Letters to the Editor

Am. J. Hum. Genet. 66:309–312, 2000

Isoform-Specific Imprinting of the Human *PEG1/MEST* Gene

To the Editor:

Mouse *Peg1/Mest* encodes a protein with sequence homology to the alpha/beta-hydrolase (Sado et al. 1993). The gene maps to an imprinted region of mouse chromosome 6 and is expressed monoallelically from the paternal allele (Kaneko-Ishino et al. 1995). When the null allele is paternally transmitted, the offspring exhibit severe intrauterine growth retardation (Lefebvre et al. 1998). Uniparental disomy of mouse chromosome 6 is associated with a similar phenotype, presumably as a result of lack of expression of *Peg1/Mest* (Ferguson-Smith et al. 1991). The human homologue, *PEG1/MEST,* has been mapped to 7q31.3, within a region of conserved synteny corresponding to mouse chromosome 6, and is monoallelically expressed from the paternal allele in a wide variety of tissues during prenatal and postnatal development. Uniparental disomy of chromosome 7 in humans is associated with phenotypic features of Russell-Silver syndrome (MIM 180860), characterized by intrauterine growth retardation with dysmorphic features such as triangular facies. *PEG1/MEST,* as the only known imprinted gene on chromosome 7, has been considered a candidate gene for the syndrome (Kobayashi et al. 1997; Lefebvre et al. 1997; Riesewijk et al. 1997).

Imprinting of *PEG1/MEST* is apparently lost in lymphocytes and transformed lymphoblastoid cell lines. In these tissues, *PEG1/MEST* is apparently expressed from both the paternal allele and the maternal allele (Riesewijk et al. 1997). Furthermore, *PEG1/MEST* is transcribed in lymphoblastoid cell lines from patients with maternal uniparental disomy of chromosome 7, or "upd(7)mat" (Cuisset et al. 1997; Riesewijk et al. 1997). Because upd(7)mat cells lack a paternal allele of *PEG1/MEST,* the transcript must derive from the maternal allele.

The purpose of this report is to delineate the underlying mechanism of apparent loss of imprinting in lymphocytes,to better understand the control of imprinting of the human *PEG1/MEST* gene. In general, loss of imprinting may be accounted for by several mechanisms. First, imprinting can be regulated in a tissue-specific way. Relaxation of imprinting or biallelic expression of imprinted genes is observed in some tissues. Examples include insulin (*Ins*) *1* and *Ins2* (Giddings et al. 1994; Deltour et al. 1995), and *Ube3a* (Albrecht et al. 1997). Second, imprinting may be controlled in a promoter-specific manner. Such promoter-specific imprinting was first identified in the *IGF2* gene (Vu and Hoffman 1994; Ekstroem et al. 1995). In liver and chondrocytes, the *IGF2* transcript from the P1 promoter is always derived from both the paternal allele and the maternal allele, whereas transcripts from other promoters (P2–P4) are expressed solely from the paternal allele. This finding demonstrated that both imprinting and a lack of imprinting could occur within a single gene in a single tissue, suggesting that regional imprinting factors might be important. Third, imprinting can be governed in an isoform-specific way when a single transcription unit encodes different proteins. Maternally derived (e.g., from *NESP55*), paternally derived (e.g., from *XLALPHAS*), and biallelically derived (e.g., from *GSALPHA*) proteins are produced by different patterns of promoter use and alternative splicing of a single transcription unit, *GNAS1* (Hayward et al. 1998; Peters et al. 1999).

Comparison of the 5′ end of the expressed-sequence tag (EST) sequences assembled as the *PEG1/MEST* UniGene cluster (Hs. 79284) revealed that six EST clones—AA305098 and AA305289 (colon carcinoma), AA337069 (endometrial tumor), R18211 (infant brain), AA095601 (8-wk fetal heart), and AA092738 (10-wk fetal heart)—share a novel sequence joined to exon 2 of *PEG1/MEST* (Cuisset et al. 1997), suggesting transcription of an alternative isoform (fig. 1). We first characterized the alternative isoform of *PEG1/MEST* and examined expression of each of the original and novel isoforms independently. In the following discussion, the original isoform and the alternative one will be referred to as "isoform 1" and "isoform 2," respectively. To delineate the genomic structure of the *PEG1/MEST* transcription unit containing the two isoforms, finished genomic-sequence contigs of 7q31.3, deposited at the University of Washington Genome Center were surveyed and aligned against the isoform 1–specific and the isoform 2–specific cDNA sequences, by means of Sequencher software (Gene Codes). Because a mapping

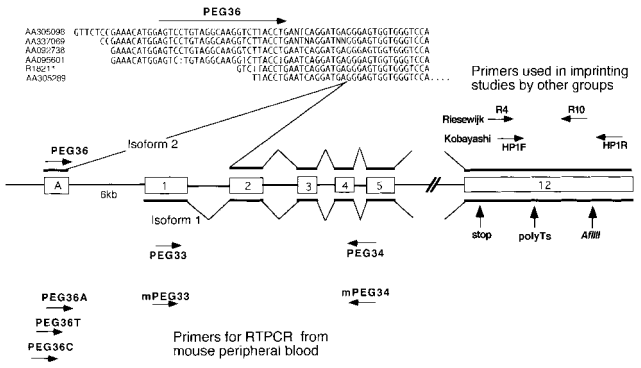

Figure 1 Alternative splicing of human *PEG1/MEST*. Arrows indicate primers used for expression studies. RT was followed by PCR amplification by use of exon specific primers. *Top,* Human EST sequences that matched with the first exon of the alternative isoform. *Middle,* Exon organization of the *PEG1/MEST* transcription unit that transcribes two isoforms. In the present study, RT-PCR was done either with PEG36 and PEG34 or with PEG33-PEG34, to detect isoform 2 and isoform 1, respectively. Poly T's and *Afl*II correspond to polymorphic sites in the 3′ UTR that have been described elsewhere. Riesewijk et al. (1997) and Kobayashi et al. (1997) used primer pair R4 and R10 and primer pair HP1F and HP1R, respectively. *Bottom,* Primers used for RT-PCR, from mouse peripheral blood.

study had indicated that genetic distance between *PEG1/MEST* and D7S649 (also known as "sWSS1203") was <1 cM (Kobayashi et al. 1997), sequence contigs flanking the PAC clone djs213 containing D7S649 were analyzed (GenBank accession number AC007938). Comparison of the genomic sequence of PAC djs201 and cDNA sequences of isoform 1 and isoform 2 revealed the following: (1) the two isoforms have distinctive first exons (the first exon of isoform 2 will be referred to as "exon A," the first exon of isoform 1 as "exon 1"); and (2) exon A is located 6 kb upstream of exon 1. Exon A contains a stop codon only 6 bases 5′ of the exon-intron boundary. It is likely that the start codon of isoform 2 is within exon 2 and that exon A comprises the 5′ UTR of isoform 2. Exon A is ⩾57 bp in length. Expression of isoform 1 and isoform 2 in lymphoblastoid cells was detected by means of reverse transcription–coupled PCR (RT-PCR) assay. Either the forward PCR primer (PEG36 [5′-agtcctgtaggcaaggtcttacctg]), based on the isoform 2–specific sequence in exon A, or the forward primer specific for the exon 1 of the isoform 1 (PEG33 [5′-atgggataacgcggccatggtg-3′]), was used with the reverse primer that anneals to the portion of the cDNA sequence shared between the two isoforms (PEG34 [5′-atagtgatgtggtctcggtttgtcactg-3′]) (fig. 1). A upd(7)mat lymphoblastoid cell line (GM11496) (Spence et al. 1988) and a paternal uniparental disomy of chromosome 7, or "upd(7)pat," lymphoblastoid cell line (Pan et al. 1998) were obtained from the National Institute of General Medical Sciences (NIGMS Coriell Cell Repositories) and from the tissue culture core at Baylor College of Medicine, respectively. The cells were cultured under standard conditions, and total RNA was extracted by means of an RNA purification kit (QIAGEN). One microgram of total RNA was used to synthesize cDNA with the Superscript preamplification system (GIBCO/BRL), and 1% of the resulting material was used for RT-PCR. The cycling conditions were 94°C for 10 min (1 cycle); 94°C for 1 min, 58°C for 1 min, and 72°C for 2 min (40 cycles); and 72°C for 10 min (1 cycle). RT-PCR revealed that the upd(7)mat cell line expressed isoform 2 but not isoform 1, whereas normal lymphocytes and the upd(7)pat cell line expressed both isoform 1 and isoform 2 (fig. 2).

In this study, we have demonstrated that (1) an alternative isoform of *PEG1/MEST* is expressed concurrently with the original isoform in adult lymphocytes and lymphoblastoid cell lines and (2) isoform 1 (the original isoform) is expressed only from the paternal allele, whereas isoform 2 (the alternative isoform) is expressed from both the paternal allele and the maternal allele. These results are discordant with the results of previous studies, which support biallelic expression of the *PEG1/MEST* in lymphocytes. In retrospect, it is understandable why the previous studies failed to identify such differential imprinting: the primers used for RT-PCR in other studies would not have allowed discrimination between the imprinted isoform and the nonimprinted isoform (fig. 1). In lymphocytes, recognition of an imprinted isoform (isoform 1) was masked by the presence of the nonimprinted form (isoform 2).

Other studies have demonstrated that, in upd(7)mat lymphocytes, only the methylated allele is present at the

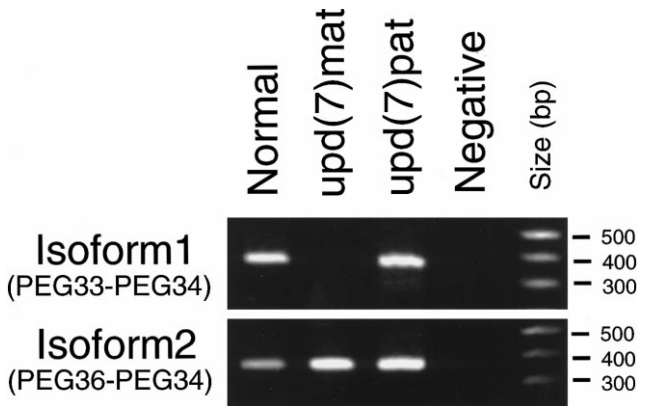

Figure 2 Expression patterns of the *PEG1/MEST* isoforms. Expression of isoforms 1 and 2 in normal, upd(7)mat, and upd(7)pat lymphoblastoid cell lines were analyzed by means of RT-PCR. Isoform 1 is expressed in upd(7)pat but not in upd(7)mat. Isoform 2 is expressed in both cell lines. Hence, isoform 1 is imprinted, whereas isoform 2 is not.

promoter of the isoform 1 of *PEG1/MEST,* whereas both methylated and unmethylated alleles are present in normal lymphocytes (Riesewijk et al. 1997). We now conclude, on the basis of findings from the present study, that parental-of-origin–specific loss of isoform 1 expression is strictly correlated with the methylation of the promoter of isoform 1. Documentation of this tight correlation validates the use of methylation analysis of *PEG1/MEST* gene in lymphocytes as a diagnostic assay for upd(7)mat.

Identification of isoform-specific imprinting illustrates several important issues with respect to imprinting studies in general. First, effort should be made to identify isoforms when one is evaluating new potentially imprinted genes. A potentially imprinted gene could be mistakenly disregarded if isoform-specific imprinting is overlooked. As shown in this study, use of the EST database can be very helpful in the identification of alternative isoforms. Second, imprinted genes that are allegedly subject to tissue-specific imprinting may need further evaluation. As shown with *PEG1/MEST* and *GNAS1,* nonimprinted or reciprocally imprinted isoforms may be expressed in tissues in which imprinting is apparently lost (Hayward et al. 1998; Peters et al. 1999). Third, the concept of leaky expression needs to be challenged. Examples of leaky expression include *p57kip2* (Reik and Maher 1997) and *IMPT1/ORCTL2* (Cooper et al. 1998; Dao et al. 1998). With respect to *PEG1/MEST,* a minimal but detectable level of expression from the maternal allele was observed in early (6–9 wk) human embryos, and this was considered to be leaky expression from the imprinted inactive maternal allele (Kobayashi et al. 1997). It is probable that these leaky transcripts from the maternal allele represent isoform 2, in light of the fact that isoform 2 is expressed as early as 8–10 wk in fetal heart (EST sequences AA092738 and AA095601). Similarly, the concept of interspecific imprinting differences may need revision. In contrast to the human *PEG1/MEST* gene, the mouse gene is not expressed from the paternal allele in lymphocytes (Riesewijk et al. 1997), nor is leaky expression from the maternal allele observed in mouse embryos (Kaneko-Ishino et al. 1995). Hence, a difference, in imprinting patterns, between mice and humans may simply reflect absence of isoform 2 in the mice. In fact, evaluation of the mouse *Peg1/Mest* UniGene cluster (Mm. 1089), consisting of 181 mouse EST sequences, revealed no evidence of alternative splicing: all 40 ESTs that contained exon 2 were flanked by exon 1 sequence, not by exon A–like sequence. Furthermore, RT-PCR, done with cDNA obtained from mouse peripheral blood by means of mouse-specific primer positioned within exon 4/5 of mouse *Peg1/Mest* gene (mPEG34 [5′-atgtggtctcggcttgtcactg-3′]), in combination with any of the four human-specific primers positioned within exon A (PEG36, PEG36A [5′-agtcctgtaggcaaggtcttacctga-3′], PEG36T [5′-gagtcctgtaggcaaggtcttacct-3′], and PEG36C [5′-gagtcctgtaggcaaggtcttacc-3′]) failed to amplify, whereas primer mPEG33 (5′-gggataatgcggccatggtg-3′), designed on the basis of mouse exon 1 sequence (the exon unique to isoform 1), yielded a specific PCR product when used with mPEG34 (fig. 1). These observations support the contention that isoform 2 may not be expressed in mouse peripheral blood and/or lymphocytes. In summary, human *PEG1/MEST* is imprinted in an isoform-specific manner rather than in a tissue-specific manner in lymphocytes.

