Letters to the Editor

Am. J. Hum. Genet. 64:895, 1999

Psoriasis Linkage in the HLA Region

RICHARD O. LEDER AND SUSAN E. HODGE[1,2,3]
[1]*Division of Clinical-Genetic Epidemiology, New York State Psychiatric Institute,* [2]*Department of Psychiatry, Columbia University School of Physicians & Surgeons, and* [3]*Division of Biostatistics, Columbia University School of Public Health, New York*

To the Editor:

We were pleased to see the article by Jenisch et al. (1998), providing additional evidence for linkage between (familial) psoriasis and human leukocyte antigen (HLA). This confirms, in partially new data, what we (Leder et al. 1998) and others (Nair et al. 1997; Trembath et al. 1997) have already demonstrated—that there is a psoriasis susceptibility locus (PSORS1) on chromosome 6, closely linked to HLA. Unlike genome scans (Nair et al. 1997; Trembath et al. 1997), the analyses of Jenisch et al. (1998) and our own analyses show that support for linkage is enhanced by use of haplotypes to account for HLA associations in linkage analysis.

However, we note that Jenisch et al. (1998) still hedge on the finding of linkage. Readers should be aware that our analyses of previously published data (Leder et al. 1998) already support tight linkage beyond any reasonable doubt (LOD score of 23.7 under the assumption of dominant inheritance).

The assertion by Jenisch et al. (1998) that "Previous studies based on limited numbers of families found only weak evidence . . . for linkage to the HLA region" is also belied by our analysis of the previously published data. For example, the 31 families reported by Civatte et al. 1977 (cited in Jenisch et al. 1998) actually yield a LOD score of 6.29 between psoriasis and HLA-B, and family data from Sun et al. 1987 (not cited by Jenisch et al. 1998) yield a LOD score of 4.29. When the Civatte and Sun data are combined with data from other families reported previously (including all of those cited in Jenisch et al. 1998), evidence for linkage to psoriasis in the HLA region is overwhelming (summarized in table 3 of Leder et al. 1998). Finally, we suggest that all investigators of HLA and psoriasis should use the official nomenclature for this psoriasis-susceptibility locus, PSORS1 (Leder et al. 1998; White et al. 1998; see also White et al. 1999). Now that HLA linkage in psoriasis is conclusively settled, it remains both to identify the HLA-linked psoriasis gene(s) and to solidify evidence for the putative role(s) of non-HLA genes in the genetic etiology of psoriasis.

References

Civatte J, Lazarovici C, Ganas P, Leon S, Hors J, Contu L, Dausset J (1977) HLA system in psoriasis: study of 31 families. Ann Dermatol Venereol 104:525–532

Jenisch S, Henseler T, Nair R, Guo S, Westphal E, Stuart P, Krønke M, et al (1998) Linkage analysis of human leukocyte antigen (HLA) markers in familial psoriasis: strong disequilibrium effects provide evidence for a major determinant in the HLA-B/-C region. Am J Hum Genet 63:191–199

Leder RO, Mansbridge JN, Hallmayer J, Hodge SE (1998). Familial psoriasis and HLA-B: unambiguous support for linkage in 97 published families. Hum Hered 48:198–211

Nair R, Henseler T, Jenisch S, Stuart P, Bichakjian CK, Lenk W, Westphal E, et al (1997) Evidence for two psoriasis susceptibility loci (HLA and 17q) and two novel candidate regions (16q and 20p) by genome-wide scan. Hum Mol Genet 6:1349–1356

Sun Y, An J, Li S, Li H, Chen W, Lin L, Xu L, et al (1987) HLA antigen distribution and haplotype segregation in Chinese patients with psoriasis. Chin Med J 100:236–241

Trembath RC, Clough RL, Rosbotham JL, Jones AB, Camp RDR, Frodsham A, Browne J, et al (1997) Identification of a major susceptibility locus on chromosome 6p and evidence for further disease loci revealed by a two-stage genome-wide search in psoriasis. Hum Mol Genet 6:813–820

White J, Maltais L, Nebert D (1998) Networking nomenclature. Nat Genet 18:209

——— (1999) An increasingly urgent need for standardized gene nomenclature. Nat Genet, http://genetics.nature.com/web_specials/nomen/nomen_article.html

Address for correspondence and reprints: Richard O. Leder, 725 Los Robles Avenue, Palo Alto, CA 94306. E-mail: diwpri@leland.stanford.edu

0002-9297/99/6403-0027$02.00

Am. J. Hum. Genet. 64:896–897, 1999

Reply to Leder and Hodge

To the Editor:

On the basis of human leukocyte antigen (HLA) association studies, workers in the field of psoriasis have long been aware that the HLA complex plays an important role in determining psoriasis susceptibility. The question has always been why many families appear not to show linkage to HLA. We share in the pleasure of Drs. Leder and Hodge (1999 [in this issue]) now that the genetics of the HLA region in psoriasis is coming into sharper focus.

The general agreement between Leder and Hodge's studies (Leder et al. 1998), our own work (Jenisch et al. 1998 and in press), and the recent studies of Trembath et al. (1997) and Burden et al. (1998) provides welcome insight into this long-standing puzzle. By optimizing LOD scores over a variety of penetrance functions, assuming Hardy-Weinberg equilibrium, Leder and Hodge (Leder et al. 1998) found the highest LOD scores for dominant models specifying high disease allele frequency and low penetrance. We reached essentially the same conclusion, following the suggestions of Risch et al. (1989) for complex-trait data. It is well appreciated that power to detect linkage is diminished when the disease allele frequency is high and the penetrance is low. Given the smaller sample sizes of earlier studies, it is not surprising that linkage to the HLA region was not always apparent.

We have reported that linkage to HLA is more readily detected when marker-trait disequilibrium is taken into account, in part because of more-accurate specification of phase (Jenisch et al. 1998). Leder et al. (1998) and Trembath et al. (1997) report similar results. This effect was first pointed out 15 years ago (Clerget-Darpoux 1982) but has not been widely exploited in the genetic analysis of other common HLA-associated disorders. Even without incorporation of disease-marker haplotype frequencies, Leder et al. (1998) found strong evidence for linkage to HLA under a dominant model, whereas we did not. Leder et al.'s study made use of previously published pedigrees, and concerns regarding ascertainment bias in favor of linkage are inevitable in such a study. However, it is also possible that our sample yielded lower LOD scores because it contained a number of small pedigrees, thereby increasing the number of phase-unknown individuals.

We concur with Leder and Hodge (1999) that there is now excellent agreement regarding the importance of the HLA region in familial psoriasis and that this locus should now be referred to as PSORS1. We would emphasize that, because the HLA loci yielding the highest LOD scores in familial psoriasis are so similar to those observed in prior case-control association studies, there is unlikely to be any difference between familial and "sporadic" juvenile-onset psoriasis with respect to the involvement of PSORS1. We can also infer that genetic differences between juvenile- and adult-onset psoriasis must exist, because of their different HLA associations (Henseler and Christophers 1985). Whether an HLA locus different from PSORS1 is involved in the adult-onset form of this disease remains to be determined.

High disease allele frequencies and low penetrance values are likely to be the rule rather than the exception in common multifactorial diseases. We hope that these recent insights into the genetics of the HLA region in psoriasis will be of benefit to other groups studying complex genetic disorders.

James Elder,[1] Rajan Nair,[1] Philip Stuart,[1] John Voorhees,[1] Sun-Wei Guo,[2] Stefan Jenisch,[3] Eckhard Westphal,[3] Martin Krönke,[3] Tilo Henseler,[4] and Enno Christophers[4]

[1]*Department of Dermatology, University of Michigan, Ann Arbor;* [2]*Institute of Human Genetics and Division of Epidemiology,University of Minnesota, Minneapolis; and Departments of* [3]*Immunology and* [4]*Dermatology, University of Kiel, Kiel, Germany*

References

Burden AD, Javed S, Bailey M, Hodgins M, Connor M, Tillman D (1998) Genetics of psoriasis: paternal inheritance and a locus on chromosome 6p. J Invest Dermatol 110:958–960

Clerget-Darpoux F (1982) Bias of the estimated recombination fraction and LOD score due to an association between a disease gene and a marker gene. Ann Hum Genet 46: 363–372

Henseler T, Christophers E (1985) Psoriasis of early and late onset: characterization of two types of psoriasis vulgaris. J Am Acad Dermatol 13:450–456

Jenisch S, Henseler T, Nair RP, Guo S-W, Westphal E, Stuart P, Krönke M, et al (1998) Linkage analysis of HLA markers in familial psoriasis: strong disequilibrium effects provide evidence for a major determinant in the HLA-B/-C region. Am J Hum Genet 63:191–199

Jenisch S, Westphal E, Nair RP, Stuart P, Voorhees JJ, Christophers E, Krönke M, et al. Linkage disequilibrium analysis of familial psoriasis: identification of multiple disease-associated MHC haplotypes. Tissue Antigens (in press)

Leder RO, Hodge SE (1999) Psoriasis linkage in the HLA region. Am J Hum Genet 64:895–895 (in this issue)

Leder RO, Mansbridge JN, Hallmayer J, Hodge SE (1998) Familial psoriasis and HLA-B: unambiguous support for linkage in 97 published families. Hum Hered 48:198–211

Risch N, Claus E, Giuffra L (1989) Linkage and mode of inheritance in complex traits. Prog Clin Biol Res 329: 183–188

Trembath RC, Clough RL, Rosbotham JL, Jones AB, Camp

RDR, Frodsham A, Browne J, et al (1997) Identification of a major susceptibility locus on chromosome 6p and evidence for further disease loci revealed by a two stage genome-wide search in psoriasis. Hum Mol Genet 6:813–820

Address for correspondence and reprints: Dr. James Elder, Department of Dermatology, 3312 CCGC, Box 0932, University of Michigan, Ann Arbor, MI 48109-0932. E-mail: jelder@umich.edu

0002-9297/99/6403-0028$02.00

Am. J. Hum. Genet. 64:897–900, 1999

Protein-Truncation Mutations in the *RP2* Gene in a North American Cohort of Families with X-Linked Retinitis Pigmentosa

To the Editor:

X-linked forms of retinitis pigmentosa (XLRP) are a genetically heterogeneous group of retinal dystrophies that result in relatively severe clinical manifestations (Bird 1975; for a review, see Aldred et al. 1994). The two major XLRP loci, *RP2* (MIM 312600) and *RP3* (MIM 312610), have been mapped to Xp11.32-11.23 and Xp21.1, respectively (for a review see Aldred et al. 1994; Fujita et al. 1996; Fujita and Swaroop 1996; Thiselton et al. 1996). The *RP15* locus (MIM 300029) has been mapped to Xp22.13-22.11 in a single family with retinal degeneration (McGuire et al. 1995), and some evidence exists for a fourth locus, *RP6* (MIM 312612), at Xp21.3 (Musarella et al. 1990; Ott et al. 1990). We recently localized another genetic locus, *RP24* (MIM 300155), at Xq26-27 by using linkage analysis in an XLRP family (Gieser et al. 1998). In addition, the disease in some retinitis pigmentosa (RP) families with apparently X-linked inheritance does not seem to be linked to markers in the region of mapped XLRP loci (Teague et al. 1994; L. Gieser, R. Fujita, and A. Swaroop, unpublished data). It therefore appears that mutations in several genes on the X chromosome may lead to RP.

The first XLRP gene, *RPGR* (retinitis pigmentosa GTPase regulator), was isolated from the *RP3* region (Meindl et al. 1996; Roepman et al. 1996). Genetic analysis has suggested that *RP3* accounts for 70% of XLRP (Ott et al. 1990; Teague et al. 1994; Fujita et al. 1997). However, *RPGR* mutations are detected in only 20% of XLRP (and genetically defined *RP3*) families (Buraczynska et al. 1997; Fujita et al. 1997; M. Guevara-Fujita, S. Fahrner, and A. Swaroop, unpublished data). The *RP2* gene has recently been isolated by a positional cloning strategy (Schwahn et al. 1998) and is predicted to encode a protein of 350 amino acids with homology to cofactor C, which is involved in folding of β-tubulin (Tian et al. 1996). The *RP2* locus is believed to represent 20%–30% of XLRP in Europe (Ott et al. 1990; Teague et al. 1994), but little or no genetic evidence exists for an *RP2* subtype in the XLRP families from North America (Musarella et al. 1990; Ott et al. 1990). Because our haplotype analysis provided suggestive evidence for *RP2* in two North American families (R. Fujita, L. Gieser, S. G. Jacobson, P. A. Sieving, and A. Swaroop, unpublished data), we examined the genomic DNA from our cohort of XLRP patients for causative mutations in the *RP2* gene.

