studies identified *COL11A2* as a candidate for OPLL, and the functional intron 6 (−4A) polymorphism of *COL11A2* affects splicing of exon 6 (Koga et al. 1998; Maeda et al. 2001). RT-PCR analyses of *COL6A1* on cultured interspinous ligament cells from 10 patients with OPLL were performed, to study the splicing variants of *COL6A1,* as described in the analysis of *COL11A2* published elsewhere (Maeda et al. 2001). Thus far, we have not detected a splicing variant of *COL6A1* due to the intron 32 (−29) genotype (data not shown). The functional impact of the polymorphisms of *COL6A1* is uncertain, but *COL6A1* may lead to increased bone mass. Two genes (*COL6A1* and *COL11A2*) encode extracellular matrix proteins, which may provide a scaffold for osteoblastic or preosteoblastic cells or chondrocytes that subsequently proceed to membranous or endochondral ossification (Zhang and Chen 2000; Gowen et al. 2003). Therefore, molecular variants of the extracellular proteins may be implicated in the ectopic ossification observed in OPLL.

The etiologies of all complex diseases potentially involve gene-gene and gene-environment interactions. In the case of OPLL, interactions of *COL6A1* and *COL11A2* must be taken into account in efforts to further understand the etiology of the disease. Elucidation of the molecular etiology of OPLL will lead to the development of new therapeutic approaches, not only for bone-forming diseases but also for diseases of bone loss, such as osteoporosis.

Acknowledgments

We thank DNA donors for making this study possible, and we thank Dr. Koga (Imakiire Hospital) and Dr. Sakou (Sakou Clinic) for recruiting families for the study. We are grateful to Ms. Eguchi and Ms. Nakamura for their technical efforts. This work was supported, in part, by a Research for the Future Program grant from the Japan Society for the Promotion of Science (to I.I.), a Research Grant for Specific Diseases (to I.I., S.I., and S.T.) and the Japan Foundation of Aging and Health from the Ministry of Public Health and Welfare (to I.I. and S.T), and a grant-in-aid for scientific research from the Japanese Ministry of Education, Science, Sports and Culture (to I.I.).

Electronic-Database Information

URLs for data presented herein are as follows:

Arlequin, http://anthropologie.unige.ch/arlequin (for the Arlequin program)
Center for Medical Genetics, Marshfield Medical Research

Foundation http://research.marshfieldclinic.org/genetics/ (for genetic linkage maps)
GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for gene sequences)
Genome Database, The, http://www.gdb.org/ (for primer design of microsatellite genotyping)
Institute of Medical Science–Japan Science and Technology Corporation database of Japanese SNPs, http://snp.ims.u-tokyo.ac.jp/ (for dbSNP selection)
NCBI dbSNP Home Page, http://www.ncbi.nlm.nih.gov/SNP/ (for dbSNP selection)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for OPLL, Bethlem myopathy, Ullrich syndorome, *COL11A2, COL6A1, COL6A2,* and *NPPS*)
Primer 3, http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi (for designing PCR primers)
Pyrosequencing Technical Support, SNP Primer Design, http://techsupport.pyrosequencing.com/v2/index.asp (for designing sequencing primers for pyrosequencing)

References

Bethlem J, Wijngaarden GK (1976) Benign myopathy, with autosomal dominant inheritance: a report on three pedigrees. Brain 99:91–100

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:861–869

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: Primer modifications that facilitate genotyping. Biotechniques 20:1004–1006, 1008–1010

Camacho Vanegas O, Bertini E, Zhang RZ, Petrini S, Minosse C, Sabatelli P, Giusti B, Chu ML, Pepe G (2001) Ullrich scleroatonic muscular dystrophy is caused by recessive mutations in collagen type VI. Proc Natl Acad Sci USA 98: 7516–7521

Furushima K, Shimo-onoda K, Maeda S, Nobukuni T, Ikari K, Koga H, Komiya S, Nakajima T, Harata S, Inoue I (2002) Large-scale screening for candidate genes of ossification of the posterior longitudinal ligament of the spine. J Bone Miner Res 17:128–137

Gowen LC, Petersen DN, Mansolf AL, Qi H, Stock JL, Tkalcevic GT, Simmons HA, Crawford DT, Chidsey-Frink KL, Ke HZ, McNeish JD, Brown TA (2003) Targeted disruption of the osteoblast/osteocyte factor 45 gene (OF45) results in increased bone formation and bone mass. J Biol Chem 278: 1998–2007

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Ikari K, Onda H, Furushima K, Maeda S, Harata S, Takeda J (2001) Establishment of an optimized set of 406 microsatellite markers covering the whole genome for the Japanese population. J Hum Genet 46:207–210

Jobsis GJ, Keizers H, Vreijling JP, de Visser M, Speer MC, Wolterman RA, Baas F, Bolhuis PA (1996) Type VI collagen mutations in Bethlem myopathy, an autosomal dominant myopathy with contractures. Nat Genet 14:113–115