Acknowledgments

We thank Mr. Taichi Suzuki, from Tokyo Technical College, for excellent laboratory assistance. This work was supported, in part, by a grant from the Pharmacia-Upjohn Fund for Growth & Development Research.

Kenjiro Kosaki,[1,2] Rika Kosaki,[1,3] William J. Craigen,[4] and Nobutake Matsuo[1]
[1]*Department of Pediatrics and* [2]*Pharmacia-Upjohn Fund for Growth & Development Research, Keio University School of Medicine, and* [3]*Keio University Health Center, Tokyo; and* [4]*Department of Molecular and Human Genetics, Baylor College of Medicine, Houston*

Electronic-Database Information

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

GenBank, http://www.ncbi.nlm.nih.gov/Web/Genbank
NIGMS Coriell Cell Repository, http://locus.umdnj.edu/nigms
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for Russell-Silver syndrome [MIM 180860])
UniGene, http://www.ncbi.nlm.nih.gov/UniGene/
University of Washington Genome Center, http://www.genome.washington.edu/UWGC/chr-7/c7project.htm

References

Albrecht U, Sutcliffe JS, Cattanach BM, Beechey CV (1997) Imprinted expression of the murine Angelman syndrome gene, Ube3a, in hippocampal and Purkinje neurons. Nat Genet 17:75–78

Cooper PR, Smilinich NJ, Day CD, Nowak NJ, Reid LH, Pearsall RS, Reece M, et al (1998) Divergently transcribed overlapping genes expressed in liver and kidney and located in the 11p15.5 imprinted domain. Genomics 49:38–51

Cuisset L, Le Stunff C, Dupont JM, Vasseur C, Cartigny M, Despert F, Delpech M, et al (1997) PEG1 expression in maternal uniparental disomy 7. Ann Genet 40:211–215

Dao D, Frank D, Qian N, O'Keefe D, Vosatka RJ, Walsh CP, Tycko B (1998) IMPT1, an imprinted gene similar to poly-

specific transporter and multi-drug resistance genes. Hum Mol Genet 7:597–608
Deltour L, Montagutelli X, Guenet JL, Jami J, Paldi A (1995) Tissue- and developmental stage-specific imprinting of the mouse proinsulin gene, Ins2. Dev Biol 168:686–688
Ekstroem TJ, Cui H, Li X, Ohlsson R (1995) Promoter specific IGF2 imprinting status and its plasticity during human liver development. Development 121:309–316
Ferguson-Smith AC, Cattanach BM, Barton SC, Beechey CV, Surani MA (1991) Embryological and molecular investigations of parental imprinting on mouse chromosome 7. Nature 351:667–670
Giddings SJ, King CD, Harman KW, Flood JF, Carnaghi LR (1994) Allele specific inactivation of insulin 1 and 2, in the mouse yolk sac, indicates imprinting. Nat Genet 6:310–313
Hayward BE, Moran V, Strain L, Bonthron DT (1998) Bidirectional imprinting of a single gene: GNAS1 encodes maternally, paternally, and biallelically derived proteins. Proc Natl Acad Sci USA 95:15475–15480
Kaneko-Ishino T, Kuroiwa Y, Miyoshi N, Kohda T, Suzuki R, Yokoyama M, Viville S, et al (1995) Peg1/Mest imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. Nat Genet 11:52–59
Kobayashi S, Kohda T, Miyoshi N, Kuroiwa Y, Aisaka K, Tsutsumi O, Kaneko-Ishino T, et al (1997) Human PEG1/MEST, an imprinted gene on chromosome 7. Hum Mol Genet 6:781–786
Lefebvre L, Viville S, Barton SC, Ishino F, Keverne EB, Surani MA (1998) Abnormal maternal behaviour and growth retardation associated with loss of imprinted gene Mest. Nat Genet 20:163–169
Lefebvre L, Viville S, Barton SC, Ishino F, Surani MA (1997) Genomic structure and parent-of-origin-specific methylation of Peg1. Hum Mol Genet 6:1907–1915
Pan Y, McCaskill CD, Thompson KH, Hicks J, Casey B, Shaffer LG, Craigen WJ (1998) Paternal isodisomy of chromosome 7 associated with complete situs inversus and immotile cilia. Am J Hum Genet 62:1551–1555
Peters J, Wroe SF, Wells CA, Miller HJ, Bodle D, Beechey CV, Williamson CM, et al (1999) A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2. Proc Natl Acad Sci USA 96:3830–3835
Reik W, Maher ER (1997) Imprinting in clusters: lessons from Beckwith-Wiedemann syndrome. Trends Genet 13:330–334
Riesewijk AM, Hu L, Schulz U, Tariverdian G, Hoglund P, Kere J, Ropers HH, et al (1997) Monoallelic expression of human PEG1/MEST is paralleled by parent-specific methylation in fetuses. Genomics 42:236–244
Sado T, Nakajima N, Tada M, Takagi N (1993) A novel mesoderm-specific cDNA isolated from a mouse embryonal carcinoma cell line. Dev Growth Differ 35:551–560
Spence JE, Perciaccante RG, Greig GM, Willard HF, Ledbetter DH, Hejtmancik JF, Pollack MS, et al (1988) Uniparental disomy as a mechanism for human genetic disease. Am J Hum Genet 42:217–226
Vu TH, Hoffman AR (1994) Promoter-specific imprinting of the human insulin-like growth factor-II gene. Nature 371: 714–717

Address for correspondence and reprints: Dr. Kenjiro Kosaki, Division of Medical Genetics, Department of Pediatrics, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan. E-mail: kkosaki@med.keio.ac.jp

0002-9297/2000/6601-0031$02.00

Am. J. Hum. Genet. 66:312–319, 2000

Involvement of the HLXB9 Homeobox Gene in Currarino Syndrome

To the Editor:

Anorectal malformations (ARMs) are among the most common congenital anomalies, accounting for 25% of digestive malformations that require neonatal surgery. ARMs have been found associated with sacral anomalies ~29% of the time (Rich et al. 1988). When ARMs are combined with lumbosacral anomalies, they fall into the spectrum of the caudal regression syndrome (CRS), which can also exhibit additional features such as partial or total sacrococcygeal agenesis, neural changes, and urogenital malformations (Lerone et al. 1997). The incidence of CRS is ~1 in 7,500 (Kallen et al. 1974). A detailed clinical characterization of patients affected by ARMs with partial or total sacrococcygeal agenesis revealed significant differences in the phenotypes, leading to the differentiation of five specific categories (Kalitzki 1965; Cama et al. 1996): (1) total sacral agenesis with normal or short transverse pelvic diameter and some lumbar vertebrae possibly missing (fig. 1*a* and *b*), (2) total sacral agenesis without involvement of lumbar vertebrae, (3) subtotal sacral agenesis or sacral hypodevelopment (with S1 present), (4) hemisacrum (fig. 1*c*), and (5) coccygeal agenesis.

In 1981, Guido Currarino described a form of CRS with hemisacrum (type IV sacral malformation), anorectal malformation, and presacral mass (anterior meningocele, teratoma, and/or rectal duplication) (fig. 1*d;* Currarino et al. 1981). The Currarino syndrome (CS; also called "Currarino triad") was observed to segregate in an autosomal dominant manner that often displayed phenotypic variability. As defined in the original reports, patients affected by true CS always exhibit the typical hemisacrum, with intact first sacral vertebra (sickle-shaped sacrum), which makes this specific sacral anomaly distinct to this syndrome.

Genetic studies suggested that a locus involved in normal sacral and anorectal development mapped to the terminal end (q36) of human chromosome 7 (Lynch et al. 1995; Seri et al. 1999). Mutations within the HLXB9 gene were identified in six cases, collectively grouped as

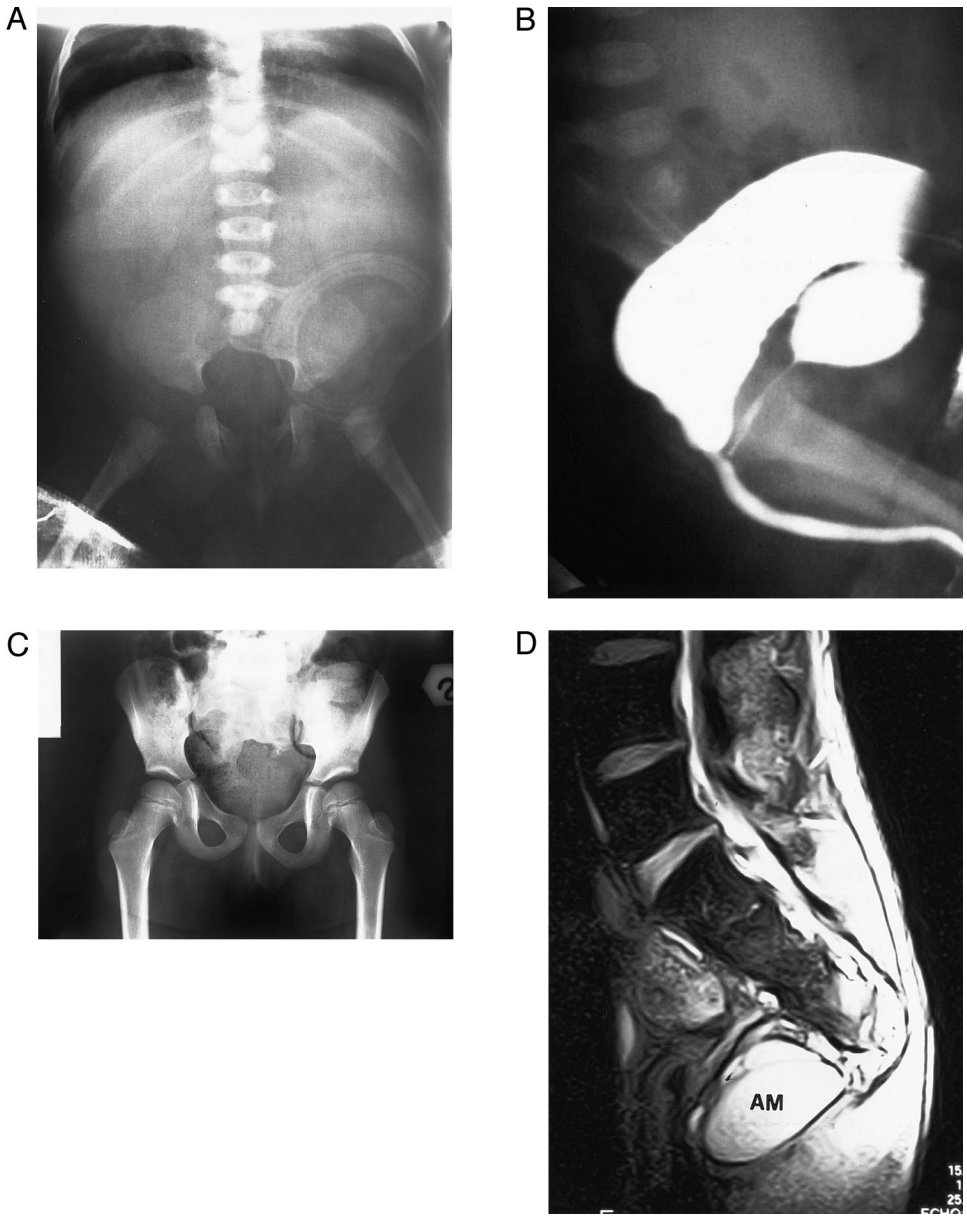

Figure 1 *a,* ARM with total sacral agenesis and L5 hypoplasia in patient 020. *b,* Distal cologram, showing presence of a rectobulbar fistula. *c* and *d,* Hemisacrum as observed in CS (*c*) and in patient 015 (*d*). The MRI shows the presence of the anterior meningocele (AM).

having dominantly inherited sacral agenesis (Ross et al. 1998). To define a more precise involvement of HLXB9 in the different phenotypic subgroupings of ARMs associated with sacral abnormalities, we screened for the presence of mutations in 27 individuals showing different sacral conditions, as described in table 1.

The HLXB9 gene contains three exons: double-gradient–denaturing gradient-gel electrophoresis (DG-DGGE) (Cremonesi et al. 1997) was performed on exons 2 and 3 (the homeodomain is coded by a portion of these two exons) (fig. 2). Patients showing an abnormal electrophoretic pattern were further analyzed by DNA sequencing (see conditions in the note to table 1). The DG-DGGE methodology yielded a high mutation-detection efficiency for exons 2 and 3, as reported elsewhere for the analysis of other genes (Cremonesi et al. 1999). Nevertheless, because of an extremely high melting profile for exon 1, it was necessary to sequence it in all 27 samples.