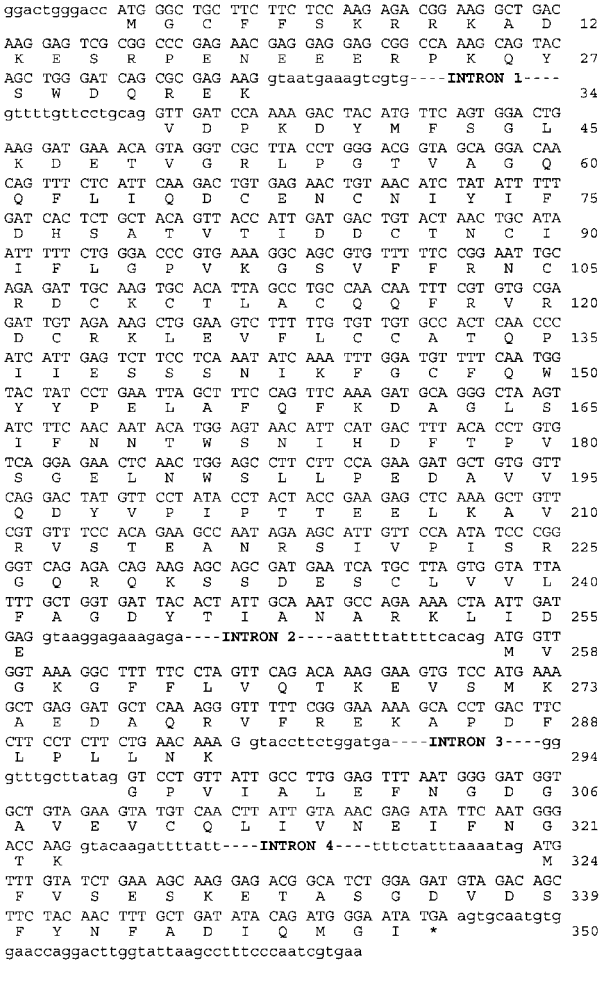

Figure 1 Composite nucleotide sequence showing *RP2* exons, including the coding region, and the exon-intron boundaries. The numbers on the right refer to the amino acid residues of the predicted RP2 protein.

The procedures for clinical ascertainment of patients, obtaining blood samples, and preparation of genomic DNA have been reported elsewhere (Fujita et al. 1997). The families included in the present study showed an apparent X-linked inheritance and no male-to-male transmission. Affected male individuals had a clinical

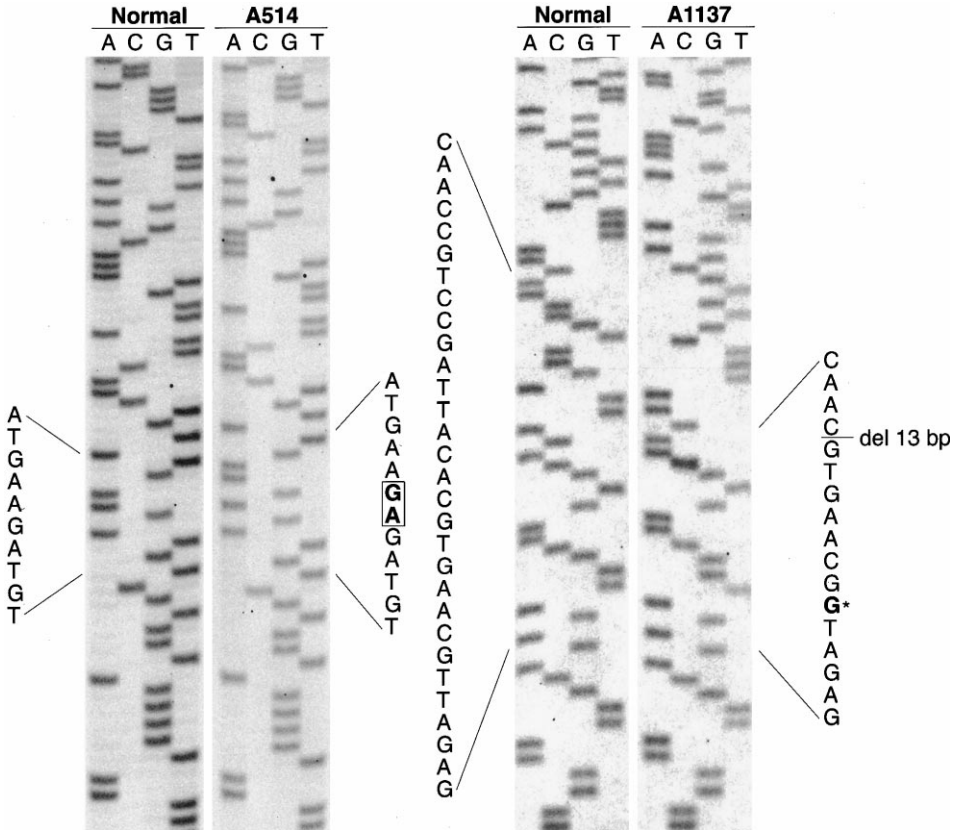

Figure 2 Representative sequencing gels showing two of the *RP2* mutations identified in this report. Sequences in the region of causative mutations are shown. The boxed sequence indicates the 2-bp insertion in patient A514. The location of the 13-bp deletion in patient A1137 is indicated by the horizontal bar. This patient also has a nucleotide substitution, indicated by an asterisk (*).

diagnosis of RP. Initially, one affected male each from 51 XLRP families was included in the *RP2* screening project. This cohort did not include families with a causative *RPGR* mutation or those in which the disease was genetically mapped to the *RP3* locus (see Buraczynska et al. 1997 and Fujita et al. 1997). Oligonucleotide primers flanking each of the five *RP2* exons (Schwahn et al. 1998) were used to amplify products from genomic DNA. PCR products were sequenced with various primers (Schwahn et al. 1998), either directly or after gel purification, by means of the ^{33}P-Thermosequenase cycle-sequencing kit (Amersham Life Science). The composite nucleotide sequence of the *RP2* exons and at the exon-intron boundaries is shown in figure 1. The derived sequence of RP2 polypeptide was identical to that reported elsewhere (Schwahn et al. 1998).

The complete sequencing of *RP2* exons and their corresponding exon-intron junction regions in 51 North American XLRP patients revealed sequence changes in five individuals (fig. 2 and table 1). All of the alterations were identified in the coding region: a 2-bp insertion in exon 1, a 13-bp deletion in exon 2, a nonsense mutation in exon 2, a 7-bp insertion in exon 2, and a 2-bp insertion in exon 4. Except for the C→T change at nucleotide 358 (arginine codon 120 in exon 2), resulting in a nonsense codon, the remaining four changes are deletions or insertions that would cause a frameshift. Therefore, all changes are predicted to result in a truncated RP2 protein. One of the patients (A1137) has an additional sequence alteration (T→G at nucleotide 322, leading to a Cys108Gly change); however, because this individual also has a 13-bp deletion nearby, we did not determine whether the T→G alteration may represent a disease-causing substitution. Each sequence change segregated in complete concordance with the disease in the respective family members that were available for the study (table 1). We suggest, on the basis of the nature of mutations and their cosegregation in respective families, that these sequence changes are causative *RP2* mutations.

This is the first report demonstrating the presence of the *RP2* subtype in North American families with XLRP. In addition to reporting five novel *RP2* mutations, our study addresses several significant issues:

1. The *RP2* mutations that we identified in our North

Table 1

***RP2* Mutations in Patients with X-Linked Retinitis Pigmentosa**

Patient Number	Exon	Nucleotide Sequence Change	Effect of Mutation	Meioses Examined
A2240	1	77/78insCA	Frameshift, 305 amino acids missing	8
A1137	2	T→G at 322 and del 330-342	Cys108Gly and a frameshift, 200 amino acids missing	1
A1135	2	C→T at 358	Arg120Stop, 230 amino acids missing	4
A512	2	483/484insGGGCTAA	Frameshift, 176 amino acids missing	2
A514	4	925/926insAG	Frameshift, 35 amino acids missing	3

NOTE.—Nucleotide positions are indicated according to the *RP2* coding sequence (National Center for Biotechnology Information accession number AJ007590; Schwahn et al. 1998).

American cohort of XLRP families are different from the seven reported in European families (Schwahn et al. 1998), suggesting a high rate of new mutations and a lack of founder effect. Similar observations have been made for *RPGR* mutations in XLRP-RP3 families (Buraczynska et al. 1997).

2. All five mutations reported here are predicted to result in a truncated RP2 protein. Except for Arg118His, the other six mutations identified by Schwahn et al. (1998) would also result in a shorter, or no, RP2 protein. We therefore suggest that the clinical phenotype in most if not all affected XLRP-RP2 families is due to the loss of RP2 function.

3. Our results suggest that it should be possible to identify a majority of *RP2* mutations in XLRP families by a protein-truncation test. Because RP2 protein is widely expressed, a relatively inexpensive diagnostic assay based on immunoblot analysis with RP2-specific antibody (when available) can also be developed. It should be noted that a protein-based diagnostic test has been established for choroideremia, another X-linked retinal dystrophy (MacDonald et al. 1998). Such a test, however, would be hard to develop for RPGR because of the diverse nature of mutations spanning a larger region of protein (Buraczynska et al. 1997) and multiple mRNA and protein isoforms (Yan et al. 1998).

4. Most of the mutations (Schwahn et al. 1998; present article) are detected in exon 2, which can be amplified as a 799-bp product. Additional mutations are present in two small exons—1 and 4. Of interest, no mutation has so far been detected in exon 3 or 5. This clustering of mutations might have significant implications for functional analysis of the RP2 protein and for prenatal and presymptomatic diagnosis.

5. Thus far it appears that screening of both *RPGR* and *RP2* genes leads to identification of disease-causing mutations in fewer than half of XLRP families. The five reported *RP2* mutations were identified by direct sequencing of coding region and exon-intron boundaries. Analysis of the *RP2* promoter region and/or the *RP2* genomic DNA by Southern blotting might reveal additional causative mutations.

Although much of the genetic and phenotypic complexities of XLRP have yet to be resolved, the cloning of *RPGR* and *RP2* genes represents a milestone in RP research. Identification of mutations in these two genes in many XLRP families provides renewed hope for more-precise diagnosis and better genetic counseling for this devastating disease.

Acknowledgments

We thank Drs. Sten Andreasson, David Birch, Nancy Carson, Bernie Chodirker, Mark Evans, Gerald Fishman, John Heckenlively, Dennis Hoffman, Maria Musarella, and Beth Spriggs and Mr. Eric L. Krivchenia for some of the patient samples that were included in the mutation screening. We acknowledge the assistance of Dr. Wolfgang Berger for providing the *RP2* primer sequences. We thank Dr. Monika Buraczynska for organization of the patient registry; Dr. Radha Ayyagari for discussions; Dr. Beverly Yashar for counseling; Ms. Cara Coats for assistance in patient collection; Mr. Jason Cook, Ms. Patricia Forsythe, and Ms. Eve Bingham for technical assistance; and Ms. D. Giebel for secretarial assistance. This research was supported by National Institutes of Health (NIH) grants EY05627, EY06094, and EY07961 and by grants from the Foundation Fighting Blindness, the Chatlos Foundation, the Kirby Foundation, the Mackall Trust, and Research to Prevent Blindness. We also acknowledge NIH grants EY07003 (core) and M01-RR00042 (General Clinical Research Center) and a shared equipment grant from the Office of Vice President for Research (University of Michigan). A.S. is recipient of a Lew R. Wasserman Merit Award, and P.A.S., a Senior Scientific Investigator Award, both from Research to Prevent Blindness.