Koga H, Sakou T, Taketomi E, Hayashi K, Numasawa T, Harata S, Yone K, Matsunaga S, Otterud B, Inoue I, Leppert M (1998) Genetic mapping of ossification of the posterior longitudinal ligament of the spine. Am J Hum Genet 62:1460–1467

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Lamande SR, Bateman JF, Hutchison W, McKinlay Gardner RJ, Bower SP, Byrne E, Dahl HH (1998) Reduced collagen VI causes Bethlem myopathy: a heterozygous COL6A1 nonsense mutation results in mRNA decay and functional haploinsufficiency. Hum Mol Genet 7:981–989

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations: heterotic models. Genetics 49:49–67

Maeda S, Ishidou Y, Koga H, Taketomi E, Ikari K, Komiya S, Takeda J, Sakou T, Inoue I (2001) Functional impact of human collagen α2 (XI) gene polymorphism in pathogenesis of ossification of the posterior longitudinal ligament of the spine. J Bone Miner Res 16:948–957

Matsunaga S, Sakou T (1997) Epidemiology of ossification of the posterior longitudinal ligament. In: Yonenobu K, Sakou T, Ono K (eds) Ossification of the posterior longitudinal ligament. Springer, Tokyo, pp 11–17

Mohire MD, Tandan R, Fries TJ, Little BW, Pendlebury WW, Bradley WG (1988) Early-onset benign autosomal dominant limb-girdle myopathy with contractures (Bethlem myopathy). Neurology 38:573–580

Nakamura I, Ikegawa S, Okawa A, Okuda S, Koshizuka Y, Kawaguchi H, Nakamura K, Koyama T, Goto S, Toguchida J, Matsushita M, Ochi T, Takaoka K, Nakamura Y (1999) Association of the human NPPS gene with ossification of the posterior longitudinal ligament of the spine (OPLL). Hum Genet 104:492–497

Okawa A, Nakamura I, Goto S, Moriya H, Nakamura Y, Ikegawa S (1998) Mutation in Npps in a mouse model of ossification of the posterior longitudinal ligament of the spine. Nat Genet 19:271–273

Onda H, Kasuya H, Yoneyama T, Takakura K, Hori T, Takeda J, Nakajima T, Inoue I (2001) Genomewide-linkage and haplotype-association studies map intracranial aneurysm to chromosome 7q11. Am J Hum Genet 69:804–819

Pape G, Giusti B, Bertini E, Brunelli T, Saitta B, Comeglio P, Bolognese A, Merlini L, Federici G, Abbate R, Chu M-L (1999) A heterozygous splice site mutation in COL6A1 leading to an in-frame deletion of the α1 (VI) collagen chain in an Italian family affected by Bethlem myopathy. Biochem Biophys Res Commun 258:802–807

Resnick D, Guerra J, Robinson CA, Vinton CV (1978) Association of diffuse idiopathic skeletal hyperostosis (DISH) and calcification and ossification of the posterior longitudinal ligament. Am J Roentgenol 131:1049–1053

Saito M, Saito A, Kamatani N (2002) Web-based detection of genotype errors in pedigree data. J Hum Genet 47:377–379

Sakou T, Taketomi E, Matsunaga S, Yamaguchi M, Sonoda S, Yashiki S (1991) Genetic study of ossification of the posterior longitudinal ligament in the cervical spine with human leukocyte antigen haplotype. Spine 16:1249–1252

Scacheri PC, Gillanders EM, Subramony SH, Vedanarayanan V, Crowe CA, Thakore N, Bingler BS, Hoffman EP (2002) Novel mutations in collagen VI genes: expansion of the Bethlem myopathy phenotype. Neurology 58:593–602

Terayama K (1989) Genetic studies on ossification of the posterior longitudinal ligament of the spine. Spine 14:1184–1191

Terwilliger JD, Shannon WD, Lathrop GM, Nolan JP, Goldin LR, Chase GA, Weeks DE (1997) True and false positive peaks in genomewide scans: applications of length-biased sampling to linkage mapping. Am J Hum Genet 61:430–438

Trojan DA, Pouchot J, Pokrupa R, Ford RM, Adamsbaum C, Hill RO, Esdaile JM (1992) Diagnosis and treatment of ossification of the posterior longitudinal ligament of the spine: report of eight cases and literature review. Am J Med 92: 296–306

Van Eerdewegh P, Little RD, Dupuis J, Del Mastro RG, Falls K, Simon J, Torrey D et al. (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 418:426–430

Weinfeld RM, Olson PN, Maki DD, Griffiths HJ (1997) The prevalence of diffuse idiopathic skeletal hyperostosis (DISH) in two large American Midwest metropolitan hospital populations. Skeletal Radiol 26:222–225

Yamauchi T, Taketomi E, Matsunaga S, Sakou T (1999) Bone mineral density in patients with ossification of the posterior longitudinal ligament in the cervical spine. J Bone Miner Metab 17:296–300

Zhang Y, Chen Q (2000) Changes of matrilin forms during endochondral ossification: molecular basis of oligomeric assembly. J Biol Chem 275:32628–32634