Analysis of the HLXB9 coding region and intron-exon

Table 1

Collection of Patients with Anorectal and Sacral Anomalies in This Study

Patient (Sex)	Clinical Diagnosis (Category)	Sacral Phenotype[a]	Associated Anomalies	HLXB9 Status/Predicted Protein Change	CGC Status
001 (F)	CS (IV)	Hemisacrum	ARM with rectoperineal fistula, presacral teratoma, tethered cord, anterior meningocele	Apparent hemizygous deletion of HLXB9	11/del
002 (F)	CS (IV)	Hemisacrum	ARM with rectoperineal fistula, anterior meningocele	384delG (exon 1)/truncated protein	11/11
011 (F)	CS (IV)	Hemisacrum	ARM with rectoperineal fistula, tethered cord, anterior meningocele, hydromyelia	R295W (exon 3/a), amino acid substitution within homeodomain	11/9
015 (M)	CS (IV)	Hemisacrum	ARM with rectoperineal fistula, anterior meningocele	Not detected	11/11
019 (M)	CS (IV)	Hemisacrum	ARM without fistula, hyposadia, Down syndrome	Not detected	11/9
059 (F)	CS (IV)	Hemisacrum	ARM with rectoperineal fistula, decreased bladder capacity	858+1G→A (exon 2)/splicing defect	11/11
060 (F)	CS (IV)	Hemisacrum	ARM with rectoperineal fistula	T248S (exon 2)/amino acid substitution within the homeodomain	11/11
066 (F)	CS (IV)	Hemisacrum	ARM with rectoperineal fistula, presacral teratoma	Not detected	11/11
069 (F)	CS (IV)	Hemisacrum	ARM, tethered cord, presacral mass, rectal duplication, bifid clitoris,lipoma of conus, holoprosencephaly	Hemizygous deletion of HLXB9	12/del
070 (M)	CS (IV)	Hemisacrum	ARM, presacral mass	Hemizygous deletion of HLXB9	11/del
003 (M)	CRS (III)	Agenesia of coccix and one sacral vertebra; SR = .43	ARM with rectobulbar fistula, lipoma of filum, hydromyelia, vertebral anomalies (L3–L4), syndactyly of second and third toes	Not detected	11/11
004 (M)	CRS (III)	Agenesia of coccix and one sacral vertebra; SR = .70	ARM with rectobulbar fistula, bilateral vesicoureteral reflux, vertebral anomalies (L2–L5)	Not detected	11/11
007 (M)	CRS (III)	SR = .55	ARM with rectoprostatic fistula, lipoma of terminal filum, hydromyelia, tethered cord, intra-atrial defect, pulmonary artery stenosis, syndactyly of the third and fourth fingers	Not detected	11/11
014 (M)	CRS (III)	CRS; SR = .45	ARM with rectovesical fistula	Not detected	11/11

(*continued*)

Table 1 Continued

Patient (Sex)	Clinical Diagnosis (Category)	Sacral Phenotype[a]	Associated Anomalies	HLXB9 Status/Predicted Protein Change	CGC Status
016 (F)	CRS (III)	Agenesia of coccix and one sacral vertebra; SR = .35	ARM with rectocloacal fistula, duplication of uterus and vagina	Not detected	11/11
018 (M)	CRS (III)	CRS	ARM with rectobulbar fistula, posterior meningocele	Not detected	11/11
024 (F)	CRS (III)	SR = .24	ARM with rectocloacal, lipoma of terminal filum, tethered cord, right ureterocele, intra-abdominal testis, bilateral congenital clubfoot	Not detected	11/10
025 (M)	CRS (III)	Agenesia of coccix and one sacral vertebra; SR = .54	ARM with rectobulbar fistula, intra-ventricular defect	Not detected	11/8
029 (F)	CRS (III)	Agenesia of coccix and three sacral vertebrae; SR = .33	ARM with rectoperineal fistula, high medullary cone (D11), vertebral fusion (L4–L5), annular pancreas, inferior limb arthrogryposis, bilateral congenital clubfoot	Not detected	11/11
030 (F)	CRS (III)	Agenesia of coccix and one sacral vertebra; SR = .58	ARM with rectocloacal fistula, tethered cord, L5 schisis, bicornate uterus	Not detected	11/11
031 (F)	CRS (III)	SR = .54	ARM with rectovestibular fistula, macrocrania	Not detected	11/11
033 (F)	CRS (III)	SR = .55	ARM with rectovestibular fistula, thickened filum terminale, tethered cord, hemangiomatosis	Not detected	11/9
035 (F)	CRS (III)	SR = .56	ARM with rectobulbar fistula, mild atrophy of cerebral cortex, hypoparathyroidism	Not detected	11/11
036 (M)	CRS (III)	Agenesia of coccix; SR = .54	ARM with rectobulbar fistula, low medullary cone (L4), open foramen ovale, right Hydronephrosis, anomalies of cervical spine and first right rib	Not detected	11/11
068 (F)	CRS (III)	SR = .27	ARM with rectocloacal fistula, monolateral renal dysgenesis	Not detected	11/11
009 (M)	Sacral agenesis (I)	Complete sacral agenesis	ARM with rectoprostatic fistula, high medullary cone (D11), hydromyelia, right multicystic kidney, left mega-ureter, right hernia and undescended testis	Not detected	11/11

(*continued*)

Table 1 Continued

Patient (Sex)	Clinical Diagnosis (Category)	Sacral Phenotype[a]	Associated Anomalies	HLXB9 Status/Predicted Protein Change	CGC Status
020 (M)	Sacral agenesis (I)	Agenesis of sacral vertebrae and one lumbar vertebra	ARM with rectobulbar fistula, bilateral congenital clubfoot, monolateral renal dysgenesis, hypospadias	Not detected	11/11

NOTE.—In the familial cases (001, 015, 059, and 060) where mutations are detected, they segregate with the phenotype. The exception is the mother of patient 060, who has the T248S change but no detectable phenotype. The following polymorphic markers were used to define the extent of the deletion in patients 069 and 070: cen-NOS3-D7S1829-GATAP6678-D7S1491-D7S798-D7S2546-AFM175yg1-D7S637-D7S2462-D7S550-SHH-D7S2465-D7S559-HLXB9-D7S2423-D7S594-tel (see Chromosome 7 Database; for a high-resolution map around SHH and HLXB9, see the work of Belloni et al. [1996] and Heus et al. [1999]). The breakpoints in patients 069 and 070 were located between markers D7S550–D7S2465 (centromeric) and D7S2423–D7S594 (telomeric) and between markers NOS3–D7S1829 (centromeric) and D7S2423–D7S594 (telomeric), respectively, indicating that HLXB9 was deleted in both patients. DG-DGGE conditions were as follows: exon 2—60%–100% urea-formamide (in all cases, 7M urea and 40% formamide represent the 100% denaturant), 6.5%–12% acrylamide, 75 V, 16 h, for primers HB9-2F (5′-TGTAGTGGTACAATCAGCAACGGGA-3′) and HB9-2R (5′-GCCCGCCCCCGCCGCCCTGCCCGCGCCCCGCGCCGCCCGCTCGCCGCCCGCCGCCCGCAAAGGTAACAGTGTCCCATGGGA-3′) (351-bp product); the PCR was 94°C for 5 min (1 cycle); 94°C for 1 min, 60°C for 45 s, and 72°C for 1 min (35 cycles); and 72°C for 10 min (1 cycle); exon 3/a—40%–90% urea, 6.5%–15% acrylamide, 50 V, 14 h, for primers HB9-3/aF (5′-GCCCGCCCCCGCCGCCCTGCCCGCGCCCCGCGCCGCCCGCTCGCCGCCCGCCGCCCCCTTCTGTTTCTCCGCTTCCTGCG-3′) and HB9-3aR (5′ CACCCTGAGGCCATTCCAGGGCCGA-3′) (266-bp product); the PCR was 94°C for 5 min (1 cycle); 94°C for 1 min, 60°C for 45 s, and 72°C for 1 min (35 cycles); and 72°C for 10 min (1 cycle); exon 3/b—30%–100% urea, 10%–15% acrylamide, 65 V, 20 h, for primers HB9-3/bF (5′CCCGCCGCCCGCCGCTCGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCCGCCCAAAAAAGGCCAAAGAGCAGG-3′) and HB9-3/bR (5′-TGCGGGCGCCGGGGCCTCCGGGAGA-3′) (446-bp product). The PCR was 94°C for 5 min (1 cycle); 94°C for 1 min, 62°C for 45 s, and 72°C for 1 min (35 cycles); and 72°C for 10 min (1 cycle). Primer pairs used for amplification and sequencing of exon 1 (with an ABI-373) were HB9-1F (5′-CCGCACACGGCCGCGTCGCCCGCCACCGGG-3′) and HB9-1R (5′-CGGCGGCGGCAGCGGCCGCTGCGCCCGGAT-3′) (439-bp product) (the PCR was 94°C for 5 min [1 cycle]; 94°C for 1 min, 60°C for 45 s, and 72°C for 1 min [35 cycles]; and 72°C for 10 min [1 cycle]) and HB9-1FaM13 (5′-TGTAAAACGACGGCCA-GTCGGCACGGGCGGCGGGCACGGGGGGCCCCA-3′) and HB9-1Rc (5′-GCAGCTCTTCCCCCGCTCGCTGGGAGCCAA-3′) (518-bp product); the PCR was 94°C for 5 min (1 cycle); 94°C for 1 min, 62°C for 45 s, and 72°C for 1 min (35 cycles); and 72°C for 10 min (1 cycle). PCR-mediated site-directed mutagenesis experiments were designed to detect the 384delG mutations by use of the following primers: HB9-1F-PSDM (5′-GCCCGCCGACCGCCTGCGCGCCGAGAGGCC-3′) and HB9-1R (179-bp product); the PCR was 94°C for 5 min (1 cycle); 94°C for 1 min, 58°C for 45 s, and 72°C for 1 min (30 cycles); and 72°C for 10 min (1 cycle). *Stu*I restriction digestion of the precipitated PCR products recognizes the presence of the restriction site only on the mutated allele (50 μl of PCR product were precipitated and digested with *Stu*I; undigested fragment, 179 bp; digested fragments, 28 and 151 bp). Analysis was on a 2% agarose/1% Nu-Sieve gel. Each PCR reaction was performed on 500 ng of genomic DNA in a 50-μl volume containing 50 pmol of each primer, 400 mmol of each dNTP, 10% dimethyl sulfoxide, 1 U of Dynazyme DNA polymerase, 10 mM TRIS-HCl (pH 8.8), 1.5 mM $MgCl_2$, 50 mM KCl, and 1.5% Triton X-100. Analysis of the CGC repeat coding for the polyalanine tract was accomplished by PCR of the region (Eppendorf *Pfu*), followed by electrophoresis on 6% polyacrylamide. Primers HB9-1Fa and HB9-1Rrep (5′-AGGGTGCAGCCCCAGCGCCAGGCC-3′) were used. One of the two primers was end-labeled with $\gamma[^{32}P]$-dATP, by means of NEB T4 kinase (in a 10-μl reaction: primer, 100 pmol; T4 kinase, 5 U; T4 buffer, 1 μl; and $\gamma[^{32}P]$-dATP, 10 mCi; the labeling was for 1 h at 37°C, followed by precipitation such that the primer was at the final concentration of 4 pmol/ml); the PCR was 94°C for 5 min (1 cycle); 94°C for 1 min, 60°C for 45 s; and 72°C for 1 min (30 cycles); 60°C for 45 s; and 72°C for 10 min (1 cycle) and was performed on 250 ng of genomic DNA in a 25-μl reaction mixture containing 4 pmol of labeled primer, 20 pmol of unlabeled primer, 400 mmol of each dNTP, 10% dimethyl sulfoxide, and 1.25 U of *Pfu,* 20 mM TRIS-HCl (pH 8), 2 mM $MgCl_2$, 10 mM KCl, 6 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, and 10 mg nuclease-free BSA/ml. Four microliters of each reaction was run on polyacrylamide gels for 4 h at 1,800 V. M13mp18 sequence was used as a length control for the PCR product that was 124 bp in the presence of the CGC_{11} repeated unit.

[a] SR = sacral ratio. This is obtained by comparison of the sacrum size with the fixed bony parameter of the pelvis, creating a ratio between different segments (Pena 1995); the normal SR is .77.

boundaries in our cohort of patients detected mutations in only those clinically characterized to have CS. Four patients (002, 011, 059, and 060) harbored, within the coding sequence, heterozygous point mutations that would be predicted to cause deleterious changes in the protein (table 1). The entire HLXB9 gene in two patients (069 and 070) and possibly a third (001; see below) was found to be deleted. For each mutation, a similar change was not observed in any of 100 normal chromosomes examined either through DG-DGGE (T248S, 858 +1G→A, R295W) or PCR-mediated site-directed mutagenesis (384delG) (Haliassos et al. 1989). DG-DGGE of exon 2 detected the presence of an abnormal pattern in 1 of 100 normal and 1 of 54 patient chromosomes examined: direct sequencing showed that the sequence variation is intronic and not involved in the splicing. Moreover, the DNA sequence of exon 1 in patient 059, already affected by the 858+1G→A mutation, contained, at position 357, a second nucleotide change, which does not cause an amino acid change (P119P).

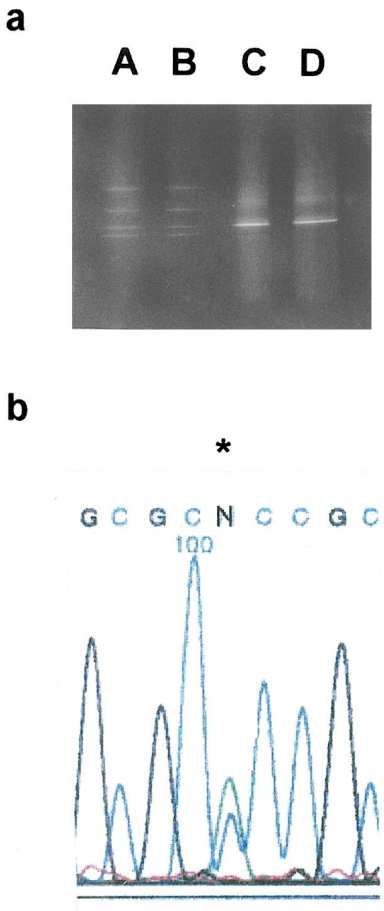

Figure 2 Example of DG-DGGE (*a*) and direct DNA sequencing (*b*). *a,* Lanes A and B, Replicated samples frm patient 060. Lanes C and D, Controls. *b,* Electrophoregram showing two peaks (marked with an asterisk [*]) at the T248S mutation in patient 060.