ALAN J. MEARS,[1,*] LINN GIESER,[1,*] DENISE YAN,[1,*] CYNTHIA CHEN,[1,*] STACEY FAHRNER,[1] SUJA HIRIYANNA,[1] RICARDO FUJITA,[1,†] SAMUEL G. JACOBSON,[3] PAUL A. SIEVING,[1] AND ANAND SWAROOP[1,2]

Departments of [1]Ophthalmology and [2]Human Genetics, Kellogg Eye Center, University of Michigan, Ann Arbor, and [3]Department of Ophthalmology, Scheie Eye Institute, University of Pennsylvania, Philadelphia

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for *RP2* [MIM 312600], *RP3* [MIM 312610], *RP6* [MIM 312612], *RP15* [MIM 300029], and *RP24* [MIM 300155])

National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/(for RP2 sequence, accession number AJ007590)

References

Aldred MA, Jay M, Wright AF (1994) X-linked retinitis pigmentosa. In: Wright AF, Jay B (eds) Molecular genetics of inherited eye disorders. Harwood Academic Publishers, Chur, Switzerland, pp 259–276

Bird AC (1975) X-linked retinitis pigmentosa. Br J Ophthalmol 59:177–199

Buraczynska M, Wu W, Fujita R, Buraczynska K, Phelps E, Andreasson S, Bennett J, et al (1997) Spectrum of mutations in the *RPGR* gene that are identified in 20% of families with X-linked retinitis pigmentosa. Am J Hum Genet 61: 1287–1292

Fujita R, Bingham E, Forsythe P, Segal M, Aita V, Navia BA, Dry K, et al (1996) A recombination outside the BB deletion refines the location of the X-linked retinitis pigmentosa locus RP3. Am J Hum Genet 59:152–158

Fujita R, Buraczynska M, Gieser L, Wu W, Forsythe P, Abrahamson M, Jacobson SG, et al (1997) Analysis of the *RPGR* gene in 11 pedigrees with the retinitis pigmentosa type 3 genotype: paucity of mutations in the coding region but splice defects in two families. Am J Hum Genet 61:571–580

Fujita R, Swaroop A (1996) RPGR: part one of the X-linked retinitis pigmentosa story. Mol Vis 2:4

Gieser L, Fujita R, Goring HHH, Ott J, Hoffman DR, Cideciyan AV, Birch DG, et al (1998) A novel locus (*RP24*) for X-linked retinitis pigmentosa maps to Xq26-27. Am J Hum Genet 63:1439–1447

MacDonald IM, Mah DY, Ho YK, Lewis RA, Seabra MC (1998) A practical diagnostic test for choroideremia. Ophthalmology 105:1637–1640

McGuire RE, Sullivan LS, Blanton SH, Church MW, Heckenlively JR, Daiger SP (1995) X-linked dominant cone-rod degeneration: linkage mapping of a new locus for retinitis pigmentosa (RP15) to Xp22.13-p22.11. Am J Hum Genet 57:87–94

Meindl A, Dry K, Herrmann K, Manson F, Ciccodicola A, Edgar A, Carvalho MRS, et al (1996) A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). Nat Genet 13:35–42

Musarella MA, Anson-Cartwright L, Leal SM, Gilbert LD, Worton RG, Fishman GA, Ott J (1990) Multipoint linkage analysis and heterogeneity testing in 20 X-linked retinitis pigmentosa families. Genomics 8:286–296

Ott J, Bhattacharya SS, Chen JD, Denton MJ, Donald J, Dubay C, Farrar GJ, et al (1990) Localizing multiple X-chromosome–linked retinitis pigmentosa loci using multilocus homogeneity tests. Proc Natl Acad Sci USA 87:701–704

Roepman R, van Duijnhoven G, Rosenberg T, Pinckers AJLG, Bleeker-Wagemakers LM, Bergen AAB, Post J, et al (1996) Positional cloning of the gene for X-linked retinitis pigmentosa 3: homology with the guanine-nucleotide–exchange factor RCC1. Hum Mol Genet 5:1035–1041

Schwahn U, Lenzner S, Dong J, Feil S, Hinzmann B, van Duijnhoven G, Kirschner R, et al (1998) Positional cloning of the gene for X-linked retinitis pigmentosa 2. Nat Genet 19: 327–332

Teague PW, Aldred MA, Jay M, Dempster M, Harrison C, Carothers AD, Hardwick LJ, et al (1994) Heterogeneity analysis in 40 X-linked retinitis pigmentosa families. Am J Hum Genet 55:105–111

Thiselton DL, Hampson RM, Nayudu M, Maldergem LV, Wolf ML, Saha BK, Bhattacharya SS, et al (1996) Mapping the RP2 locus for X-linked retinitis pigmentosa on proximal Xp: a genetically defined 5-cM critical region and exclusion of candidate genes by physical mapping. Genome Res 6: 1093–1102

Tian G, Huang Y, Rommelaere H, Vandekerckhove J, Ampe C, Cowan NJ (1996) Pathways leading to correctly folded β-tubulin. Cell 86:287–296

Yan D, Swain PK, Breuer D, Tucker RM, Wu W, Fujita R, Rehemtulla A, et al (1998) Biochemical characterization and subcellular localization of the mouse retinitis pigmentosa GTPase regulator (mRpgr). J Biol Chem 273:19656–19663

Address for correspondence and reprints: Dr. Anand Swaroop, Kellogg Eye Center, University of Michigan, 1000 Wall Street, Ann Arbor, MI 48105. E-mail: swaroop@umich.edu

*Drs. Mears and Yan, Ms. Gieser, and Ms. Chen contributed equally to this work.

†Present affiliation: Facultad de Medicina Humana, Universidad San Martin de Porres, Lima, Peru.

0002-9297/99/6403-0029$02.00

Am. J. Hum. Genet. 64:900–904, 1999

A Fifth Locus for Bardet-Biedl Syndrome Maps to Chromosome 2q31

To the Editor:

Bardet-Biedl syndrome (BBS) is a rare autosomal recessive disorder with major clinical manifestations of retinal dystrophy, obesity, dysmorphic extremities, hypogenitalism, and renal structural and functional abnormalities. It is distinguished from Laurence-Moon syndrome (MIM 245800), Biemond syndrome II (MIM 210350), and Alstrom syndrome (MIM 203800) by the absence of paraplegia, iris coloboma, and perceptive deafness, respectively. Four genetic loci for BBS have been mapped to distinct chromosomes, but the finding, in three recent population surveys, of several unlinked families with

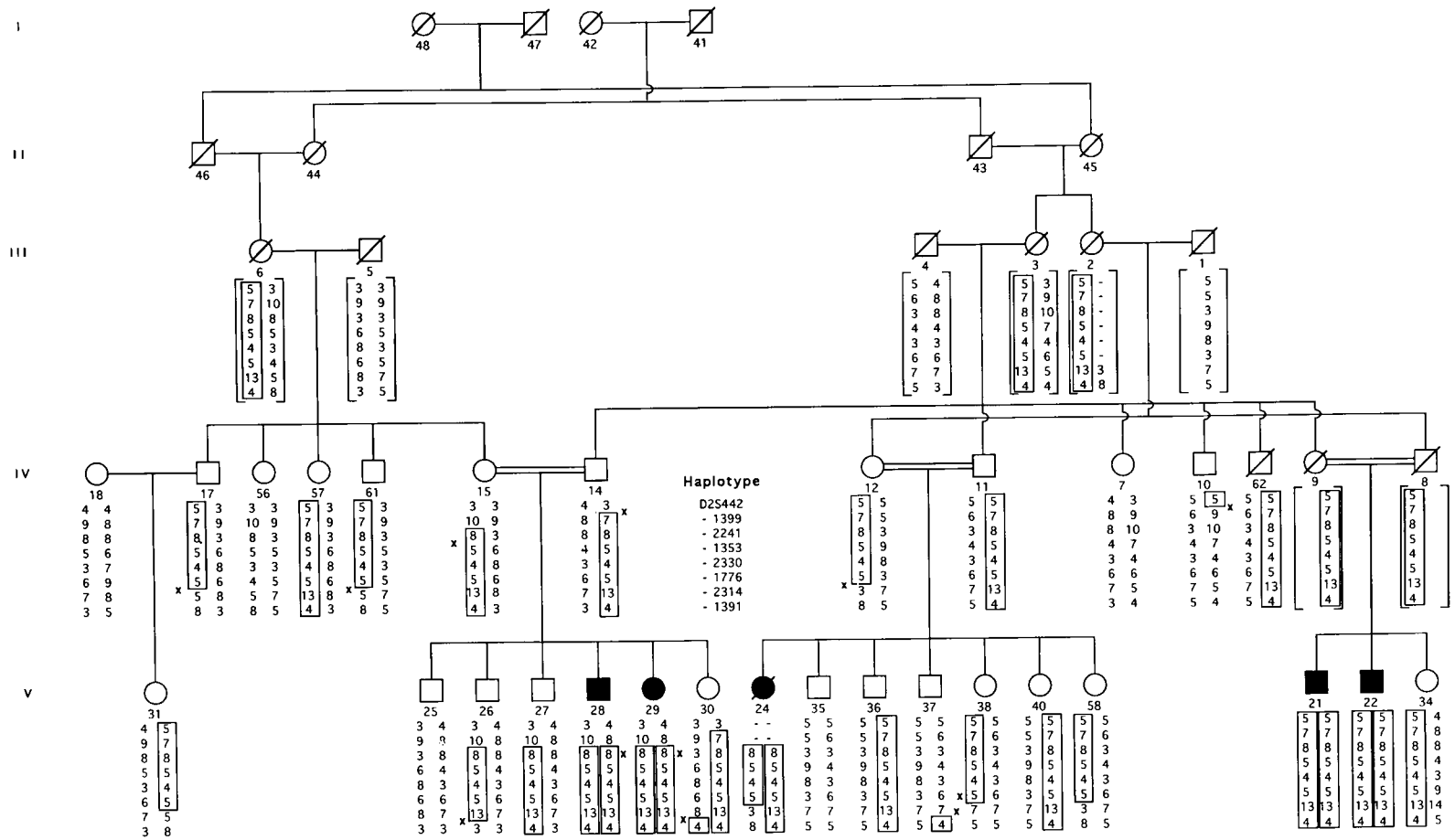

Figure 1 Cosegregation of BBS and an ancestral haplotype on chromosome 2q31 in kindred B9. Double marriage lines indicate consanguineous unions. Haplotypes were constructed manually and represent the minimal number of recombinations (x). The DNA sample from person 24 was extracted from a paraffin sample and often fails to amplify, but this person appears to be nonrecombinant for both parental chromosomes. The family members with BBS have inherited two copies of the ancestral haplotype (*boxed*) from a great-great-grandparent in generation I. The minimal region of homozygosity in relatives with BBS includes the markers pter-D2S124, D2S2330, D2S1776, and D2S335-qter.

Table 1

Clinical Manifestations of *BBS5* in a Newfoundland Kindred

Patient (Sex/Age)	Body-Mass Index[a] (wt [kg]/ht [m]^2)	Polydactyly/Other[b]	Visual Acuity	Retinal Appearance	Small Penis
21 (M /31 years)	31.7	Absent/present	Light perception only	Advanced retinitis pigmentosa	Present
22 (M /25 years)	34.6	Absent/present	No light perception	Retinal degeneration	Not examined
24 (F /29 years)	49.4	Absent/present	No light perception	Atypical retinitis pigmentosa	Not applicable
28 (M /25 years)	40.4	Absent/present	Ability to count fingers	Retinitis pigmentosa	Present
29 (F /21 years)	42.8	Absent/present	20/300	Macular dystrophy	Not applicable

[a] A value >27 is considered to indicate obesity (Nelson et al. 1994).
[b] Other = brachydactyly and/or syndactyly.

BBS provides convincing evidence for at least a fifth BBS locus (Beales et al. 1997; Bruford et al. 1997; Woods et al. 1999).

BBS is a relatively rare disease, with an estimated world prevalence of 1/125,000–160,000 (Klein and Ammann 1969; Beales et al. 1997). However, two genetically isolated and distinct populations have been identified that provide a resource of large inbred families with BBS. One in 13,500 individuals has BBS in the Bedouin-Arab tribes of the Negev region of Israel, where the custom of consanguineous marriages is still practiced by >50% of the population and where two-thirds of these marriages are between first cousins (Farag and Teebi 1989; Sheffield et al. 1998). Half a world away, on the island portion of Newfoundland, the prevalence of BBS is 1/17,500 (Green et al. 1989). Matings between distant cousins in the Newfoundland population are frequent because of three historical factors: the geographic isolation of coastal fishing villages, the low rate of immigration to these communities, and the religious restrictions on mate selection between the Protestant English and Catholic Irish settlers (Bear et al. 1988). Three of the four BBS loci—*BBS2* (Kwitek-Black et al. 1993), *BBS3* (Sheffield et al. 1994), and *BBS4* (Carmi et al. 1995)—were identified by homozygosity mapping in individual Bedouin families. We have used a similar methodology to map the fifth genetic locus for BBS to chromosome 2q31 in an inbred Newfoundland family of European ancestry.

A recently completed population-based survey of 17 BBS families from Newfoundland has identified six families in which the four known BBS loci were unambiguously excluded (Woods et al. 1999). Family B9, the largest of these kindreds, has five affected members who are the products of three consanguineous unions interrelated through two founding couples in generation I (fig. 1). The methods used in the clinic assessment of these patients have been described elsewhere (Green et al.1989). The five patients with BBS surpass the minimal criteria of three major clinical manifestations for a BBS diagnosis, because of the presence of obesity, brachydactyly and/or syndactyly, retinal dystrophy, and male hypogenitalism, in the absence of paralysis, iris coloboma, or deafness (table 1). This pedigree met the requirements for the localization of *BBS5* by homozygosity mapping (Lander and Botstein 1987; Carmi et al. 1995). We anticipated that the affected individuals would be homozygous by descent for an ancestral haplotype inherited from one of the four pedigree founders.

A genomewide scan of pooled DNA samples was performed with microsatellite markers (Cooperative Human Linkage Consortium human screening set, Weber version 8; Research Genetics), as described elsewhere (Sheffield et al. 1994). Two control pools of DNA from

Table 2

Two-Point Linkage Analysis, for BBS and 2q31 Markers

	Two-Point LOD Score at θ = [b]							Maximum
Marker[a]	.000	.010	.050	.100	.200	.300	Z_{max}	θ
D2S442	−3.080	−2.205	−1.224	−.699	−.211	−.042	.009	.439
D2S1399	−∞	−.416	0.982	1.321	1.174	.722	1.345	.122
D2S1353	5.675	5.556	5.075	4.463	3.214	1.970	5.675	.000
D2S1776	4.691	4.578	4.121	3.543	2.377	1.275	4.691	.000
D2S1391	−∞	1.617	2.558	2.600	2.066	1.298	2.631	.078

Note.—A disease model of autosomal recessive inheritance, 100% penetrance, and a gene frequency of .008 were invoked.
[a] Listed according to physical order (pter-qter) on chromosome 2q31.
[b] Calculated by MLINK and ILINK from the FASTLINK package (version 4.0P).

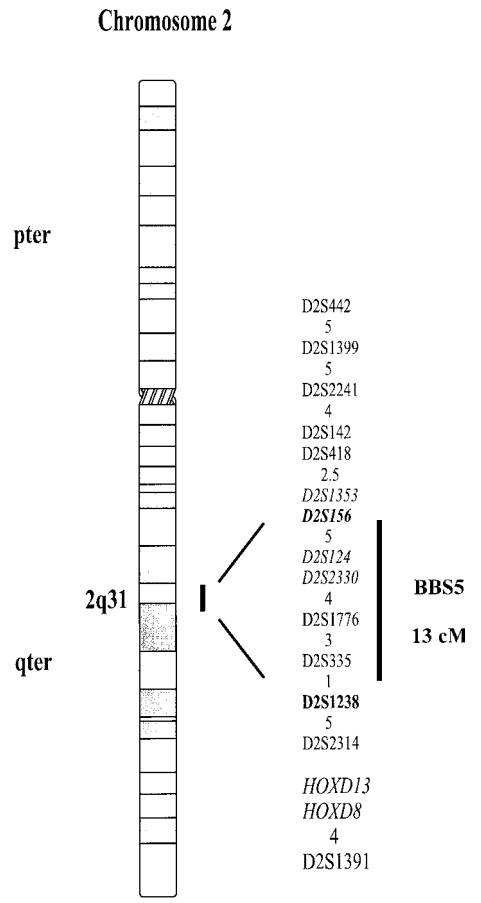

Figure 2 Critical region of *BBS5* on chromosome 2q31. Markers used to refine the position of BBS5 are identified in terms of their "D" numbers. The marker order and distances were based on the Marshfield chromosome 2 (sex-averaged) linkage map (Center for Medical Genetics, Marshfield Medical Research Foundation) and the Chromosome 2 Workshop Consensus Map, 1996 (Genome Database). The *HOXD*-gene cluster lies downstream of the *BBS5* critical region.

4 living parents and 11 unaffected siblings, as well as a test pool of DNA from the 4 surviving patients, were amplified. Of the first 322 markers successfully amplified, 6 showed a reduction in the number of alleles (allele shift) in the test pool, compared with the control pools. Subsequent genotyping of these markers on the extended family proved that they were not linked to BBS, resulting in a false-positive rate of 1.9%. However, the 323d marker, D2S1353, gave a 4:1 allele shift, from the control pools to the test pool. Genotyping of D2S1353 on the pedigree showed it to be exclusively homozygous in patients with BBS. Two-point analysis showed significant linkage between BBS and D2S1353, with no recombination (maximum LOD score [Z_{max}] 5.675; recombination fraction [θ] 0). Genotyping of markers flanking D2S1353 confirmed linkage to 2q31 (table 2) and showed an ancestral haplotype that is homozygous by descent in all affected relatives (fig. 1).

The initial assignment of the BBS phenotype in close proximity to the *HOXD*-gene cluster on chromosome 2q31 suggested that these nine homeobox genes of the *Drosophila antennapedia* class and other closely located genes (*EVX2* and *DLX1/DLX2*) that are involved in patterning of the embryo are candidate genes for *BBS5*. Recent findings that duplication of the *HOXD13* gene causes synpolydactyly (Akarsu et al. 1996) focused our attention on it as the most promising gene candidate, given that syndactyly and/or polydactyly are congenital manifestations of BBS. However, refined mapping of two key recombinant ancestral chromosomes in patients 15 and 12 placed *BBS5* within the 13-cM interval D2S156–D2S1238 (fig. 2), several centimorgans upstream from the *HOXD13* gene that is positioned at the proximal end of the *HOXD*-gene cluster (Spurr et al. 1996). Refined mapping of the recombinant ancestral chromosome excludes all genes within the *HOXD*-gene cluster as being candidate genes for *BBS5*.

Acknowledgments

We thank the family for their participation in this study. Financial support for this work came from the Kidney Foundation of Canada and the Medical Research Council of Canada. T.-L.Y. and M.O.W. are recipients of Graduate Fellowships from the Faculty of Medicine at Memorial University. J.S.G is a recipient of a grant from the Canadian Genetic Diseases Network.

Terry-Lynn Young,[1] Lynette Penney,[1] Michael O. Woods,[1] Patrick S. Parfrey,[1] Jane S. Green,[1] Donna Hefferton,[1] and William S. Davidson[2]

[1]*Faculty of Medicine and* [2]*Department of Biochemistry, Memorial University, St. John's, Newfoundland*

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

Cooperative Human Linkage Consortium, http://www.chlc.org/HomePage.html (for microsatellite markers)

Genome Database, http://gdbwww.gdb.org (for marker order and distance)

Center for Medical Genetics, Marshfield Medical Research Foundation, http://www.marshmed.org/genetics (for marker order and distance)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim (for Laurence-Moon syndrome [MIM 245800], Biemond syndrome II [MIM 210350], and Alstrom syndrome [MIM 203800])

References

Akarsu AN, Stoilov I, Yilmaz E, Sayli BS, Sarfarazi M (1996) Genomic structure of HOXD13 gene: a nine polyalanine duplication causes synpolydactyly in two unrelated families. Hum Mol Genet 5:945–952

Beales PL, Warner AM, Hitman GA, Thakker R, Flinter FA (1997) Bardet-Biedl syndrome: a molecular and phenotypic study of 18 families. J Med Genet 34:92–98

Bear JC, Nemec TF, Kennedy JC, Marshall WH, Power AA, Kolonel VM, Burke GB (1988) Inbreeding in outport Newfoundland. Am J Med Genet 29:649–660

Bruford EA, Riise R, Teague PW, Porter K, Thomson KL, Moore AT, Jay M, et al (1997) Linkage mapping in 29 Bardet-Biedl syndrome families confirms loci in chromosomal regions 11q13, 15q22.3–q23, and 16q21. Genomics 41: 93–99

Carmi R, Rokhlina T, Kwitek-Black AE, Elbedour K, Nishimura D, Stone EM, Sheffield VC (1995) Use of a DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. Hum Mol Genet 4:9–13

Farag TI, Teebi AS (1989) High incidence of Bardet-Biedl syndrome among the Bedouin. Clin Genet 36:463–465

Green JS, Parfrey PS, Harnett JD, Farid NR, Cramer BC, Johnson G, Heath O, et al (1989) The cardinal manifestations of Bardet-Biedl syndrome, a form of Laurence-Moon-Biedl syndrome. N Engl J Med 321:1002–1009

Klein D, Ammann F (1969) The syndrome of Lawrence-Moon-Bardet-Biedl and allied diseases in Switzerland. J Neurol Sci 9:479–513

Kwitek-Black AE, Carmi R, Duyk GM, Buetow KH, Elbedour K, Parvari R, Yandava C, et al (1993) Linkage of Bardet-Biedl syndrome to chromosome 16q and evidence for nonallelic genetic heterogeneity. Nat Genet 5:392–396

Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236:1567–1570

Nelson JK, Moxness KE, Jensen M, Gastmean C (1994) Mayo Clinic diet manual of nutrition practice, 7th ed. CW Mosby, St Louis, pp 186, 657

Sheffield VS, Carmi R, Kwitek-Black A, Rokhlina T, Nishimura D, Duyk GM, Elbedour K, et al (1994) Identification of a Bardet-Biedl syndrome locus on chromosome 3 and evaluation of an efficient approach to homozygosity mapping. Hum Mol Genet 3:1331–1335

Sheffield VS, Stone EM, Carmi R (1998) Use of isolated inbred human populations for identification of disease genes. Trends Genet 14:391–396

Spurr NK, Bashir R, Bushby K, Cox A, Cox S, Hildebrandt F, Hill N, et al (1996) Report and abstracts of the Fourth International Workshop on Human Chromosome 2 Mapping 1996. Cytogenet Cell Genet 73:255–273

Woods MO, Young TL, Parfrey PS, Hefferton D, Green JS, Davidson WS (1999) Genetic heterogeneity of Bardet-Biedl syndrome in a distinct Canadian population: evidence for a fifth locus. Genomics 55:2–9

Address for correspondence and reprints: Dr. Terry-Lynn Young, Department of Biochemistry, Memorial University of Newfoundland, St. John's, Newfoundland, Canada, A1B 3X9. E-mail: tlyoung@morgan.ucs.mun.ca

0002-9297/99/6403-0030$02.00

Am. J. Hum. Genet. 64:904–908, 1999

Autosomal Dominant (Beukes) Premature Degenerative Osteoarthropathy of the Hip Joint Maps to an 11-cM Region on Chromosome 4q35

To the Editor:

We have previously reported the clinical and radiographic features of affected individuals from a large kindred who have an autosomal dominant form of bilateral dysplasia of the hip joints with severe secondary osteoarthrosis (Cilliers and Beighton 1990). This family came to the attention of one of us (H.C.) because of the number of patients with the family name, Beukes, who presented to the Department of Orthopaedic Surgery, University of Orange Free State, South Africa, for prosthetic hip-joint replacement as a consequence of bilateral premature degenerative osteoarthropathy. Genealogical studies subsequently revealed that all the affected individuals were members of an extended family that could be traced back to a single Dutch immigrant to South Africa who arrived in 1685 (Cilliers and Beighton 1990). Our continued investigation of this family has now traced 55 individuals in eight generations who, on the basis of either their medical histories or clinical and radiographic presentation of the disorder, appear to have inherited the disorder. The disorder clearly has an autosomal dominant mode of inheritance, but there is some evidence of nonpenetrance in that apparently unaffected individuals have had affected offspring. The clinical and radiographic manifestations have been described in detail elsewhere (Cilliers and Beighton 1990). In brief, the presenting symptom is hip-joint discomfort, which usually develops during childhood at age <2 years but may develop either later in childhood or, as in one instance, as late as the age of 35 years. After onset of symptoms, the hip joints deteriorate progressively, gait is disturbed, and, by early adulthood, affected persons are crippled by degenerative arthropathy. The earliest radiological changes are broadening of the femoral necks, late appearance of the secondary ossification centers of the femoral head, and an irregular appearance of the proximal epiphyseal line of the femur. By mid childhood, the femoral heads are flat (coxa plana), with broadening of the femoral necks, adaptation of the acetabulum to the mal-