The missense mutations all affect the homeodomain, suggesting that an amino acid change in this region is relevant to the normal functioning of the protein. The 858+1G→A mutation affects a nucleotide in the donor splicing site, which is 100% conserved, suggesting that an abnormal protein will be obtained in this case.

We also analyzed the region of exon 1, coding for the 16-alanine stretch, since variation in polyalanine residues has been reported in other homeobox genes involved in disease (Goodman et al. 1997; Mundlos et al. 1997; Brais et al. 1998). This region, which includes a CGC repeat, was examined to test if length variation was associated with any phenotype in our collection (87 unaffected individuals, or 174 chromosomes, served as controls). We determined that the CGC_{11} allele was the most common in the general population, accounting for the 90.23% of the chromosomes analyzed. Other alleles observed were CGC_{12} (1.7%), CGC_9 (7.47%), and CGC_8 (0.6%), and these were all heterozygous changes. Only one sample (including both affected individuals and controls) revealed a homozygous change in the polyalanine tract (CGC_9/CGC_9). Although this sample was from the control collection, closer investigation revealed that the individual lacked fusion of the posterior arch of the vertebrae and had scoliosis on the left side. No association between the CGC-repeat length and the presence of sacral malformations could be established, since the alleles observed in the 27 patients were all detected in controls.

The CGC-repeat length was important in the analysis of CS sample 001, the one family, of our collection, demonstrating linkage to 7q36 (Seri et al. 1999). In this family, the CGC_9 allele has not been passed from the 2d to the 3d generation, which suggests that a microdeletion involving HLXB9 or an expansion of the allele from CGC_9 to CGC_{11} was present in two affected brothers of the 2d generation and then was transmitted to their descendants. Since an expansion would not likely be pathogenic (on the basis of our other findings), it is more likely that a microscopic deletion occurred. FISH analysis, with BACs and cosmids encompassing HLXB9, give the expected two signals on metaphase analysis (data not shown), which would suggest that the microdeletion is small. In three patients with CS (015, 019, and 066), no DNA sequence alterations were detected.

Therefore, we identified HLXB9 mutations in only those individuals diagnosed as having CS. This observation is not inconsistent with the findings of the original report (Ross et al. 1998), which described six different mutations in unrelated individuals who either were diagnosed with CS or had partial characteristics reminiscent of CS. (These patients, however, were all collectively grouped as having sacral agenesis.) Instead, it refines the involvement of HLXB9 in specific anorectal and sacral malformations. Thus, we have not been able to demonstrate a role for HLXB9 in caudal regression (categories III and V) and total sacral agenesis (categories I and II). Although it is possible that a mutation or DNA sequence variation in the noncoding region of HLXB9 could be present in these individuals, it is perhaps more likely that these malformations are due to other factors or defects in other gene(s). In three patients with CS (015, 019, and 060), no mutation could be found, which suggests genetic heterogeneity in CS or some other nongenetic determinant (since these cases are all sporadic and not familial). To our knowledge, no linkage studies of families affected exclusively by CRS or total sacral agenesis have been reported so far. The use of mouse models to study the human disease will likely be limited, since mice homozygous for a null mutation in Hlxb9 apparently fail to show any sacral defect (Harrison et al. 1999; Li et al. 1999)

In the mutation screen, we paid particular attention to the polyalanine region in exon 1, since variation in

lengths of these tracts in other homeodomain genes have been described in specific pathologies such as synpolydactyly (Mundlos et al. 1997), oculopharyngeal muscular dystrophy (Brais et al. 1998), and cleidocranial dysplasia (Goodman et al. 1997). Although we could determine that the CGC_{11} allele was the most common in the population (~90%), there was no specific correlation with any heterozygous variants (which led to either an increase or a decrease in CGC length) in the sacral conditions studied. A homozygous reduction in the length of the repeat (CGC_9/CGC_9) was identified in a single control sample, and, although this individual exhibited spinal anomalies, a clear correlation could not be established and will require further investigation.

In most ARMs and sacral-defect cases, the cause is unknown, and no familial recurrence is noted. In spite of the small number of familial cases reported, it is possible that specific genes cause defined phenotypes. In this study, we have presented evidence that the HLXB9 may be exclusively involved in CS. This observation is immediately relevant for clinical investigation and genetic screening, and the criteria and protocols to perform these studies have been outlined in this report.

Acknowledgments

This work has been supported by "Fondazione Telethon Italia" grant 297/bi, the Italian Ministry of Health, "Ricerca Finalizzata, 1996," and the Medical Research Council of Canada (MRC). We are grateful to patients and families who provided samples for our analysis. The authors would also like to acknowledge Peter Heutink, Carol Stayton, Massimiliano Agnelli, Silvia Presi, Elena Rossi, Roberta Cinti, and the PRIMM Sequencing Facility, for technical support, and Drs. Jean Michel Guys, Armando Cama, Margherita Lerone, Romeo Carrozzo, Victoria Maria Siu, and Hironao Numabe, for clinical evaluation. L.-C. T. is a Senior Scientist, and S.W.S. is a Scholar, of the MRC.

E. Belloni,[1,*] G. Martucciello,[2] D. Verderio,[1] E. Ponti,[1] M. Seri,[2] V. Jasonni,[2] M. Torre,[2] M. Ferrari,[1] L.-C. Tsui,[3] and S. W. Scherer[3]

[1]*Istituto di Ricovero e Cura a Carattere Scientifica, Hospital San Raffaele, Genetics and Molecular Diagnostic Unit, Milan;* [2]*"Giannina Gaslini" Institute, University of Genoa, Genoa; and* [3]*Department of Genetics, The Hospital for Sick Children, Toronto*

Electronic-Database Information

The URL for data in this article is as follows:

Chromosome 7 Database, The, http://www.genet.sickkids.on.ca/chromosome7

References

Belloni E, Muenke M, Roessler E, Traverso G, Siegel-Bartelt J, Frumkin A, Mitchell HF, et al (1996) Identification of *Sonic hedgehog* as a candidate gene responsible for holoprosencephaly. Nat Genet 14:353–356

Brais B, Bouchard JP, Xie YG, Rochefort DL, Chretien N, Tome FM, Lafreniere RG, et al (1998) Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nat Genet 18:164–167

Cama A, Palmieri A, Capra V, Piatelli GL, Ravegnani M, Fondelli P (1996) Multidisciplinary management of caudal regression syndrome. Eur J Pediatr Surg 6:44–45

Cremonesi L, Firpo S, Ferrari M, Righetti PG, Gelfi C (1997) Double-gradient DGGE for optimized detection of DNA point mutations. Biotechniques 22:326–330

Cremonesi L, Carrera P, Fumagalli A, Lucchiari S, Cardillo E, Ferrari M, Righetti SC, et al (1999) Validation of double gradient denaturing gradient gel electrophoresis through multigenic retrospective analysis. Clin Chem 45:35–40

Currarino G, Coln D, Votteler T (1981) Triad of anorectal, sacral, and presacral anomalies. Am J Roentgenol 137: 395–398

Goodman FR, Mundlos S, Muragaki Y, Donnai D, Giovannucci-Uzielli ML, Lapi E, Majewski F, et al (1997) Synpolydactyly phenotypes correlate with size of expansions in HOXD13 poly-alanine tract. Proc Natl Acad Sci USA 94: 7458–7463

Haliassos A, Chomel JC, Tesson L, Baudis M, Kruh J, Kaplan JC, Kitzis A (1989) Modification of enzymatically amplified DNA for the detection of point mutations. Nucleic Acids Res 17:3606

Harrison KA, Thaler J, Pfaff SL, Gu H, Kehrl JH (1999) Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice. Nat Genet 23:71–75

Heus HC, Hing A, van Baren MJ, Joosse M, Breedveld GJ, Wang JC, Scherer SW, et al (1999) A physical and transcriptional map of the preaxial polydactyly locus on chromosome 7q36. Genomics 57:342–351

Kalitzki M (1965) Congenital malformations and diabetes. Lancet 2:641–642

Kallen B, Windberg I (1974) Caudal mesoderm pattern of anomalies: from renal agenesis to sirenomelia. Teratology 9:99–112

Lerone M, Bolino A, Martucciello G (1997) The genetics of anorectal malformations: a complex matter. Semin Pediatr Surg 6:170–179

Li H, Arber S, Jessell TM, Edlund H (1999) Selective agenesis of the dorsal pancreas in mice lacking homeobox gene Hlxb9. Nat Genet 23:67–70

Lynch SA, Bond PM, Copp AJ, Kirwan WO, Nour S, Balling R, Mariman E, et al (1995) A gene for autosomal dominant sacral agenesis maps to the holoprosencephaly region at 7q36. Nat Genet 11:93–95

Mundlos S, Otto F, Mundlos C, Mulliken JB, Aylsworth AS, Albright S, Lindhout D, et al (1997) Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. Cell 89:773–779

Pena A (1995) Anorectal malformations. Semin Pediatr Surg 4:35–47

Rich AM, Brock WA, Pena A (1988) Spectrum of genitourinary malformations in patients with imperforate anus. Pediatr Surg Int 3:110–113

Ross AJ, Ruiz-Perez V, Wang Y, Hagan D-M, Scherer S, Lynch SA, Lindsay S, et al (1998) A homeobox gene, HLXB9, is the major locus for dominantly inherited sacral agenesis. Nat Genet 20:358–361

Seri M, Martucciello G, Paleari L, Bolino A, Priolo M, Salemi G, Caroli F, et al (1999) Exclusion of the *Sonic Hedgehog* gene as responsible for Currarino syndrome and anorectal malformations with sacral hypodevelopment. Hum Genet 104:108–110

* Present affiliation: Department of Experimental Oncology, Institute of Oncology, Milan.

Address for correspondence and reprints: Dr. Steve Scherer, Department of Genetics, Room 9107, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. E-mail: steve@genet.sickkids.on.ca

0002-9297/2000/6601-0032$02.00

Am. J. Hum. Genet. 66:319–326, 2000

A Novel Locus for Leber Congenital Amaurosis Maps to Chromosome 6q

To the Editor:

Leber congenital amaurosis (LCA) (MIM 204000/204100) is a clinically and genetically heterogeneous retinal disorder that occurs in infancy and is accompanied by profound visual loss, nystagmus, poor pupillary reflexes, and either a normal retina or varying degrees of atrophy and pigmentary changes (Leber 1869, 1871; François 1968). The electroretinogram (ERG) is extinguished or severely reduced (Franceschetti 1954). LCA is largely a recessive disease, although autosomal dominant pedigrees have been identified (Sorsby et al. 1960; Heckenlively 1988). To date, three genes for LCA have been identified and sequenced: retinal guanylate cyclase (*GUCY2D*) on chromosome 17p13; retinal pigment epithelium protein (*RPE65*) on chromosome 1p31; and cone-rod homeobox (*CRX*) on chromosome 19q13.3. One additional locus has been identified on chromosome 14q24 (Stockton et al. 1998). We show evidence for linkage to chromosome 6q11-16 in a multigenerational kindred of Old Order River Brethren. The disease gene maps to a 23-cM interval flanked by DNA polymorphic markers D6S1551 and D6S1694, with a maximum two-point LOD score of 3.38 (recombination fraction [θ] zero) at D6S391. Two candidate genes on chromosome 6 were screened for mutations: gamma aminobutyric acid rho1 and rho2 (*GABRR1* and *GABRR2*) at 6q14-21 (Cutting et al. 1992), and interphotoreceptor matrix proteoglycan (*IMPG1*) at 6q13-15 (Gehrig et al. 1998).

The incidence of LCA is 3 in 100,000 persons and accounts for $\geqslant$5% of all inherited retinal dystrophies (Perrault et al. 1996). Clinical and genetic heterogeneity have been demonstrated (Wardenburg 1961; Camuzat et al. 1996). The phenotype has been associated with familial juvenile nephronophthisis and cone-shaped epiphyses (Saldino-Mainzer syndrome) and with kidney disease (Senior-Loken syndrome), osteoporosis, metabolic diseases, and neurological abnormalities (Loken et al. 1961; Senior et al. 1961; Dekaban 1969; Mainzer et al. 1970; Ellis et al. 1984).

The first locus for LCA was mapped to 17p13 with the use of homozygosity mapping in consanguineous families of North African descent (Camuzat et al. 1996). Mutations in the retina-specific guanylate cyclase gene (*RETGC 1*), on chromosome 17p13, involved in phototransduction, were subsequently identified (Perrault et al. 1996). Mutations in *RPE65* on chromosome 1p31, specific to the retinal pigment epithelium involved in retinoid metabolism, were reported in patients with LCA, thus establishing a second gene (*LCA2*) for this heterogeneous disease (Marlhens et al. 1997). The photoreceptor-specific homeobox gene *CRX,* on chromosome 19q13.3, has been implicated as the third gene, since mutations were demonstrated (Freund et al. 1998). A novel locus on chromosome 14q24 (*LCA3*) was identified in consanguineous Saudi Arabian families (Stockton et al. 1998).

We studied a consanguineous family belonging to the Old Order River Brethren, a religious isolate originating in eastern Pennsylvania. The Old Order River Brethren descended from the Swiss, who emigrated to America in the 1750s in pursuit of religious freedom (Breckvill 1972). The kindred includes three affected individuals in two related sibships (fig. 1) who were initially evaluated at the Johns Hopkins Center for Hereditary Eye Diseases (JHCHED) and who are being followed annually. The patients presented with visual acuities in the order of 20/100–20/400, nystagmus, high hypermetropia, poor pupillary reflexes, and normal fundi. Progressive hypermetropia and increasing peripheral retinal mottling, of varying degree, were noted. The ERG was abolished. Review of other systems was unremarkable. We report a novel locus for LCA (*LCA5*) in this pedigree, on chromosome 6q11-16, by linkage analysis and homozygosity mapping.