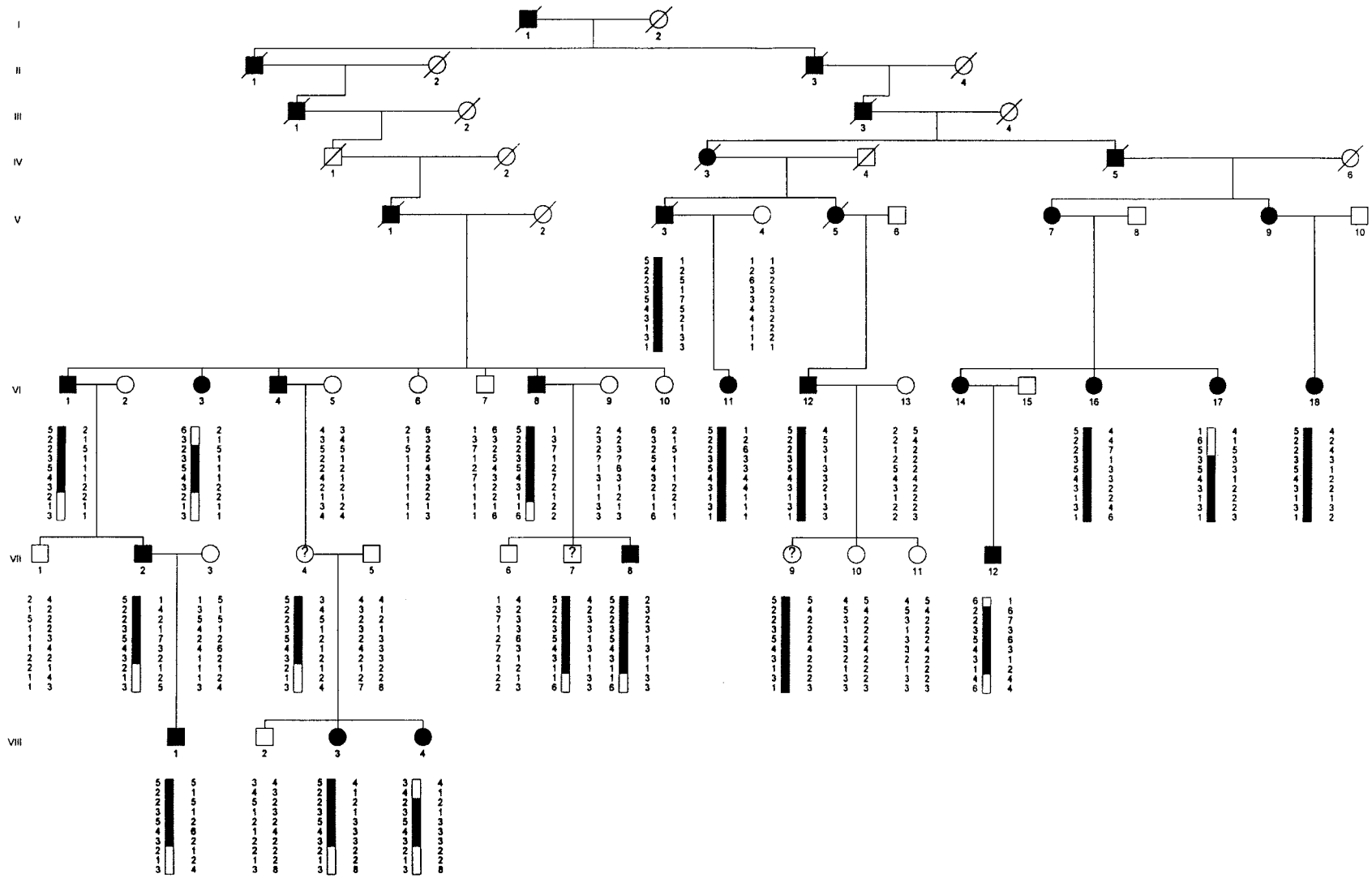

Figure 1 BHD pedigree, showing disease-linked haplotypes. Blackened circles and squares represent affected females and males, respectively. Symbols containing a question mark (?) are likely to be nonpenetrant carriers of the disease. The haplotypes for all individuals that were genotyped are given under the symbols. The DNA sample from individual V-3 was obtained before his death. The disease-linked haplotype is on the left and is indicated by a blackened box to its right. The marker order, from top to bottom, is D4S1607, D4S2951, D4S1554, D4S408, D4S2924, D4S171, D4S1540, D4S3051, D4S426, and D4S2940.

Table 1

Two-Point *Z* Values, between BHD and Chromosome 4q35 Markers

		Z AT θ =						
MARKER	Z_{max} (θ_{max})	.00	.01	.05	.10	.20	.30	.40
D4S1607	1.54 (.18)	−11.96	−2.19	.37	1.23	1.53	1.14	.48
D4S2951	1.46 (.17)	−12.20	−1.32	.63	1.26	1.43	1.07	.53
D4S1554	2.00 (.10)	−1.42	1.27	1.89	2.0	1.75	1.25	.63
D4S408	3.58 (.00)	3.58	3.55	3.37	3.05	2.23	1.30	.39
D4S2924	5.73 (.00)	5.73	5.68	5.40	4.96	3.87	2.58	1.17
D4S171	4.84 (.00)	4.84	4.79	4.55	4.15	3.16	2.01	.82
D4S1540	5.11 (.00)	5.11	5.03	4.67	4.17	3.06	1.90	.78
D4S3051	.18 (.33)	−3.14	−2.35	−1.01	−.43	.04	.17	.14
D4S426	.03 (.40)	−7.10	−2.63	−1.18	−.60	−.15	0	.03
D4S2930	.18 (.29)	−12.36	−3.64	−1.44	−.55	.07	.18	.11

NOTE.—All values are calculated under the assumption of 90% penetrance and a disease-allele frequency of .0001.

formed femoral head, superolateral displacement of the femoral head, and an irregular appearance of the greater trochanteric epiphyses. By adulthood, these features are more pronounced, and there is superior migration and superolateral displacement of the femoral head and overgrowth of the greater trochanter in a superomedial direction. In the later stages, coxa vara is a prominent finding. Signs of degenerative osteoarthrosis (periarticular cysts, periarticular sclerosis, and narrowing of the joint space) are evident in early to mid childhood and are progressive. Apart from the hip problems, the general health of affected individuals is good, their height is normal, and, other than in one instance in which a young adult had severe kyphoscoliosis that necessitated spinal fusion, involvement of the vertebral bodies and other joints is minimal. The radiographic findings, the absence of involvement of the vertebral bodies and joints other than the hip, and the normal stature of affected individuals has led to the conclusion that this disorder is distinct from other autosomal dominant forms of chondrodysplasia, in which premature degenerative osteoarthropathy of the hip joint is a major complication. It therefore seemed appropriate that this condition be categorized as a familial hip dysplasia, and it was called "Beukes hip dysplasia" (BHD [MIM 142669]), on the basis of the name of the affected family (Cilliers and Beighton 1990).

Mutations in the genes encoding components of the extracellular matrix of cartilage have been identified in families with forms of chondrodysplasia with secondary osteoarthritis (see Kuivaniemi et al. 1997; Briggs et al. 1998, and references therein). We commenced our studies to locate the gene responsible for BHD, by performing analyses of linkage to polymorphic markers within or near cartilage candidate genes. We found, however, no evidence of linkage to COL2A1 (Beighton et al. 1994), COL9A1, COL9A2, COL11A1, COL11A2, COL10A1, CRTL-1, CRTM, AGC1, or COMP (Al-Ali et al. 1994; G. Wallis, P. Roby, and S. Eyre, unpublished data). We therefore performed a genomewide screen with a panel of 290 markers with an average spacing of 11 cM (Davies et al. 1994). For this purpose, genotype data were obtained from 32 individuals from the BHD kindred, including 15 affected individuals, 11 unaffected related individuals, and 6 unrelated spouses. The affected status of the 15 individuals who were genotyped had been established on the basis of their clinical and radiological presentation of the disorder. The BHD pedigree shown infigure 1 has been condensed and includes only those individuals who were genotyped and those who were required for the linkage analysis. Despite the fact that the pedigree has been condensed, the relationships between the members of the pedigree have been retained. Genotyping was done with an ABI 373 sequencer and GENESCAN 1.2.2-1 and GENOTYPER 1.1.1 software. Two-point LOD score (*Z*) values were computed by the LINKAGE package (Lathrop and Lalouel 1984), for various recombination fraction (θ) values, with penetrance values of 90%, 95%, and 100%, and a disease frequency of .0001. One marker on chromosome 4, D4S408, had a maximum *Z* (Z_{max}) value of 3.58 at a maximum θ (θ_{max}) of .00, at 90% penetrance (see table 1). Further analysis with markers from this region, with penetrance values of 85%, 90%, and 95%, gave a two-point Z_{max} value of 5.73 for marker D4S2924, at θ = .00 and at a penetrance of 90% (see table 1). Haplotypes were constructed with the map order D4S1607–1.5 cM–D4S2951–1 cM–D4S1554–4 cM–D4S408–3.2 cM–D4S2924–0.5 cM–D4S171–0.5 cM–D4S1540–3 cM–D4S3051–1 cM–D4S426–1.3 cM–D4S2930. The order of the markers was derived from on-line genetic mapping data at the Center for Medical Genetics, Marshfield Medical Research Foundation Website. As judged on the basis of an examination of the marker haplotypes segregating with BHD (fig. 1), the closest recombinants involving affected family

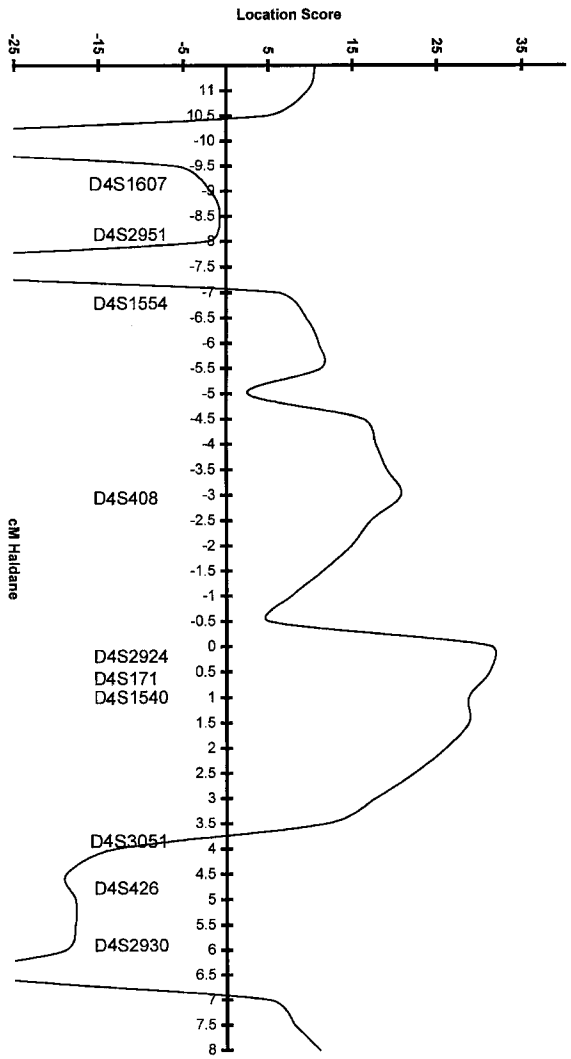

Figure 2 Graphic representation of three-point location scores based on genotype data for chromosome 4 markers and members of the BHD family. A location score is equivalent to *Z* multiplied by 4.6.

members were at D4S1554 proximally and D4S3051 distally, which limits the BHD gene locus to an interval of ~11 cM. Three apparently clinically unaffected individuals (VII-4, age 49 years; VII-7, age 33 years; and VII-9, age 33 years) were found to have inherited the disease-linked haplotype. Individual VII-4 transmitted the disorder and the disease-linked haplotype to her two affected offspring, demonstrating that she is a nonpenetrant carrier of the mutated gene. However, to date, her only potential clinical symptom of the disorder has been hip-joint pain during pregnancy. The remaining two individuals have not reported any symptoms of the disorder. Attempts are currently underway to obtain recent radiographs of these three individuals, to determine whether they have any radiological evidence of the disorder.

Multipoint analysis was done with the LINKAGE 5.1 LINKMAP program and the marker order given above. Multipoint analysis with the complete set of markers spanning the disease interval was not possible, because of both the high number of alleles per marker and the large number of individuals in the pedigree, so sequential three-point analyses were done. The combined results are shown in figure 2. The multipoint location score for the chromosome 4 markers was 30.05 (equivalent to $Z = 6.5$), and the likely location of the BHD gene was confirmed to be the 11-cM interval between D4S1554 and D4S3051. Currently, within the linked region there are no known or obvious potential candidate genes for the disease, and no other forms of familial osteochondrodysplasia are known to map to this region. Physical mapping data for this region include a single YAC contig, WC4.7, to which a number of expressed sequence tags (ESTs) have been mapped (Whitehead Institute/MIT Genome Sequencing Project).

Our finding that BHD does not map to any of the loci that have previously been identified for other forms of autosomal dominant chondrodysplasia with associated osteoarthropathy (notably, spondyloepiphyseal dysplasia [MIM 184100 and MIM 183900], multiple epiphyseal dysplasia [MIM 226900], and pseudoachondroplasia [MIM 177170]) supports the clinical and radiographic data suggesting that this disorder is a distinct form of familial hip dysplasia. Identification of the BHD gene within the linked region on 4q35 could have implications for the investigation of other, more common forms of idiopathic hip osteoarthritis.

Acknowledgments

We sincerely thank the family members who participated in this study, as well as the many genetics nurses who were involved in the collection of the DNA. We thank Mike Briggs and Mike Dixon for helpful discussions. This work was supported by grants from the Royal Society (United Kingdom), the Arthritis Research Campaign (United Kingdom), the U.C.T. Staff Research Fund, and the Mauerberger Foundation, S.A.

PHILIP ROBY,[1,2] STEPHEN EYRE,[1,2]
JANE WORTHINGTON,[2] RAJKUMAR RAMESAR,[4]
HENDRIK CILLIERS,[5] PETER BEIGHTON,[4]
MICHAEL GRANT,[1] AND GILLIAN WALLIS[1,3]

[1]*The Wellcome Trust Centre for Cell-Matrix Research, School of Biological Sciences,* [2]*Arthritis Research Campaign Epidemiology Research Unit, and* [3]*Department of Medicine, University of Manchester, Manchester;* [4]*Department of Human Genetics, University of Cape Town, Cape Town; and* [5]*University of Bloemfontein, Bloemfontein, South Africa*

Electronic-Database Information

Accession numbers and URLs for data used in this study are as follows:

Center for Medical Genetics, Marshfield Medical Research Foundation, http://www.marshmed.org/genetics/ (for marker order)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nim.nih.gov/Omim (for BHD [MIM 142669], pseudoachondroplasia [MIM 177170], spondyloepiphyseal dysplasia [MIM 183900 and MIM 184100], and multiple epiphyseal dysplasia [MIM 226900])
Whitehead Institute/MIT Genome Sequencing Project, http://carbon.wi.mit.edu (for mapping of ESTs)

References

Al-Ali M, Beighton P, Read A, Warman M, Donn R, Boot-Handford R, Wallis GA (1994) Exclusion of candidate genes in two families with inherited forms of osteoarthritis. Int J Exp Pathol 75:A71
Beighton P, Cilliers HJ, Ramesar R (1994) Autosomal dominant (Beukes) premature degenerative osteoarthropathy of the hip joint unlinked to COL2A1. Am J Med Genet 53: 348–351
Briggs MD, Mortier GR, Cole WG, King LM, Golik SS, Bonaventure J, Nuytinck L, et al (1998) Diverse mutations in the gene for cartilage oligomeric matrix protein in the pseudoachondroplasia–multiple epiphyseal dysplasia disease spectrum. Am J Hum Genet 62:311–319
Cilliers HJ, Beighton P (1990) Beukes familial hip dysplasia: an autosomal dominant entity. Am J Med Genet 36:386–390
Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, et al (1994) Genome-wide search for human type 1 diabetes susceptibility genes. Nature 371: 130–136
Kuivaniemi H, Tromp G, Prockop DJ (1997) Mutations in fibrillar collagens (types I, II, III and XI), fibril-associated collagen (type IX) and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage and blood vessels. Hum Mutat 9:300–315
Lathrop GM, Lalouel JM (1984) Easy calculations of LOD scores and genetic risks on small computers. Am J Hum Genet 36:460–465

Address for correspondence and reprints: Dr. Gillian A. Wallis, School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, United Kingdom. E-mail: gwallis@fs1.scg.man.ac.uk

0002-9297/99/6403-0031$02.00

Am. J. Hum. Genet. 64:908–910, 1999

Common Fragile Sites: G-Band Characteristics within an R-Band

To the Editor:

Common fragile sites are chromosomal loci prone to breakage and rearrangement and are considered to be part of the normal chromosome structure. They are visualized as constrictions, gaps, or breaks on metaphase chromosomes from cells exposed to specific tissue-culture conditions (Sutherland and Richards 1995). Three common fragile sites—FRA3B, FRA7H, and FRA7G—were recently cloned and identified at the molecular level (Boldog et al. 1997; Inoue et al. 1997; Huang et al. 1998; Mishmar et al. 1998). Sequence analysis of these three common fragile sites revealed no CGG or other expanded repeated sequences, such as have been found in rare fragile sites (Sutherland and Richards 1995). DNA sequence analysis of FRA3B, FRA7H, and FRA7G did not reveal any obvious feature that could account for the fragility of these sites. To shed light on the mechanism of fragility, we undertook a new approach and analyzed the available sequences of FRA3B, FRA7H, and FRA7G, for DNA structural characteristics that might be associated with their fragility (Mishmar et al. 1998). The analysis revealed several regions with a potential to form unusual DNA structures, including high flexibility, low stability, and non–B DNA–forming sequences. Thus, these unusual DNA characteristics are possibly intrinsic properties of common fragile sites, which may affect their replication, condensation, and organization and may lead to fragility.

While analyzing the sequences of FRA3B, FRA7H, and FRA7G, we noticed several features that are characteristic of G-bands (Gardiner 1995). The three cloned fragile sites have high (>57%) A/T content, and are all gene poor. FRA3B and FRA7H are rich in LINE sequences. FRA3B and markers proximal to FRA7G were shown to replicate late during S-phase (Selig et al. 1992; Huang et al. 1998; Le Beau et al. 1998). G-bands and R-bands correspond to functional subregions, represented as stained bands, that apparently reveal the basic structural organization of chromosomes. G-bands are characterized as regions with high A/T content that replicate late during S-phase, are insensitive to DNase-I, and are gene poor, Alu poor, and LINE rich. In contrast, the complementary R-bands are regions with high G/C content that replicate early during S-phase, are DNase sensitive, and are gene rich, Alu rich, and LINE poor (Gardiner 1995). Most (76/89 [>85%]) of the common fragile sites, including the cloned sites, map to R-bands (according to our analysis of the Genome Database data). These characteristics suggest that fragile sites

Table 1

Clusters of Regions with High (>13.7%) Flexibility in Common Fragile Sites and in R- and G-Band Sequences

Type of Sequence and Variable Locus	Length (bp)	Region	Flexibility Cluster
Common fragile sites, >57% A/T:			
FRA3B (U66722 and AF020503)	276,822	3p14.2	1
FRA7G (AC002066)	151,770	7q31.2	1
FRA7H (AF017104)	161,115	7q32.3	1
G-band, >57% A/T:			
HS79C4	158,548	1q24	1
HS106H8	175,825	1q24	None
HS206D15	101,574	1q24	1
HS117P20	141,589	1q24	None
AC004615	235,141	5p15.2	3
HS451B15	186,510	6p24	None
AC003091	137,817	7p21	1
AC003075	123,336	7p21	1
AC004741	112,242	7p21	None
AC004492	165,608	7q31.1	1
U66059	267,156	7q35	1
U66060	215,422	7q35	1
U66061	232,650	7q35	1[a]
AC002526	134,580	Xq23	1
HSAC002086	112,686	Xq23	None
AC005191	185,893	Xq23	None
AC002476	147,102	Xq23	2
HS75N13	141,851	Xq21.1	1
G-band, <57% A/T:			
HS934G17	107,603	1p36.21	None
HS232K4	198,161	6p22.3	None
HS257A7	127,917	6p24	None
HS445C9	131,398	22q12.1	None
R-band, >57%:			
AC003099	95,129	4q25	None
AC003100	90,430	4q25	1
HS111M5	107,526	6p21.3	None
HS172K2	131,234	6p21.3	None
HS265J14	90,547	6p21.3	None
HS167A14	132,790	6q27	None
HS155D22	148,851	6q27	2
L11910	180,388	13q14.1	None
HSU91325	135,046	16p13.11	None
HS393P12	104,597	Xp11.21	None
HSU82696	158,898	Xq28	None
HSU40455	106,000	Xq28	None
AF003627	81,007	Xq28	1
HS884M20	114,173	Xq28	None
R-band, <57% A/T:			
HSU91326	150,296	16p11.2	None
HSAF001548	145,831	16p13.11	None
HSU95738	171,368	16p13.11	None
HSU95740	149,490	16p13.1	None
HSU95737	93,481	16p13.1	None
HS339A18	132,805	Xp11.2	None
HSU07000	152,141	22q11	None
HSU52111	153,460	Xq28	None
HSU52112	174,424	Xq28	None
HSU82672	156,854	Xq28	None
AF002992	104,037	Xq28	1
AF003628	86769	Xq28	None

[a] Same cluster as in U66060 (U66060 and U66061 are in contig).

might share, with the chromatin of G-bands, structural features (as well as other features) that distinguish the fragile sites from their flanking R-band sequences. This different chromatin organization might affect the replication and condensation of the fragile sequences and thus may contribute to the fragility.

Here we focus on one of the characteristics identified in common fragile sites: high DNA flexibility. We assessed the flexibility by measuring potential local variations in the DNA structure at the twist angle (for details, see Mishmar et al. 1998). A region with potential high flexibility was defined as a region deviating significantly from the average value of the entire analyzed sequence (Mishmar et al. 1998). Because DNA flexibility appears to play an important role in protein-DNA interactions, it could well affect chromatin condensation and organization (Sarai et al. 1989). Our previous flexibility analysis of FRA7H, FRA3B, and FRA7G revealed impressive clusters of regions with high flexibility, at all three sites. A cluster of regions with high flexibility is defined as at least three high-flexibility peaks in a region of <40 kb, on the basis of the flexibility pattern identified in the three common fragile sites (Mishmar et al. 1998). In control sequences comprising 14 genomic sequences that map to chromosomal bands in which fragile sites were not described (1.1 Mb in all), regions with high flexibility appeared approximately every 100 kb. No clusters of regions with high flexibility were identified in these control sequences (Mishmar et al. 1998).

Because fragile sites appeared to share several features with G-bands, we decided to extend our analysis of flexibility patterns (Mishmar et al. 1998) to sequences of known band localization and base composition. For this purpose we have developed a user-friendly computer program (FlexStab) that enables flexibility analysis of sequences as much as 350 kb in length. The program is available from our Website at Hebrew University of Jerusalem. The analysis was performed on available GenBank sequences that map to chromosomal bands in which fragile sites have not been described. The mapping information was drawn from Genome Database 6.0, on the basis of the resolution of ~500 chromosomal bands. Our search identified ~6.9 Mb of DNA sequences of >80 kb each, 3.53 Mb from G-bands and 3.35 Mb from R-bands. Most (2.97 Mb [89%]) of the G-band sequences had, as expected, high (>57%) A/T content (table 1). Fourteen clusters of regions with high flexibility were identified in this group. No clusters of regions with high flexibility were identified in the available G-band sequences (0.56 Mb) with low (<57%) A/T content table 1). In R-band sequences with high A/T content (1.68 Mb), four clusters were identified (table 1). In R-band sequences with low A/T content (1.67 Mb), only one cluster was identified (table 1). There were significantly more high-flexibility clusters in A/T-rich G-band se-

quences than in A/T-rich R-band sequences ($P = .009$). Thus, the flexibility pattern is one of the features that differentiate R- and G-bands.

As mentioned, common fragile sites were found to be A/T rich. The pattern of high-flexibility clusters found in the identified common fragile sites (see table 1) (Mishmar et al. 1998) was significantly different ($P = .02$) from that of A/T-rich control sequences mapped to R-bands. This pattern was not different from that of A/T-rich control sequences mapped to G-bands ($P = .85$). These results might indicate that common fragile sites mapped to R-bands have the flexibility patterns characteristic of G-bands with the same A/T content.

Our previous analysis of potential unusual DNA structures in FRA7H revealed a cluster of regions with potential to form triple helixes (Mishmar et al. 1998). Previous studies, using monoclonal antibodies to triple-helix DNA, showed that G-bands are rich in triple-helix DNA (Burkholder et al. 1991). Thus, clusters of regions with potential to form triple-helix DNA might be added to the G-band characteristics found in common fragile sites.

Together, all the known molecular features of common fragile sites indicate that they might consist of DNA sequences with characteristics of G-bands embedded within R-bands. Of what significance could this feature be to the mechanism of fragility? We think that delayed replication and aberrant condensation of fragile sites might be involved. Chromosomal bands apparently represent regions with several origins of replication that are coordinately controlled to initiate the replication process. The presence of a relatively small region consisting of a common fragile site with G-band characteristics might lead to disturbances in the regional control of replication. This might involve inappropriate initiation of replication in the fragile region. The addition of aphidicolin, which inhibits DNA elongation, might further add to the interference in replication at fragile sites, leading to unreplicated sequences that might adopt abnormal chromatin organization, resulting in fragility.

Dan Mishmar,[1] Yael Mandel-Gutfreund,[2] Hanah Margalit,[2] Ayelet Rahat,[1] and Batsheva Kerem[1]

[1]*Department of Genetics, The Hebrew University, and* [2]*Department of Genetics and Biotechnology, The Hebrew University, Hadassah, Jerusalem*

Electronic-Database Information

URLs for data in this article are as follows:

Genome Database, http://www.gdb.org

Hebrew University of Jerusalem, http://leonardo.ls.huji.ac.il/departments/genesite/faculty/bkerem.htm

References

Boldog F, Gemmill RM, West J, Robinson M, Robinson L, Li E, Roche J, et al (1997) Chromosome 3p14 homozygous deletions and sequence analysis of FRA3B. Hum Mol Genet 6:193–203

Burkholder GD, Latimer LJ, Lee JS (1991) Immunofluorescent localization of triplex DNA in polytene chromosomes of Chironomus and Drosophila. Chromosoma 101:11–18

Gardiner K (1995) Human genome organization. Curr Opin Genet Dev 5:315–322

Huang H, Qian J, Proffit J, Wilber K, Jenkins R, Smith DI (1998) FRA7G extends over a broad region: coincidence of human endogenous retroviral sequences (HERV-H) and small polydispersed circular DNAs (spcDNA) and fragile sites. Oncogene 16:2311–2319

Inoue H, Ishii H, Alder H, Synder E, Druck T, Huebner K, Croce CM (1997) Sequence of the FRA3B common fragile region: implications for the mechanism of FHIT deletion. Proc Natl Acad Sci USA 94:14584–14589

Le Beau MM, Rassool FV, Neilly ME, Espinosa III R, Glover TW, Smith DI, McKeithan TW (1998) Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. Hum Mol Genet 7:755–761

Mishmar D, Rahat A, Scherer SW, Nyakatura G, Hinzmann B, Kohwi Y, Mandel-Gutfreund Y, et al (1998) Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of an SV40 integration site. Proc Natl Acad Sci USA 95:8141–8146

Sarai A, Mazur J, Nussinov R, Jernigan RL (1989) Sequence dependence of DNA conformational flexibility. Biochemistry 28:7842–7849