Venous blood samples were obtained from 27 family members of the Old Order River Brethren community and a cheek brush sample was obtained from an infant (individual 29). Consents were obtained in accordance with regulations of the Johns Hopkins Medical Institutions' Joint Committee on Clinical Investigation. DNA was isolated from whole blood by means of the QIAamp Blood Kit (Qiagen), according to the manufacturer's in-

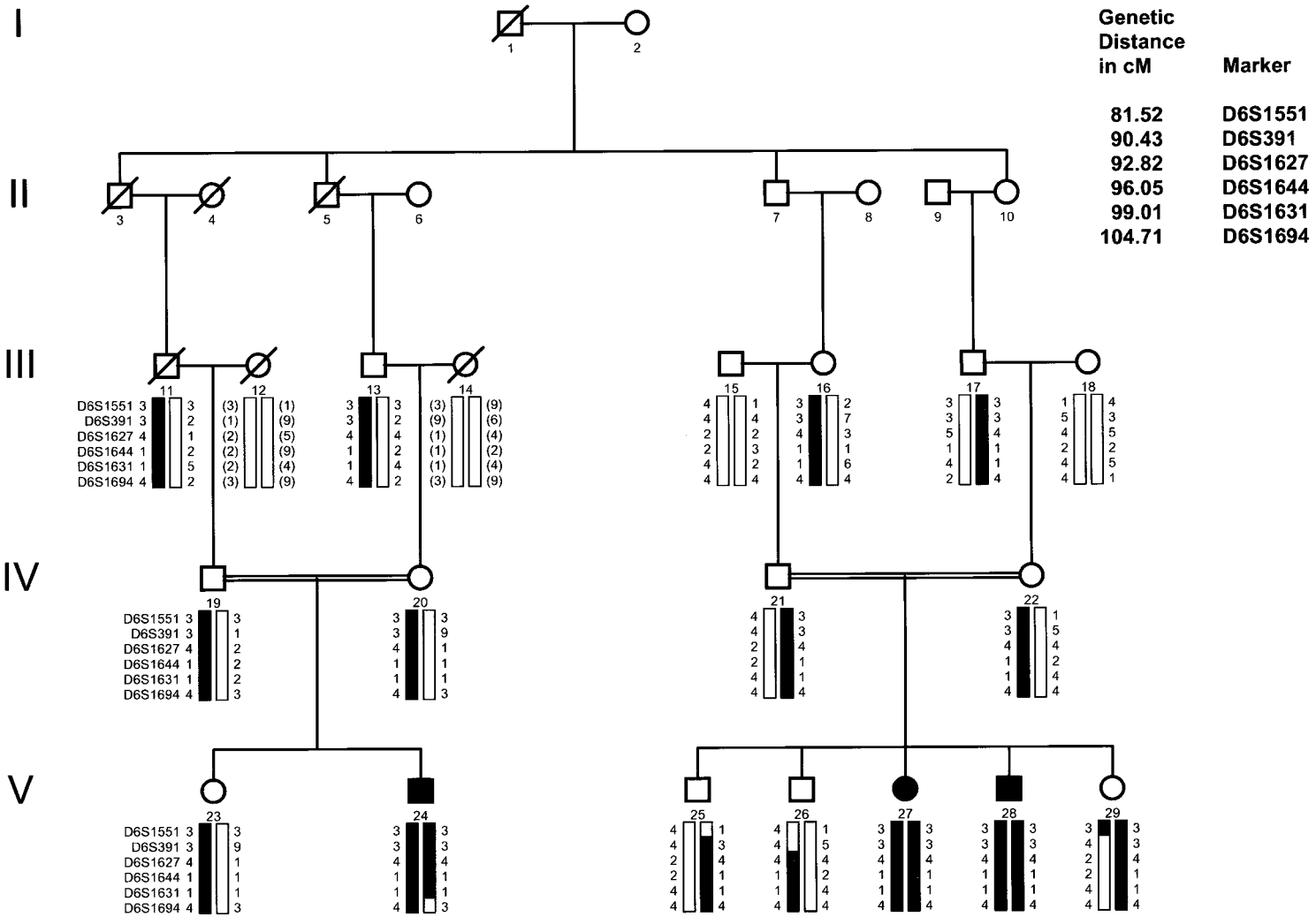

Figure 1 Pedigree of family with LCA. Blackened symbols represent affected individuals. Blackened bars indicate disease-linked haplotype. Genotype (in parentheses) is inferred. Markers are listed in order, from centromere to telomere. Critical recombination events are noted in individuals 29 and 24 at markers D6S1551 and D6S1694, respectively.

structions. The alkali method was used to obtain DNA from the single cheek sample.

Initially, the affected members and their first-degree relatives were screened to exclude linkage to the regions of the previously described genes involved in LCA on chromosomes 1, 17, and 19. The screen was then extended, by use of the whole-genome 8A multiplex version of markers spaced at 20 cM (Research Genetics). A region of homozygosity was identified on chromosome 6q. Further analysis, with additional markers in all potentially significant family members, was undertaken. Marker information was obtained from the Genome Database.

PCR-based genotyping, with fluorescent labeled markers, was performed by means of the Applied Biosystems 373 automated DNA sequencer. PCR reactions were performed in a 9600 Perkin Elmer thermocycler, and the PCR products were checked for amplification with a 3% agarose gel (Saiki et al. 1988).

The amplified PCR product was genotyped by means of the automated DNA sequencer. GENESCAN ANALYSIS 2.0.0 and GENOTYPER version 1.1 software were used, to size the PCR products and to analyze the data. Allele sizes were scored by two independent observers.

Two-point linkage analysis was performed by use of the MLINK option of the FASTLINK program, version 5.1 (Lathrop et al. 1984; Cottingham et al. 1993.) In this pedigree, LCA was analyzed as an autosomal recessive trait with complete penetrance, with an assumed allele frequency of .0032. A total of 40 microsatellite markers (Research Genetics) on chromosome 6 were analyzed, to determine the minimum region containing the new gene (Lander and Botstein 1987). Marker allele frequencies were estimated by means of the Genetic Analysis System, version 2.0 (Young), and GCONVERT (Duffy). The final LOD scores were computed by means of the allele frequencies generated by the GCONVERT program. Recombination frequencies for males and females were assumed to be equal. All inbreeding loops in the family were disconnected for computational reasons (Ott 1991) (fig. 1).

The GENEHUNTER program was used to perform multipoint linkage analysis against a fixed map of 17 informative markers, with an assumed equilibrium between marker and test loci (Kruglyak et al. 1996). These markers were selected from the original 40 microsatellite markers because they were highly polymorphic and their relative orders and map distances were well estimated in public databases. The comprehensive genetic map of the Center for Medical Genetics, Marshfield Medical Research Foundation, provided the sex-averaged genetic distances for the markers noted in figure 1. The same map provided the order for 37 markers. The remaining three markers were placed by means of the Genome Location Database. Subsequently, the genetic framework map from the Center for Medical Genetics, Marshfield Medical Research Foundation, reduced the 37 markers to 23, indicating that some markers occurred at identical positions. Thus, in the analyses in which multiple markers had identical positions, the markers with reduced information content were dropped in favor of those with better information content.

Multipoint LOD scores were computed with the same model described above for two-point analysis. Multipoint nonparametric linkage (NPL) scores were also computed with the use of only the affected individuals. Because of the inherent limitation on pedigree size in the GENEHUNTER program, the large pedigree was trimmed and broken into two separate units, accounting for some potential loss of power in the analysis. The data were examined for regions of allelic homozygosity in the affected individuals (Dib et al. 1996). Haplotype analysis was used to further define the interval containing the disease locus (Lathrop et al. 1985).

The following retina-specific genes on chromosome 6 were evaluated for the presence of disease-causing mutations: the *GABRR1* and *GABRR2* genes on chromosome 6q14.1-21 (Cutting et al. 1991, 1992) and the *IMPG1* gene on chromosome 6q13-15 (Gehrig et al. 1998) (fig. 2).

GABRR1 and *GABRR2* are assumed to have arisen by gene duplication and share a 50% homology with each other. GABA is a neuroinhibitory transmitter mediating fast synaptic inhibition by activating chloride channels. *GABRR1* is expressed largely in the retina; *GABRR2* is expressed primarily in the brain. *GABRR1* expression in the developing retina suggests its possible role as a candidate gene. In exons 1 and 4, polymorphic changes were identified. This did not change in the amino acid. These polymorphic changes were also identified in the normal population. No sequence changes were noted in *GABRR2*.

IMPG1, a novel gene encoding a major proteoglycan of the interphotoreceptor matrix, is expressed in the retina by both rods and cones and maps to 6q13-15; it was considered a further candidate gene for LCA (Gehrig et al. 1998). No significant changes were noted after the 17 coding exons of this gene were sequenced.

Linkage of LCA in the Old Order River Brethren was found at chromosome 6q11-16, supported by statistically significant two-point LOD scores with maximum LOD score (Z_{max}) 3.38 ($\theta = 0$) at D6S391 (table 1). Haplotype analysis of recombination events localizes the disease locus to a region of 23 cM, flanked by D6S1551 and D6S1694, thus identifying a new locus for LCA. Critical recombinant events were observed at marker D6S1551 in individual 29, who is unaffected, and at marker D6S1694 in individual 24, who is affected, defining the centromeric and telomeric boundaries, respectively. A common haplotype covers the region in all

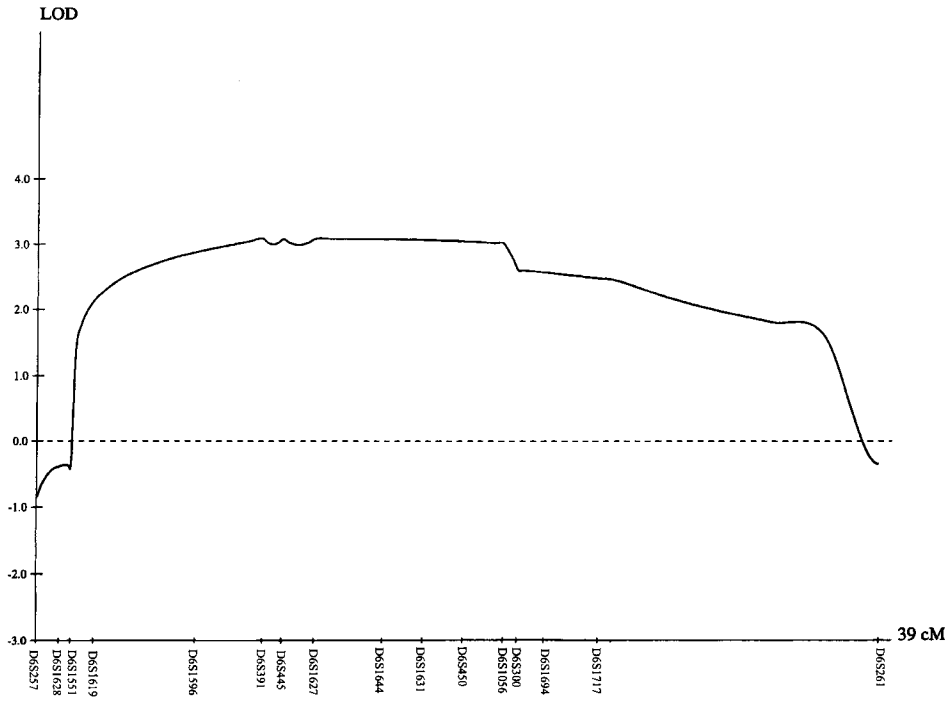

Figure 2 Results of multipoint LOD score analysis using the GENEHUNTER program. The centromere is located toward the left of the graph. Maximum LOD scores were obtained between markers D6S391 and D6S450.

the affected individuals (fig. 1). Homozygosity of other highly informative markers across the candidate region was noted.

The maximum multipoint LOD score was 3.10 between D6S391 and D6S450, a 9.5-cM interval (fig. 2). The multipoint NPL score was significant, with $P <$ .004 for a 23-cM region. Sequencing of candidate genes *GABRR1* and *GABRR2* and *IMPG1* revealed polymorphic changes only.

LCA in the Old Order River Brethren, a highly inbred community, maps to a 23-cM interval on chromosome 6q11-16, as defined by linkage analysis and homozygosity mapping. Since this population is genetically isolated, and since LCA is quite rare, we presume that a single common ancestor was a carrier for this recessive trait (Lander and Botstein 1987). The large size of the region of homozygosity in the family and the history of migration indicate the recency of the mutation in the population (fig. 1).

LCA in this pedigree was not associated with multisystem abnormalities. Renal function remains normal. Neurological and hepatic function were within normal limits. The patients are of normal stature and intelligence. Neither photophobia nor photoattraction was reported in infancy, although pressing on the globes (the digito-ocular phenomenon of Franceschetti-Bamatter) played a prominent part in childhood behavior (Franceschetti 1954). Visual dysfunction, nystagmus, and the digito-ocular phenomenon were noticed in early infancy. A high hyperopic refractive correction was noted in all the patients (Wagner et al. 1985). Ophthalmoscopic examination in infancy revealed normal fundi, but in childhood, attenuated retinal vasculature with a varying degree of pigmentary changes was noticed. Electroretinography showed a markedly reduced response in the affected individuals. Vision has been stable in all affected members of the family who have been followed clinically at JHCHED.