Selig S, Okumura K, Ward DC, Cedar H (1992) Delineation of DNA replication time zones by fluorescence in situ hybridization. EMBO J 11:1217–1225

Sutherland GR, Richards RI (1995) The molecular basis of fragile sites in human chromosomes. Curr Opin Genet Dev 5:323–327

Address for correspondence and reprints: Dr. Batsheva Kerem, Department of Genetics, The Hebrew University, Jerusalem, Israel, 91904. E-mail: kerem@leonardo.ls.huji.ac.il

0002-9297/99/6403-0032$02.00

Am. J. Hum. Genet. 64:910–915, 1999

Finite-Sample Properties of Family-Based Association Tests

To the Editor:

During the past few years, there has been much interest in the use of family-based association tests to detect linkage between marker and disease loci, since these methods avoid the problems of ascertaining the appropriate pop-

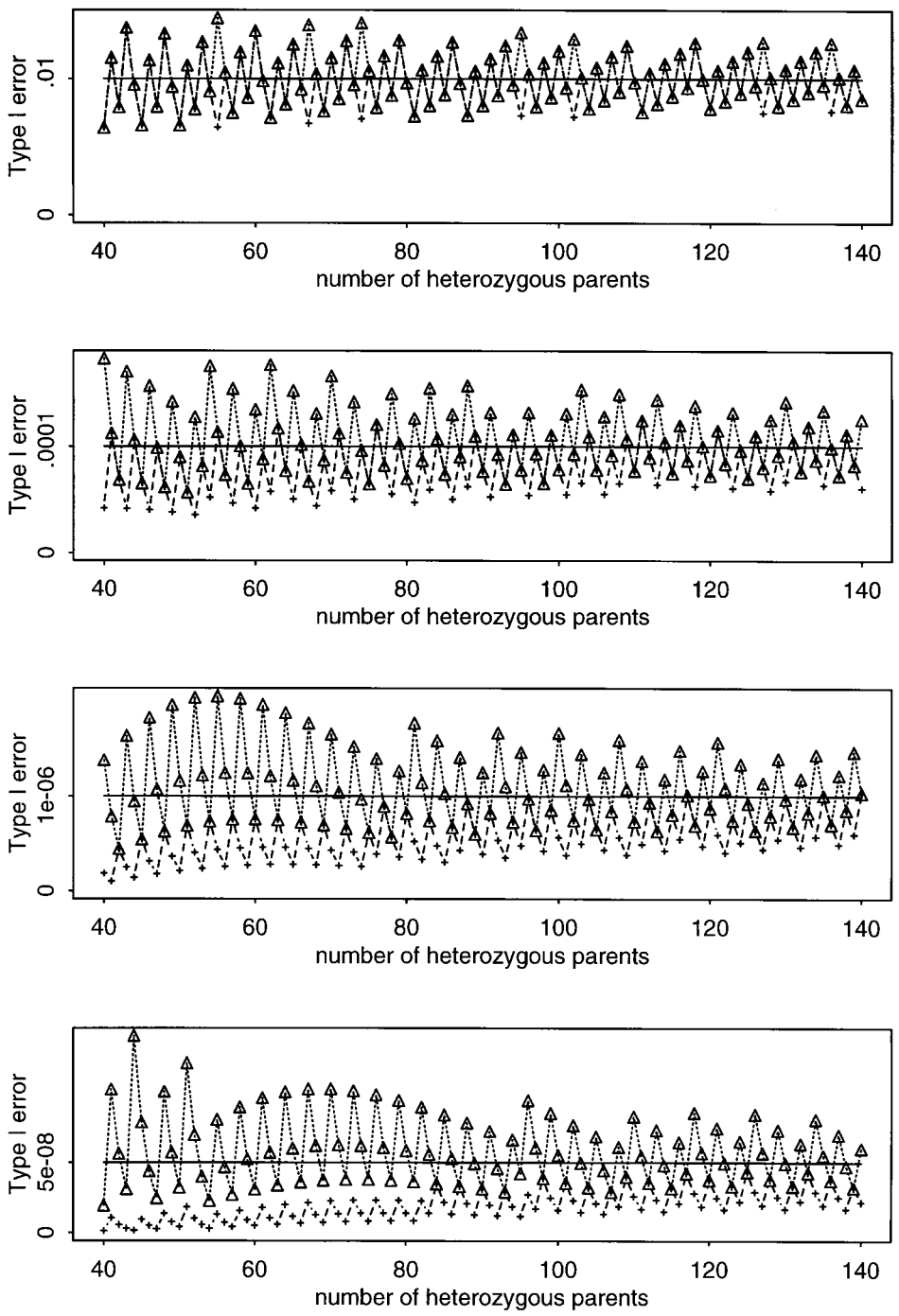

Figure 1 Achieved type I errors for $T_{TDT}(+)$ and Λ (Δ), compared with specified α (*unbroken horizontal line*), for (*from top to bottom*) $\alpha = .01, .0001, 1 \times 10^{-6}$ (1e-06), and 5×10^{-8} (5e-08).

ulations of cases and controls implicit in population association studies. Although these tests were originally developed for candidate-gene studies, the use of such methods in genome scans has recently been proposed (Risch and Merikangas 1996).

Perhaps the best-known family-based association test is the transmission/disequilibrium test (TDT) for diallelic markers, introduced by Spielman and Ewens (1993). A number of similar tests have subsequently been suggested; for reviews, see the work of Spielman and Ewens (1996) and Curnow et al. (1998). However, it has often not been clear how these various tests are related, and

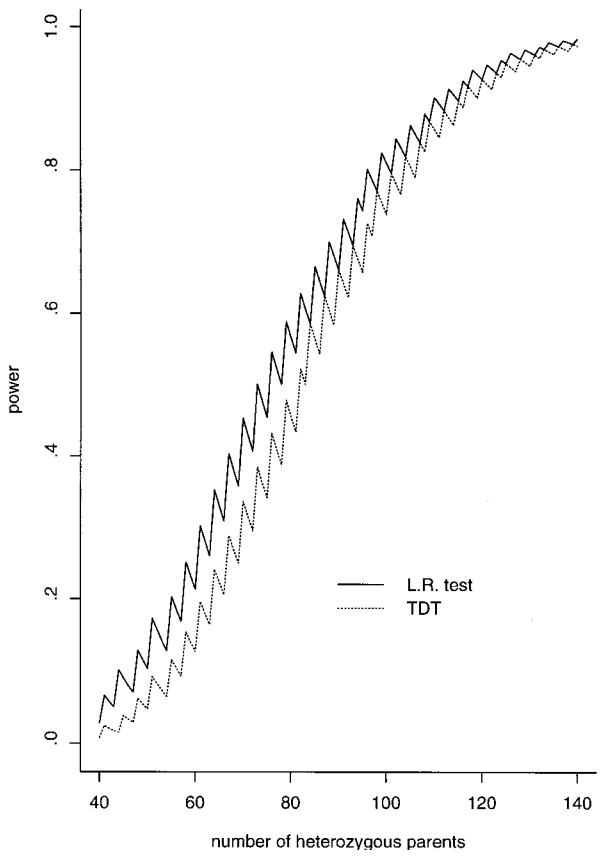

Figure 2 Power versus number of heterozygous parents for T_{TDT} and likelihood-ratio statistic Λ (L.R. test), when $\gamma = 4$ and $\alpha = 5 \times 10^{-8}$.

there has been debate about the advantages and disadvantages of several of the tests (e.g., see Kaplan et al. 1997; Sham 1997). There is a need to investigate the relationships between the suggested tests and to establish which of them should be preferred in a given situation. Determining the relative merits of competing test statistics is often difficult, because the comparisons usually rely on simulation or on asymptotic results that may be of limited relevance to finite-sample data. Here we suggest a way of avoiding these problems for diallelic markers, focusing particularly on the TDT, the extended TDT (ETDT [Sham and Curtis 1995]), the score tests introduced by Schaid (1996), and tests based on the conditional likelihood of the offspring marker types when the parental marker data are given.

We consider a sample of N families, each with a single affected child. All individuals have been genotyped at a marker locus with m alleles, labeled as "M_1," M_2,...,M_m." We wish to use information on the alleles transmitted from the parents to the affected child, to test the null hypothesis of no linkage or no association between the marker and disease. Note that we can only test this compound null hypothesis when we have a single affected child in each family; for families with multiple affected children, the tests discussed below are valid only as tests of linkage, not as tests of association (Spielman and Ewens 1996).

Consider a single family, with parental marker genotypes **g** and **h** and with the genotype of the affected child denoted by **x**. If we use C_A to denote that a child is affected, then the probability of the child's genotype, conditional on the parental genotype is, by Bayes's theorem,

$$P(\mathbf{x}|g,h,C_A) = \frac{P(C_A|\mathbf{x},\mathbf{g},\mathbf{h})P(\mathbf{x}|g,h)P(\mathbf{g},\mathbf{h})}{\sum_{x^* \in G} P(C_A|\mathbf{x}^*,\mathbf{g},\mathbf{h})P(\mathbf{x}^*|g,h)P(\mathbf{g},\mathbf{h})}$$

$$= \frac{P(C_A|\mathbf{x})P(\mathbf{x}|g,h)}{\sum_{x^* \in G} P(C_A|\mathbf{x}^*)P(\mathbf{x}^*|g,h)},$$

where G is the set of possible marker types for the affected child (Schaid 1996). We shall assume normal segregation (which requires, e.g., absence of meiotic drive), so that $P(\mathbf{x}|\mathbf{g}, \mathbf{h})$ is easily calculated, leaving only $P(C_A|\mathbf{x})$, the risk of disease for a particular marker genotype, to be discussed.

We could model the disease locus explicitly, but it is often more convenient (Self et al. 1991; Schaid 1996) to work directly with the marker genotype, by putting $f_{\mathbf{x}} = P(C_A|\mathbf{x})$. The $f_{\mathbf{x}}$ then reflects both the disease-locus penetrances and the strength of allelic association between the marker and disease loci in the affected children. Note that $m(m + 1)/2$ parameters are needed in the general model; this will be large for highly polymorphic markers. Schaid (1996) derives score tests for general $f_{\mathbf{x}}$ and for various special cases representing particular disease models. In particular, Schaid shows that, for a log-additive model—that is, one in which allelic effects combine multiplicatively at the marker, so that $f_{\mathbf{x}} = f_{x_1} f_{x_2}$—the score test for a diallelic marker is the TDT statistic.

Note that we are using this multiplicative model as a convenient approximation, rather than as something that we believe is exactly correct. If the true disease model is not multiplicative, then tests based on the multiplicative model remain valid in the sense of having the correct size, but they may not be optimal, in that there may exist other test statistics with higher power. However, the multiplicative model has the advantage of requiring $(m - 1)$ parameters, where tests derived by means of the general model need $m(m + 1)/2$. This means that tests based on the multiplicative model can be more powerful than general alternatives even when the allelic effects do not combine multiplicatively at the marker (e.g., see Schaid 1996).

Schaid (1996) comments that likelihood-ratio tests

could be used instead of score tests, but he opts for score statistics because of their ease of calculation. Here we derive the likelihood-ratio test for the multiplicative disease model discussed above.

It can be shown (e.g., see Curnow et al. 1998) that, in the presence of allelic association, the marker alleles transmitted to an affected child from the child's two parents are independent if and only if the multiplicative model holds at the marker. Therefore, under the multiplicative model, $P(\mathbf{x}|g,h,C_A) = P(x_1|\mathbf{g},C_A)P(x_2|\mathbf{h},C_A)$, where x_1 and x_2 are the alleles transmitted from parents with genotypes **g** and **h**, respectively. Now,

$$P(x_1 = i|\mathbf{g},C_A) = 0 \quad \text{if } g_1 \neq i \text{ and } g_2 \neq i ,$$

$$P(x_1 = i|\mathbf{g},C_A) = 1 \quad \text{if } g_1 = g_2 = i ,$$

$$P(x_1 = i|\mathbf{g},C_A) = \frac{f_i P(x_1 = i|\mathbf{g})}{f_{g_1}P(x_1 = g_1|\mathbf{g}) + f_{g_2}P(x_1 = g_2|\mathbf{g})} \quad \text{otherwise .}$$

so that, if $\mathbf{g} = (i,j)$ and $x_1 = i$, then

$$P(x_1|\mathbf{g},C_A) = 1 \quad \text{if } i = j ,$$

$$P(x_1|\mathbf{g},C_A) = \frac{f_i}{(f_i + f_j)} \quad \text{if } i \neq j ,$$

because $P(x_1 = i|\mathbf{g}) = P(x_1 = j|\mathbf{g}) = .5$.

Let n_{ij} be the number of transmissions of M_i from M_iM_j parents in our sample of $2N$ parents. Then $L(\mathbf{f})$, the likelihood of the child genotypes, given the parental genotypes, under the multiplicative model, is

$$L(\mathbf{f}) = \prod_{i=1}^{m}\prod_{j<i}\binom{n_{ij}+n_{ji}}{n_{ij}}\left(\frac{f_i}{f_i+f_j}\right)^{n_{ij}}\left(\frac{f_j}{f_i+f_j}\right)^{n_{ji}} ,$$

by derivation from the formula above. This can be maximized over **f**, to give L_A. The null hypothesis is no linkage or no association between marker and disease; in this case, the two parental marker alleles are equally likely to be transmitted, so that the likelihood under the null hypothesis is

$$L_0 = \prod_{i=1}^{m}\prod_{j<i}\binom{n_{ij}+n_{ji}}{n_{ij}}\left(\frac{1}{2}\right)^{n_{ij}} .$$

The likelihood-ratio statistic is $\lambda = L(\mathbf{f})/L_0$, and, by standard theory, $-2\ln(\lambda)$ has an approximate χ^2 distribution with $m - 1$ df, under the null hypothesis.

Note that $L(\mathbf{f})$ is equivalent to the likelihood derived by Sham and Curtis (1995), with our f_i being equivalent to their d_{ii}. Sham and Curtis (1995) made assumptions that, at first sight, seem to be rather different from those which we have made here: they assume that there is no recombination between marker and disease loci and that parental transmissions of marker alleles are independent. However, as we have noted above, parental transmissions are independent if and only if the multiplicative model holds at the marker locus; the two sets of assumptions are therefore directly equivalent, and we should expect to obtain the same likelihoods.

In summary, the score test for the multiplicative model is Schaid's (1996) general TDT statistic, and the likelihood-ratio test is Sham and Curtis's (1995) ETDT statistic. By standard theory (Cox and Hinkley 1974), these tests are asymptotically equivalent; we will now show that, for diallelic markers, a stronger result holds.

Remember that in this case the score test is the TDT of Spielman et al. (1993), so that, for a test of size α, we reject the null if

$$T_{\mathrm{TDT}} = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} > k_{\mathrm{TDT}} ,$$

where the critical value k_{TDT} is chosen to give the required type I error rate α. In most cases, it seems that $L(\mathbf{f})$ must be maximized numerically, but, if $m = 2$, then we have

$$L(\mathbf{f}) \propto \frac{f_1^{n_{12}}f_2^{n_{21}}}{(f_1+f_2)^{n_{12}+n_{21}}} ,$$

and it is easy to show that $L(\mathbf{f})$ is maximized when $f_1/f_2 = n_{12}/n_{21}$. The likelihood-ratio test with size α is therefore likely to reject the null hypothesis if

$$\Lambda = 2\ln\left\{\frac{n_{12}^{n_{12}}n_{21}^{n_{21}}}{[0.5(n_{12}+n_{21})]^{n_{12}+n_{21}}}\right\} > k_\Lambda ,$$

where, again, the critical value k_Λ is chosen to give the required type I error rate α. Usually, asymptotic results are relied on, so that $k_\Lambda = k_{\mathrm{TDT}} = \chi^2_{1,1-\alpha}$.

Now suppose that there are H heterozygote parents in the sample. The values of T_{TDT} and Λ are completely determined by n_{12}, because $n_{21} = H - n_{12}$. Both of the aforementioned tests can be rewritten with rejection region $\{c_l > n_{12}\} \cup \{n_{12} > c_u\}$, where $c_l = H - c_u$ by symmetry; in fact, it is clear that any reasonable test statistic must have this form. Moreover, under the multiplicative disease model discussed above, the probability that a heterozygote parent will transmit the M_1 allele rath-

er than the M_2 is $\nu = f_1/(f_1 + f_2)$, so that, conditional on H heterozygote parents in the sample, $n_{12} \sim \text{binomial}(H,\nu)$. The constant c_u can be chosen to give the approximate type I error rate, by use of this binomial distribution; any c_u corresponds to a particular k_Λ and k_{TDT}, and, clearly, for a particular c_u, the test statistics will be exactly equivalent. It follows that, provided that type I error rates are properly controlled by appropriate choice of the critical values k_Λ and k_{TDT}, Λ and T_{TDT} will have identical power. However, the appropriate critical values will be the same for the two tests only if Λ and T_{TDT} have the same distribution, for then we will have $k_\Lambda = k_{TDT}$ for any c_u; if not, then setting $k_\Lambda = k_{TDT} = \chi^2_{1,1-\alpha}$ will give different rejection regions—and, therefore, different sizes and powers—for Λ and T_{TDT}. We now consider the properties of the test statistics if the χ^2 approximation is used.

It is convenient to put $\gamma = f_1/f_2$ so that $\nu = \gamma/(1 + \gamma)$ and to condition on there being H heterozygote parents, so that the distribution of the tests statistic depends only on the two parameters H and ν. Note that the null hypothesis that the two parental maker alleles are equally likely to be transmitted corresponds to $\gamma = 1$. We can investigate the *actual* size and power of the test directly, because, for any α, it is easy to calculate the probability that T_{TDT} or Λ is greater than $\chi^2_{1,1-\alpha}$, via the above binomial distribution. We consider $\alpha = .01, .0001, 1 \times 10^{-6}$, and 5×10^{-8}. The first two values of α might be appropriate for candidate loci, whereas the last has been suggested, by Risch and Merikangas (1996), for genome scans using the TDT.

First, consider the actual type I error rates, which are shown in figure 1, for $\alpha = .01, .0001, 1 \times 10^{-6}$, and 5×10^{-8}. Achieved type I errors for Λ oscillate about the asymptotic size, α, with the amplitude of the oscillation being relatively greatest for small α. Achieved type I errors for T_{TDT} are less variable, and T_{TDT} is conservative for small α. Overall, the null distributions of T_{TDT} and Λ are well approximated by χ^2_1, for most α, with the approximation much less satisfactory in the extreme tails of the distribution, despite reasonably large sample sizes. This suggests that the χ^2 approximation should not be used in genome scans, unless sample sizes are very large. The oscillation of the type I errors about the asymptotic size is caused by the underlying discreteness of the data; for example, the critical value $\chi^2_{1,1-\alpha}$ is 29.72 when $\alpha = 5 \times 10^{-8}$, and, if $H = 138$ or $H = 139$, this is exceeded by T_{TDT} only if $n_{21} \geq 102$ or $n_{12} \geq 102$. The type I error rate for T_{TDT} for $H = 138$ or $H = 139$ is therefore $P(n_{12} \geq 102) + P(n_{21} \geq 102))$ when $\gamma = 1$, which is greater if $H = 139$ than if $H = 138$. However, if $H = 140$, then T_{TDT} exceeds 29.72 only if $n_{12} \geq 103$ or $n_{21} \geq 103$, so that the type I error rate for T_{TDT} is $P(n_{12} \geq 103) + P(n_{21} \geq 103)$, for $\gamma = 1$, and this is less than the type I error rate for $H = 139$. Note that the effect on the error rate can be large, even for reasonable sample sizes; for example, if $H = 118$, then the achieved type I error rate for Λ is 8.55×10^{-8}, nearly twice the nominal 5×10^{-8}.

We stress again that, if critical values are correctly set, then T_{TDT} and Λ have identical size and power. However, the χ^2 approximation does not provide the correct critical values, and we will now show that this also leads to misleading power results. In particular, the fact that Λ tends to have higher type I error rates than does T_{TDT}, when the χ^2 approximation is used, can result in Λ appearing to have more power than T_{TDT} does; for example, consider figure 2, which plots the power of Λ and T_{TDT} as a function of H, for $\alpha = 5 \times 10^{-8}$ and $\gamma = 4$. We see that, although T_{TDT} and Λ have similar power for small γ, Λ can have considerably higher power for large γ, particularly for small α. The fact that power does not increase monotonically with sample size is, at first sight, surprising but, as with the oscillations in the type I error that have been noted above, is explained by the underlying discreteness of the data.

We have derived the likelihood-ratio test for a sample of families, in each of which there is a single affected child and all individuals have been genotyped at a particular marker locus, under the assumption that allelic effects combine multiplicatively at the marker. This test proves to be the ETDT (Sham and Curtis 1995), because the alleles transmitted to an affected child are independent if and only if the multiplicative model holds. For diallelic markers, the finite-sample properties of this statistic can be investigated by exact calculation, because then, for H heterozygous parents, only $H + 1$ outcomes need be considered, and these are easily enumerated. The computations become more complicated if the marker under consideration has more than two alleles, but, as a way of assessing the properties of test statistics, this type of exact calculation may be preferable to simulation, because it gives the exact sampling distribution of the test statistic under investigation, at any point in the parameter space.

For diallelic markers, the likelihood-ratio test Λ and the corresponding score test T_{TDT} are exactly equivalent, provided that type I error rates are correctly controlled. Exact calculation of the distribution of Λ and T_{TDT} shows that type I error rates are adequately controlled by reliance on asymptotic distributions for α that are appropriate for candidate loci, but not for the very small α required for genome screens. It thus seems that, if Λ or T_{TDT} is to be used in genome scans, then either significance levels must be calculated exactly, by means of the binomial distribution given here, or Monte Carlo approaches (Kaplan et al. 1997; Morris et al. 1997) must be used. Note also the exact test of Cleves et al. (1997).

It is easy to derive corresponding likelihood tests for other models, such as those for recessive or dominant

diseases, or for the general model in which no relationship is assumed between the f_x, and it is also easy to extend the test to cope with multiple affected or unaffected sibs (Thompson 1997). The properties of such tests deserve further study; comparisons with the corresponding score tests (Schaid 1996) would be of particular interest. By analogy with the results given here, reliance on asymptotic null distributions should be avoided for such tests; also see the work of Chapman (1976). Monte Carlo approaches are always available, and they provide a simple alternative to asymptotic approximations.

Finally, note that the results given here illustrate the problems of comparing the test statistics by stochastic simulation, especially when asymptotic distributional results are relied on. When the χ^2 approximation is used, the properties of Λ and T_{TDT} are very sensitive to the value of H, and to exhaustively survey the relevant parameter space by simulation would be very time consuming. On the basis of simulation results produced by the asymptotic χ^2 distribution, it would have been easy to conclude, according to the values of α, H, and γ studied, either that Λ is preferable to T_{TDT}, because it has higher power for some parameter values, or that T_{TDT} is preferable to Λ, because Λ can be anticonservative for certain values of H. For example, if $\alpha = 5 \times 10^{-8}$ and $H = 100$, then T_{TDT} is conservative, whereas the achieved type I error rate for Λ is close to the nominal value, so we might conclude that Λ is preferable. However, for $\alpha = 5 \times 10^{-8}$ and $H = 110$, T_{TDT} is still conservative but Λ is unacceptably anticonservative, and we would prefer T_{TDT}. As we have seen, we would be mistaken in both cases, because the differences are due to failure of the asymptotic χ^2 approximation, rather than to differences between the test statistics.

Acknowledgments

We are grateful to Robert Curnow and Andrew Morris for helpful discussions of this work and to two anonymous referees who made useful comments on this letter.

J. C. Whittaker[1] and D. J. Thompson[2]
[1]*Department of Applied Statistics, University of Reading, Reading, United Kingdom; and* [1]*CRC Genetic Epidemiology Unit, University of Cambridge, Cambridge*

References

Chapman JW (1976) A comparison of the χ^2, $-2\log R$ and multinomial probability criteria for significance tests when expected frequencies are small. J Am Stat Assoc 71:854–863

Cleves MA, Olson JM, Jacobs KB (1997) Exact transmission-disequilibrium tests with multiallelic markers. Genet Epidemiol 14:337–347

Cox DR, Hinkley DV (1974) Theoretical statistics. Chapman & Hall, London

Curnow RN, Morris AP, Whittaker JC (1998) Locating genes involved in human diseases. Appl Stat 47:63–76

Kaplan NL, Martin ER, Weir BS (1997) Power studies for transmission/disequilibrium tests with multiple alleles. Am J Hum Genet 60:691–702

Morris AP, Curnow RN, Whittaker JC (1997) Randomisation tests of disease-marker association. Ann Hum Genet 61: 49–60

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Schaid DJ (1996) General score tests for associations of genetic markers with disease using cases and their parents. Genet Epidemiol 13:423–449

Self SG, Longton G, Kopecky KJ, Liang KY (1991) On estimating HLA-disease association with application to a study of aplastic anemia. Biometrics 47:53–61

Sham P (1997) Transmission/disequilibrium tests for multiallelic loci. Am J Hum Genet 61:774–778

Sham PC, Curtis D (1995) An extended transmission/disequilibrium test (TDT) for multiallele marker loci. Ann Hum Genet 59:323–336

Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 59:983–989

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516

Thompson DJ (1997) Using information from unaffected sibs in genetic association studies. MS diss, Department of Applied Statistics, University of Reading, Reading, UK

Address for correspondence and reprints: Dr. John Whittaker, Department of Applied Statistics, University of Reading, P.O. Box 240, Whiteknights Road, Reading, RG6 2FN, United Kingdom. E-mail: j.c.whittaker@reading.ac.uk

0002-9297/99/6403-0033$02.00