Genetic studies have identified a large region on chromosome 6q responsible for several retinal dystrophies (Small et al. 1992, 1993, 1997; Stone et al. 1994; Kelsell et al. 1995, 1998; Sauer et al. 1997; Griesinger et al. 1998; Rabb et al. 1998; Ruiz et al. 1998). *LCA5* lies in the overlapping region of autosomal recessive RP at 6cen-q16 (Ruiz et al. 1998), progressive bifocal chorioretinal dystrophy (*PBCRA*) at 6q12-21 (Kelsell et al. 1995), North Carolina macular dystrophy at 6q14-16.2 (Small et al. 1993), Stargardt-like dominant macular degeneration (*STGD3*) at 6q13 (Stone et al. 1994; Griesinger et al. 1998), and dominant cone-rod dystrophy (*CORD7*) at 6q13-15 (Kelsell et al. 1998) (fig. 3). The occurrence of multiple loci so closely spaced in the genome could indicate the presence of a number of retinal genes in continuum, since the phenotype of all these

retinal dystrophies, their ophthalmologic appearance, age at onset, and the extent and pattern of visual loss are varied. On the other hand, like the *ABCR* gene mutations that cause autosomal recessive retinitis pigmentosa (Martinez-Mir et al. 1997, 1998), juvenile and late-onset fundus flavimaculatus (Allikmets et al. 1997*a*), cone-rod dystrophy (Cremers et al. 1998), age-related macular disease, and recessive Stargardt disease (Kaplan et al. 1993; Gerber et al. 1995, 1998; Allikmets et al. 1997*b*), leading to distinct phenotypes (Lewis et al. 1999), it is possible that a single large gene in the proximal centromeric portion of the long arm of chromosome 6 could cause a myriad of retinal dystrophies, LCA being the most severe.

Perhaps *LCA5* is allelic with *STGD3, RP25, CORD7, MCDR,* and *PBCRA*. It is conceivable that mutations in different sites cause different structural alterations in the predicted protein, predisposing to varying phenotypes (Rozet et al. 1998).

The three other genes causing LCA are known to cause other phenotypically varied retinal dystrophies as well, raising the possibility of a similar situation in the *LCA5* gene. *GUCY2D* mutations (*LCA1*) have been identified in autosomal dominant cone-rod dystrophy (Kelsell et al. 1998), although *RPE65* (*LCA2*) mutations cause autosomal recessive retinitis pigmentosa as well as LCA (Morimura et al. 1998). Mutations in the cone-rod homeobox gene are now known to cause autosomal dominant cone-rod dystrophy, LCA, and late-onset dominant retinitis pigmentosa (Sohocki et al. 1998). It is currently possible to identify mutations of the known LCA genes in less than one-third of the patients with LCA (Dharmaraj et al. 1999). The isolation of another locus for this retinal disorder, *LCA5,* on chromosome 6q, will account for an additional proportion of patients with

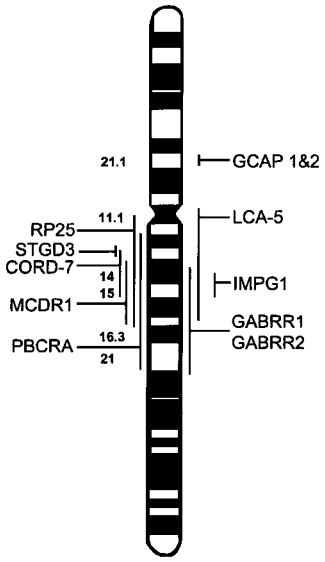

Figure 3 Chromosome 6 ideogram, showing the location of candidate genes screened and retinal disease loci in the region.

Table 1

Two-Point LOD Scores for Linkage between LCA and Chromosome 6 Markers

	LOD Score at θ =								
Order	.00	.01	.50	.10	.20	.30	.40	θ_{max}	Z_{max}
D6S257	−1.33	−.17	.67	.92	.93	.71	.39	.145	.97
D6S1628	−1.24	1.00	1.49	1.52	1.29	.93	.49	.084	1.53
D6S1658	−1.23	−.63	−.15	.04	.19	.18	.11	.239	.20
D6S430	−1.17	1.41	1.85	1.82	1.46	1.00	.51	.066	1.86
D6S1551	−1.32	.27	.82	.94	.88	.67	.37	.117	.95
D6S1619	1.20	1.17	1.03	.87	.59	.36	.17	.001	1.20
D6S1596	1.23	1.20	1.10	.96	.71	.47	.23	.001	1.23
D6S391	3.38	3.30	3.00	2.62	1.87	1.16	.55	.001	3.38
D6S1707	3.15	3.07	2.76	2.37	1.63	.98	.47	.001	3.15
D6S251	3.22	3.14	2.83	2.45	1.71	1.06	.50	.001	3.22
D6S445	2.97	2.89	2.59	2.23	1.56	.98	.48	.001	2.97
D6S1627	2.33	2.28	2.07	1.82	1.33	.87	.43	.001	2.33
D6S1644	2.20	2.15	1.93	1.66	1.17	.74	.35	.001	2.20
D6S1631	2.70	2.63	2.38	2.07	1.47	.93	.45	.001	2.70
D6S450	2.28	2.23	2.01	1.74	1.24	.79	.38	.001	2.28
D6S1056	1.47	1.42	1.22	1.02	.70	.46	.23	.001	1.47
D6S300	1.24	1.33	1.43	1.36	1.07	.71	.35	.050	1.43
D6S1716	.08	.10	.13	.14	.13	.11	.08	.090	.14
D6S1694	−.40	−.31	−.14	−.07	−.07	−.07	−.04	.848	.19
D6S1717	−.73	−.74	−.77	−.77	−.65	−.43	−.20	.827	.39
D6S261	−1.18	−.60	−.10	.10	.23	.21	.13	.231	.23

an identifiable gene mutation. Recruitment of additional families with LCA to further narrow the critical region is under way, and candidate gene analysis continues.

Acknowledgments

We are grateful to the patients and their families for their cooperation and participation in the study. We would like to acknowledge the help of Garry Cutting, Marina Kniazeva, Joan Bailey-Wilson, Reza Vagefi, Naba Bora, Carrie Gruver, Kang Zhang, Suzanne M. Leal, Mary Anderson, Gregory Leppert, Karen A. Klima, and Pamela Maskell. This work was supported in part by grants from the Foundation for Retinal Research, The Grousbeck Foundation, and The Edel and Krieble Funds of The Johns Hopkins Center for Hereditary Eye Diseases.

SHAROLA DHARMARAJ,[1] YINGYING LI,[2] JOHANE M. ROBITAILLE,[2] EDUARDO SILVA,[1] DANPING ZHU,[1] THOMAS N. MITCHELL,[1] LARA P. MALTBY,[3] AGNES B. BAFFOE-BONNIE,[4,5] AND IRENE H. MAUMENEE[1]

[1]*The Johns Hopkins Center for Hereditary Eye Diseases, The Wilmer Eye Institute, The Johns Hopkins Medical Institutions, Baltimore;* [2]*IWK-Grace Health Centre, Dalhousie University, Halifax;* [3]*Women's and Children's Health, North Carolina Department of Health and Human Services, Wilmington;* [4]*Fox Chase Cancer Center, Philadelphia; and* [5]*Division of Statistical Genetics, National Human Genome Research Institute, National Institutes of Health, Bethesda*

Electronic-Database Information

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

Center for Medical Genetics, Marshfield Medical Research Foundation, http://marshmed.org/genetics/

Duffy LD (1995) Sib-pair program, http://www.qimr.edu.au/davidD/davidd.html

Genome Database, http://www.gdb.org/ (for primer information)

Genome Location Database, http://cedar.genetics.soton.ac.uk/pub/PROGRAMS/ldb (for placing markers)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for LCA [MIM 204000/204100])

Research Genetics, http://www.resgen.com/

Young A (1995) Genetic Analysis System, version 2.0, http://users.ox.ac.uk/~ayoung/gas.html

References

Allikmets R, Shroyer NF, Singh N, Seddon JM, Lewis RA, Bernstein PS, Peiffer A, et al (1997*a*) Mutation of the Stargardt disease gene (*ABCR*) in age-related macular degeneration. Science 277:1805–1807

Allikmets R, Singh N, Sun H, Shroyer NF, Hutchinson A, Chidambaram A, Gerrad B, et al (1997*b*) A photoreceptor cell-specific ATP-binding transporter gene (*ABCR*) is mutated in recessive Stargardt macular dystrophy. Nat Genet 15:236–246

Breckvill LT (1972) History Old Order River Brethren. Breckvill and Stickler, Pennsylvania

Camuzat A, Rozet JM, Dollfus H, Gerber S, Perrault I, Weissenbach J, Munnich A, et al (1996) Evidence of genetic heterogeneity of Leber's congenital amaurosis (LCA) and mapping of *LCA1* to chromosome 17p13. Hum Genet 97: 798–801

Cottingham RW Jr, Idury RM, Schaffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53: 252–263

Cremers FP, van de Pol DJ, Van Driel M, den Hollander AI, van Haren FJ, Knoers NV, Tijmes N, et al (1998) Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene *ABCR*. Hum Mol Genet 7:355–362

Cutting GR, Curristin S, Zoghbi H, O'Hara B, Seldin MF, Uhl GR (1992) Identification of a putative gamma-aminobutyric acid receptor subunit rho 2 cDNA and localization of the genes rho 2 (*GABRR2*) and rho 1 (*GABRR1*) to human chromosome 6q14-q21 and mouse chromosome 4. Genomics 12:801–806

Cutting GR, Lu L, O'Hara BF, Kasch LM, Montrose-Rafizadeh C, Donovan DM, Shimada S, et al (1991) Cloning of the gamma-aminobutyric acid *GABA RHO 1* cDNA: a receptor subunit highly expressed in the retina. Proc Natl Acad Sci USA 88:2673–2677

Dekaban AS (1969) Hereditary syndrome of congenital retinal blindness (Leber), polycystic kidneys and maldevelopment of the brain. Am J Ophthalmol 68:1029–1037

Dharmaraj S, Silva E, Li YY, Loyer M, Koenekoop RK, Maumenee IH (1999) Mutational analysis in one hundred consecutive patients with Leber congenital amaurosis. Invest Ophthalmol Vis Sci 40:A2983

Dib C, Faune S, Fizames C, Samson D, Dvout N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5262 microsatellites. Nature 380:152–154

Ellis DS, Heckenlively JR, Martin CL, Lachman RS, Sakati NA, Rimoin DL (1984) Leber's congenital amaurosis associated with familial juvenile nephronophthisis and cone-shaped epiphyses of the hands (the Saldino-Mainzer syndrome). Am J Ophthalmol 97:233–239

Franceschetti A, Dieterle P (1954) L'importance diagnostique de l'electrorétinogramme dans les dégénérescences tapétorétinennes avec rétrécissement du champ visuel et héméralopie. Conf Neurol 14:184–186

François J (1968) Leber's congenital tapeto-retinal degeneration. Int Ophthalmol Clin 8:929–947

Freund CL, Wang QL, Chen S, Muskat B, Wiles CD, Sheffield VC, Jacobson SG, et al (1998) De novo mutations in the *CRX* homeobox gene associated with Leber congenital amaurosis. Nat Genet 18:311–312

Gehrig A, Felbor U, Kelsell R, Hunt DM, Maumenee IH, Weber BHF (1998) Assessment of the interphotoreceptor matrix proteoglycan-1 (*IMPG1*) localized to 6q13-q15 in

autosomal dominant Stargardt-like disease (ADSTGD), progressive bifocal chorioretinal atrophy (PBCRA), and North Carolina macular dystrophy (MCDR1). J Med Genet 35: 641–645

Gerber S, Rozet JM, Bonneau D, Souied E, Camuzat A, Dufier JL, Amalric P, et al (1995) A gene for late-onset fundus flavimaculatus with macular dystrophy maps to chromosome 1p13. Am J Hum Genet 56:396–399

Gerber S, Rozet JM, van de Pol TJ, Hoyng CB, Munnich A, Blankenagel A, Kaplan J, et al (1998) Complete exon-intron structure of the retina-specific ATP binding transporter gene (*ABCR*) allows the identification of novel mutations underlying Stargardt disease. Genomics 48:139–142

Griesinger IB, Sieving PA, Chandrasekharappa SC, Ayyagari R (1998) Macular degeneration with highly variable phenotype localized to chromosome 6q. Am J Hum Genet 63: A30

Heckenlively JR (1988) Retinitis pigmentosa. Philadelphia, Lippincott, pp 125–149

Kaplan J, Gerber S, Larget-Piet D, Rozet JM, Dollfus H, Dufier JL, Odent S, et al (1993) A gene for Stargardt's disease (fundus flavimaculatus) maps to the short arm of chromosome 1. Nat Genet 5:308–311

Kelsell RE, Godley BF, Evans K, et al (1995) Localization of the gene for progressive bifocal chorioretinal atrophy (PBCRA) to chromosome 6q. Hum Mol Genet 4:1653–1656

Kelsell RE, Gregory-Evans K, Gregory-Evans CY, Holder GE, Jay MR, Weber BH, Moore AT, et al (1998) Localization of a gene (*CORD7*) for a dominant cone-rod dystrophy to chromosome 6q. Am J Hum Genet 63:274–279

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

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

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

——— (1985) Multipoint linkage analysis in humans: detection of linkage and estimation of recombination. Am J Hum Genet 37:482–498

Leber T (1869) Über retinitis pigmentosa und angeborene amaurose. Albrecht von Graefe's Arch Klin Exp Ophthalmol 15:1–25

——— (1871) Über anormale formen der retinitis pigmentosa. Arch für Ophthalmol 17:314–341

Lewis RA, Shroyer NF, Singh N, Allikmets R, Hutchinson A, Li Y, Lupski JR, et al (1999) Genotype/phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene *ABCR*, in Stargardt disease. Am J Hum Genet 64:422–434

Loken AC, Hanssen O, Halvolsen S, Jolster NB (1961) Hereditary renal dysplasia and blindness. Acta Paediatr 50: 177-194

Mainzer F, Saldino RM, Ozonoff MB, Minagi H (1970) Familial nephropathy associated with retinitis pigmentosa, cerebellar ataxia and skeletal abnormalities. Am J Med 49: 556–562

Marlhens F, Bareil C, Griffoin JM, Zrenner E, Amalric P, Eliaou C, Liu SY, et al (1997) Mutations in *RPE65* cause Leber's congenital amaurosis. Nat Genet 17:139–141

Martinez-Mir A, Bayes M, Vilageliu L, Grinberg D, Ayuso C, Del Rio T, Garcia-Sandoval B, et al (1997) A new locus for autosomal recessive retinitis pigmentosa (*RP19*) maps to 1p13-1p21. Genomics 40:142–146

Martinez-Mir A, Palona E, Allikmets R, Ayuso C, Del Rio T, Dean M, Vilageliu L, et al (1998) Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene *ABCR*. Nat Genet 18:11–12

Morimura H, Fishman GA, Grover SA, Fulton AB, Berson EI, Dryja TP (1998) Mutations in the *RPE65* gene in patients with autosomal recessive retinitis pigmentosa or Leber congenital amaurosis. Proc Natl Acad Sci USA 95:3088–3093

Ott J (ed) (1991) Analysis of human genetic linkage. 2d ed. Johns Hopkins University Press, Baltimore

Perrault I, Rozet JM, Calvas P, Gerber S, Camuzat A, Dollfus H, Chatelin S, et al (1996) Retinal-specific guanylate cyclase gene mutations in Leber's congenital amaurosis. Nat Genet 14:461–466

Rabb MF, Mullen L, Yelchits S, Udar N, Small KW (1998) A North Carolina macular dystrophy phenotype in a Belizean family maps to the MCDR1 locus. Am J Ophthalmol 125: 502–508

Rozet JM, Gerber S, Souied E, Perrault I, Chatelin S, Ghazi I, Leowski C, et al (1998) Spectrum of *ABCR* gene mutations in autosomal recessive macular dystrophies. Eur J Hum Genet 6:291–295

Ruiz A, Borrego S, Marcos I, Antinolo G (1998) A major locus for autosomal recessive retinitis pigmentosa on 6q, determined by homozygosity mapping of chromosomal regions that contain gamma-aminobutyric acid-receptor clusters. Am J Hum Genet 62:1452–1459

Saiki RK, Gelfand DH, Stoffel S, Sharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Sauer CG, Schworm HD, Ulbig M, Blankenagel A, Rohrshneider K, Pauleikhoff D, Grimm T, et al (1997) An ancestral core haplotype defines the critical region harbouring the North Carolina macular dystrophy gene (*MCDR1*). J Med Genet 34:961–966

Senior B, Friedman AI, Brando JL (1961) Juvenile familial nephropathy with tapetoretinal degeneration. Am J Ophthalmol 52:625–633

Small KW, Puech B, Mullen L, Yelchits S (1997) North Carolina macular dystrophy phenotype in France maps to the *MCDR1* locus. Mol Vis 3:1–6

Small KW, Weber J, Roses A, Pericak-Vance P (1993) North Carolina macular dystrophy *MCDR1:* a review and refined mapping to 6q14-q16.2. Ophthalmic Paediatr Genet 14: 143–150

Small KW, Weber JL, Roses A, Lennon F, Vance JM, Pericak-Vance MA (1992) North Carolina macular dystrophy is assigned to chromosome 6. Genomics 13:681–685

Sohocki MM, Sullivan LS, Mintz-Hittner HA, Birch D, Heckenlively JR, Freund CL, McInnes RR, et al (1998) A range of clinical phenotypes associated with CRX, a photoreceptor transcription-factor gene. Am J Hum Genet 63:1307–1315

Sorsby A, Williams CE (1960) Retinal aplasia as a clinical entity. Br Med J 1:293–297

Stockton DW, Lewis RA, Abboud EB, Rejhi-Manzer AA, Anderson KH, Lupski JR (1998) A novel locus for Leber congenital amaurosis on chromosome 14q24. Hum Genet 103: 328–333

Stone EM, Nichols BE, Kimura AE, Weingeist TA, Drack A, Sheffield VC (1994) Clinical features of a Stargardt-like dominant progressive macular dystrophy with genetic linkage to chromosome 6q. Arch Ophthalmol 112:765–772

Waardenburg PJ (1961) Congenital and early infantile retinal dysfunction (high-graded) amblyopia and amaurosis Leber. In: Waardenburg PJ, Franceschetti A, Klein D (eds) Genetics and ophthalmology. 1st ed. Blackwell Scientific, Oxford, pp 1567–1581

Wagner RS, Caputo AR, Nelson LB, Zanoni D (1985) High hyperopia in Leber's congenital amaurosis. Arch Ophthalmol 103:1507–1509

Address for correspondence and reprints: Dr. Sharola Dharmaraj, Maumenee Building, Suite 517, The Johns Hopkins Center for Hereditary Eye Diseases, 600 North Wolfe Street, Baltimore, MD 21287-9237. E-mail: sdharmaraj@jhmi.edu

0002-9297/2000/6601-0033$02.00

Am. J. Hum. Genet. 66:326–330, 2000

A Gene for an Autosomal Dominant Scleroatrophic Syndrome Predisposing to Skin Cancer (Huriez Syndrome) Maps to Chromosome 4q23

To the Editor:

Huriez syndrome (MIM 181600), also referred to as "sclerotylosis," is an autosomal dominant genodermatosis, characterized by the triad of congenital scleroatrophy of the distal extremities, palmoplantar keratoderma (PPK), and hypoplastic nail changes, that was first described in two large pedigrees from northern France (Huriez et al. 1968). Several additional families have since been described (Lambert et al. 1977; Fischer 1978; Shaw et al. 1978; Hamm et al. 1996; Kavanagh et al. 1997). The development of aggressive squamous cell carcinoma (SCC) of the affected skin is a distinctive feature of the syndrome, occurring in ~15% of affected individuals. SCC in Huriez syndrome is characterized by early onset, mostly in the third to fourth decade of life, and by early metastasis formation (Hamm et al. 1996). The pathogenetic mechanism of tumorigenesis in Huriez syndrome is unknown.

Linkage to the MN–blood-type locus on chromosome 4q28-q31 was reported initially and has subsequently been refuted (Delaporte et al. 1995; Kavanagh et al. 1997). We therefore embarked on a linkage analysis for Huriez syndrome with highly polymorphic microsatellite markers, beginning on chromosome 4. After informed consent was obtained, 22 affected and 35 unaffected members of one of the families first described by Huriez (family A) and a second family originating from the same region of northern France (family B) were included in the analysis. We calculated two-point LOD scores between each marker locus and Huriez syndrome, under the assumption of autosomal dominant inheritance with complete penetrance, a frequency of .0005 for the disease allele, and equal allele frequencies for each marker allele, using the LINKAGE version 5.21 software (Lathrop et al. 1984). With reference to the Human Gene Map (Schuler et al. 1996), we identified marker D4S424 within 2 cM of the glycophorin A gene (GYPA), which is the erythrocyte membrane protein that encodes the MN–blood-group receptors. D4S424 yielded a LOD score of -10.7 at recombination fraction (θ) 0.01. In the families under investigation, the MN–blood-type locus was therefore excluded as a candidate region for Huriez syndrome. Our finding is consistent with the exclusion of the MN locus in the first English family with Huriez syndrome (Kavanagh et al. 1997). In contrast, marker D4S1560 gave evidence for linkage, with a LOD score of 4.4 at $\theta = 0$. Subsequent analysis of additional flanking markers confirmed localization of the Huriez locus to this region, with the highest LOD score (Z_{max}) of 12.22 at $\theta = 0$, with D4S2380 (table 1). For fine mapping of the candidate region, additional microsatellite markers were selected and mapped on the high-resolution Stanford Human Genome Center TNG radiation hybrid panel. Relative distances in centirays (fig. 1) were determined with reference to the Stanford G3 panel (Stewart et al. 1997).

Under the more conservative assumption of incomplete penetrance, the candidate interval is defined by two recombination events in one affected individual (family A, member 6.1). Here, the 17-cM region is delimited centromerically by D4S395 and distally by D4S411. Under the assumption of complete penetrance, two additional recombination events in two unaffected probands further limit the Huriez locus (fig. 2): one in family A (member 4.11), localizing the gene telomeric to D4S1544, and the other in family B (member 2.5), localizing the gene centromeric to D4S2966. The Huriez locus is thus confined to an 8-cM region between D4S1544 and D4S2966. Haplotypes were constructed for both families by use of 16 markers between D4S2963 and D4S1564. Comparison between the disease alleles of the two families revealed three adjacent markers identical by state (IBS) (fig. 1). For two of these markers, D4S2973 and D4S1559, the shared allele is the most common allele found among all individuals, with relative frequencies of .87 and .58, respectively. There is no known relationship between the two families investigated. However, in view of the rarity of the condition

Table 1

Combined Pairwise LOD Scores for Families A and B between Huriez Syndrome and Markers on Chromosome 4q

Marker	cM from ptel	LOD Score at θ = .0	.01	.05	.1	.2	.3	.4	Z_{max}	θ_{max}
D4S1544	97.9	−∞	6.36	6.55	6.16	4.93	3.35	1.52	6.55	.05
D4S414	99.2	6.99	6.87	6.39	5.76	4.40	2.97	1.26	6.87	.0
D4S2909	100.0	8.41	8.27	7.66	6.87	5.17	3.29	1.35	8.41	.0
D4S2380	101.0	12.22	12.02	11.19	10.10	7.75	5.16	2.39	12.22	.0
D4S2973	103.1	.86	.85	.78	.70	.51	.31	.13	.86	.0
D4S1559	103.1	3.69	3.59	3.21	2.76	1.91	1.08	.32	3.69	.0
D4S1578	103.1	6.67	6.55	6.04	5.40	4.08	2.71	1.29	6.67	.0
D4S1560	103.1	4.40	4.32	3.99	3.56	2.64	1.63	.60	4.32	.0
D4S2986	104.6	9.96	9.81	9.19	8.36	6.53	4.45	2.12	9.81	.0
D4S2966	107.8	−∞	8.48	8.51	7.93	6.30	4.32	2.09	8.51	.01
D4S424	146.4	−∞	−10.73	−3.60	−1.02	.73	.92	.40	.92	.3

Note.—Marker coordinates are noted according to the final Généthon linkage map (Dib et al. 1996). The coordinate of D4S2380 is noted with reference to the Marshfield Comprehensive Human Genetic Maps (Broman et al. 1998). D4S424 maps within 2 cM of the glycophorin A gene (GYPA, MN–blood-group receptors).

and the fact that both families originate from neighboring villages in northern France, we inferred that the IBS status of three adjacent markers in the candidate region reflects the existence of a common founder haplotype. If the common haplotype is verified, this would indicate that the Huriez gene is located in a 3.1-cM interval between D4S2909 and D4S1578. Further markers must be investigated to substantiate this finding, although we have exhausted all known microsatellite markers in this region. In both families, Huriez syndrome is linked to markers on chromosome 4q some 30 cM centromeric of D4S424. Epidermal growth factor (EGF) was a potential candidate gene that has been mapped to chromosome 4q21-24 on a human-rodent somatic cell hybrid panel (Brissenden et al. 1984). EGF induces cellular proliferation and differentiation of various epidermal and epithelial tissues and is a potent mitogen (Carpenter and Cohen 1979). Increased levels of the EGF receptor are associated with malignant transformation of squamous cells (Lee et al. 1997) and are observed in SCC of the skin (Springer and Robinson 1991), esophagus (Yano et al. 1991), head and neck (Lee et al. 1997), cervix (Kim et al. 1996), and lung (Pfeiffer et al. 1998). According to the gene map of the human genome (Schuler et al. 1996), the EGF gene is located between D4S411 and D4S1564. Haplotype analysis for both families A and B revealed recombination events in individuals 6.1 and 2.5, respectively, that exclude EGF as a candidate gene for Huriez syndrome.

Affected patients with sclerotylosis have a greatly increased risk of cutaneous SCC. The clinical presentation of patients with sclerotylosis strongly indicates that this disorder is a precancerous condition. Affected individuals carry a >100-fold higher risk for the development of aggressive SCC of the skin (Levi et al. 1995; Gray et al. 1997), the age at onset of skin cancer is much lower than in the general population, and tumors arise in areas of affected skin. Sclerosis (Nachbar et al. 1993; Ozturk et al. 1998), atrophy, and scarring (Hagiwara et al. 1996) are well-recognized risk factors for the development of SCC of the skin. Clinically, the development of SCC in Huriez syndrome bears striking similarity to Marjolin's ulcer, which refers to malignancies arising in chronic ulcers of the skin, scar tissue, and burn scars (Fleming et al. 1990). The skin fragility found in Huriez syndrome leads to scarring, and the scleroatrophic changes may represent a process similar to scarring, thus predisposing to skin cancer. In addition, the exposure to exogenous mutagens such as arsenic (Jackson and Grainge 1975) have been recognized as risk factors for the development of SCC of the skin. The atrophic changes in skin anatomy observed in this type of PPK may disturb the barrier function of the skin, thus facilitating the penetration of putative physical, chemical, and infectious mutagens. However, most PPKs are not associated with an increased risk of skin cancer (Stevens et al. 1996). In addition, there is no evidence that sclerotylosis is associated with an increased risk of skin tumors other than SCC, which one would expect to observe if the development of local malignancies were solely attributable to the defective barrier function and the increased exposure to exogenic carcinogens.

It is noteworthy that loss of heterozygosity of 4q has been reported in 81% of squamous cell neoplasms of the head and neck (Pershouse et al. 1997) and in 46% of cervical carcinoma (Mitra et al. 1994). In SCC of the head and neck with deletions on 4q, the region that was consistently involved extends distal of the Huriez gene

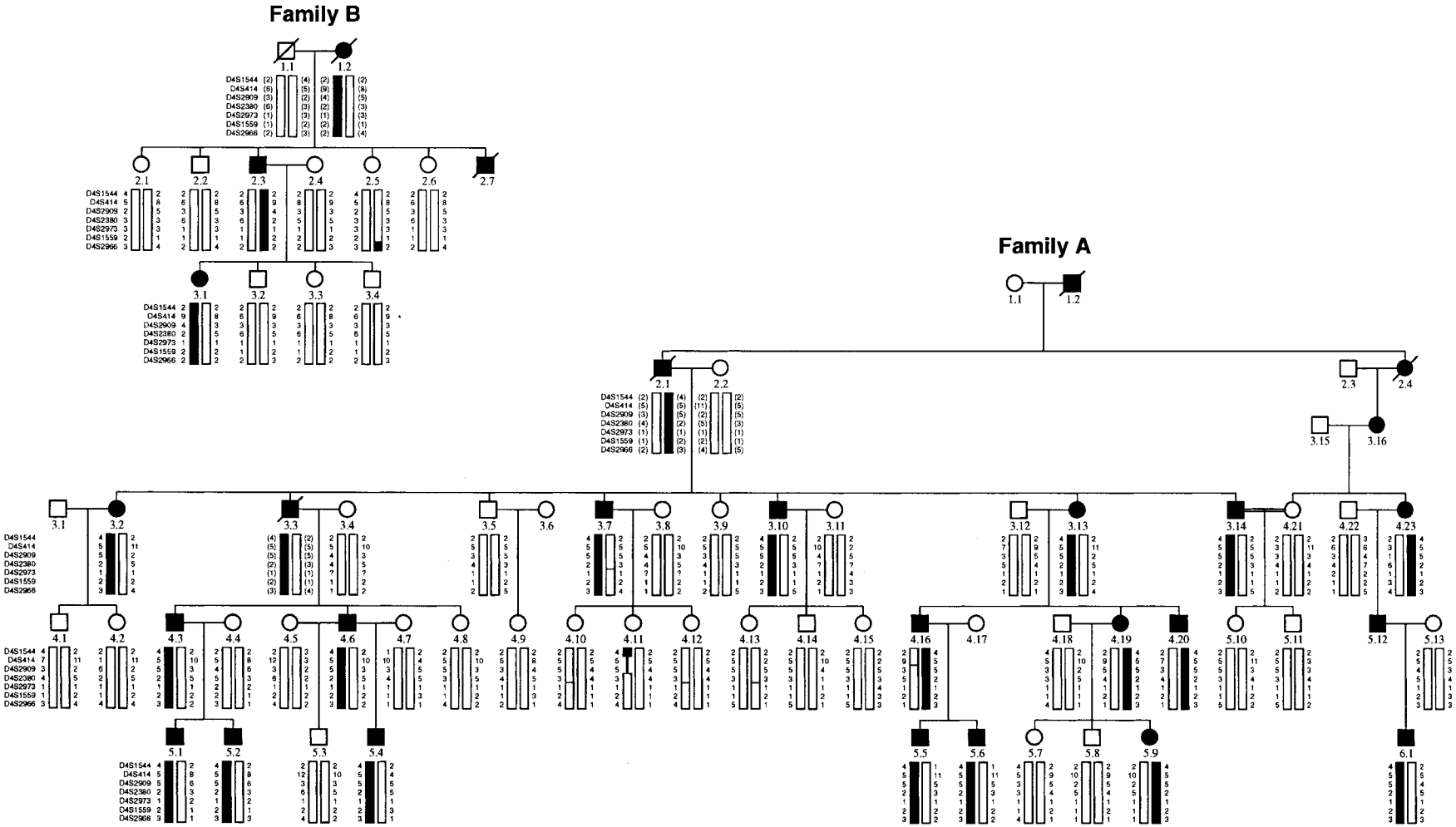

Figure 1 Pedigrees of families A and B with Huriez syndrome. Affected family members are denoted by blackened symbols. The most likely haplotype for seven markers from an 8-cM interval between D4S1544 and D4S2966 is shown below each individual.

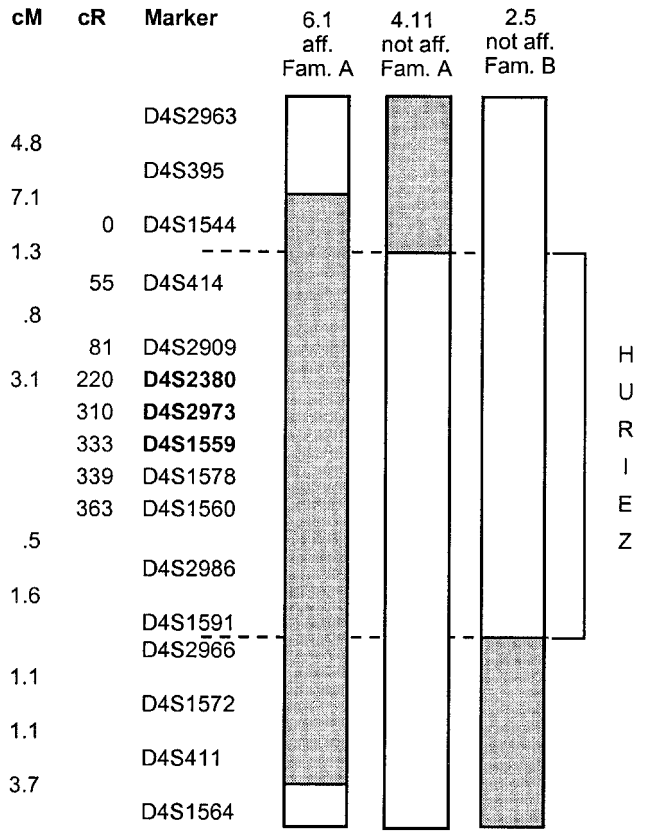

Figure 2 Linkage map of microsatellite markers from chromosome 4q23. Genetic distances are indicated in centimorgans. Markers are arranged in map order according to the final Généthon human linkage map (Dib et al. 1996). In the candidate region, markers are ordered according to their score on the TNG radiation hybrid panel. Here, relative positions are noted in cR after assigning 0 cR to marker D4S1544. The wild-type allele is represented by a white bar, the mutated allele by a gray bar. In the candidate region, markers IBS are noted in boldface type. The recombination events in members 4.11 and 6.1 of family 1 and 2.5 of family 2 are shown. The minimal candidate region for Huriez syndrome is marked. "aff." indicates affected; "not aff.," not affected.

locus and possibly overlaps with it on its centromeric end. A tumor-progression model for SCC of the skin has been proposed in which serial mutagenesis exceeds a threshold level, leading to tumor growth (Grossman and Leffell 1997). The gene mutation underlying Huriez syndrome may represent a first event in that series. Identification and characterization of the gene causing Huriez syndrome may provide important insights into the pathogenesis of skin cancer.

Acknowledgments

The kind cooperation of all members of families A and B is gratefully acknowledged. We thank G. Nürnberg for her excellent technical assistance. We are indebted to the Foundation Rene Touraine for their support. The Mikrosatellitenzentrum is supported by a grant from the German Human Genome Project (to A.R.).

YOUNG-AE LEE,[1,3] HOWARD P. STEVENS,[4,5] EMMANUEL DELAPORTE,[6] ULRICH WAHN,[3] AND ANDRÉ REIS[1,2]

[1]*Gene Mapping Center, Max-Delbrück Centrum,* [2]*Institute of Human Genetics, and* [3]*Department of Pediatrics, Pneumology, and Immunology, Charite, Campus Virchow, Humboldt-University, Berlin;* [4]*Academic Department of Dermatology, St. Bartholomew's and Royal London School of Medicine and Dentistry, London;* [5]*Department of Dermatology, Barnet General Hospital, Wellhouse Lane, Barnet, Herts, United Kingdom; and* [6]*Service de Dermatologie A, Hôpital Claude-Huriez, Lille, France*

Electronic-Database Information

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

Généthon, http://www.genethon.fr/
Human Gene Map, The, http://www.ncbi.nlm.nih.gov/SCIENCE96/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for Huriez syndrome [MIM 181600])

References

Brissenden JE, Ullrich A, Francke U (1984) Human chromosomal mapping of genes for insulin-like growth factors I and II and epidermal growth factor. Nature 310:781–784

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

Carpenter G, Cohen S (1979) Epidermal growth factor. Annu Rev Biochem 48:193–216

Delaporte E, N'guyen-Mailfer C, Janin A, Savary JB, Vasseur F, Feingold N, Piette F, et al (1995) Keratoderma with sclero-atrophy of the extremities or sclerotylosis (Huriez syndrome): a reappraisal. Br J Dermatol 133:409–416

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Fischer S (1978) La génodermatose scléro-atrophiante et kératodermique des extremités (au sujet de trois nouveaux cas familiaux). Ann Dermatol Venereol 105:1079–1082

Fleming MD, Hunt JL, Purdue GF, Sandstad J (1990) Marjolin's ulcer: a review and reevaluation of a difficult problem. J Burn Care Rehabil 11:460–469

Gray DT, Suman VJ, Su WP, Clay RP, Harmsen WS, Roenigk RK (1997) Trends in the population-based incidence of squamous cell carcinoma of the skin first diagnosed between 1984 and 1992. Arch Dermatol 133:735–740

Grossman D, Leffell DJ (1997) The molecular basis of nonmelanoma skin cancer: new understanding. Arch Dermatol 133:1263–1270

Hagiwara K, Uezato H, Miyazato H, Nonaka S (1996) Squamous cell carcinoma arising from lupus vulgaris on an old burn scar: diagnosis by polymerase chain reaction. J Dermatol 23:883–889

Hamm H, Traupe H, Brocker EB, Schubert H, Kolde G (1996) The scleroatrophic syndrome of Huriez: a cancer-prone genodermatosis. Br J Dermatol 134:512–518

Huriez C, Deminatti M, Agache P, Mennecier M (1968) Une génodysplasie non encore individualisée: la génodermatose scléro-atrophiante et kératodermique des extremités fréquemment dégénerative. Sem Hop 44:481–488

Jackson R, Grainge JW (1975) Arsenic and cancer. Can Med Assoc J 113:396–401

Kavanagh GM, Jardine PE, Peachey RD, Murray JC, De-Berker D (1997) The scleroatrophic syndrome of Huriez. Br J Dermatol 137:114–118

Kim JW, Kim YT, Kim DK, Song CH, Lee JW (1996) Expression of epidermal growth factor receptor in carcinoma of the cervix. Gynecol Oncol 60:283–287

Lambert D, Planche H, Chapuis JL (1977) La génodermatose scléro-atrophiante et kératodermique des extremités. Ann Dermatol Venereol 104:654–657

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Lee CS, Redshaw A, Boag G (1997) Epidermal growth factor receptor immunoreactivity in human laryngeal squamous cell carcinoma. Pathology 29:251–254

Levi F, Franceschi S, Te VC, Randimbison L, La-Vecchia C (1995) Trends of skin cancer in the Canton of Vaud, 1976–92. Br J Cancer 72:1047–1053

Mitra AB, Murty VV, Li RG, Pratap M, Luthra UK, Chaganti RS (1994) Allelotype analysis of cervical carcinoma. Cancer Res 54:4481–4487

Nachbar F, Stolz W, Volkenandt M, Meurer M (1993) Squamous cell carcinoma in localized scleroderma following immunosuppressive therapy with azathioprine. Acta Derm Venereol 73:217–219

Ozturk MA, Benekli M, Altundag MK, Guler N (1998) Squamous cell carcinoma of the skin associated with systemic sclerosis. Dermatol Surg 24:777–779

Pershouse MA, El-Naggar AK, Hurr K, Lin H, Yung WK, Steck PA (1997) Deletion mapping of chromosome 4 in head and neck squamous cell carcinoma. Oncogene 14:369–373

Pfeiffer P, Nexo E, Bentzen SM, Clausen PP, Andersen K, Rose C (1998) Enzyme-linked immunosorbent assay of epidermal growth factor receptor in lung cancer: comparisons with immunohistochemistry, clinicopathological features and prognosis. Br J Cancer 78:96–99

Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, et al (1996) A gene map of the human genome. Science 274:540–546

Shaw M, Formentini E, de Kaminsky AR, Kaminsky CA, Abulafia J (1978) Scleroatrophying and degenerative keratodermic genodermatosis of the extremities. Med Cutan Ibero Lat Am 6:291–295

Springer EA, Robinson JK (1991) Patterns of epidermal growth factor receptors in basal and squamous cell carcinoma. J Dermatol Surg Oncol 17:20–24

Stevens HP, Kelsell DP, Bryant SP, Bishop DT, Spurr NK, Weissenbach J, Marger D, et al (1996) Linkage of an American pedigree with palmoplantar keratoderma and malignancy (palmoplantar ectodermal dysplasia type III) to 17q24: literature survey and proposed updated classification of the keratodermas. Arch Dermatol 132:640–651

Stewart EA, McKusick KB, Aggarwal A, Bajorek E, Brady S, Chu A, Fang N, et al (1997) An STS-based radiation hybrid map of the human genome. Genome Res 7:422–433

Yano H, Shiozaki H, Kobayashi K, Yano T, Tahara H, Tamura S, Mori T (1991) Immunohistologic detection of the epidermal growth factor receptor in human esophageal squamous cell carcinoma. Cancer 67:91–98

Address for correspondence and reprints: Prof. Dr. André Reis, Institute of Human Genetics, Charité, Campus Virchow, Humboldt-University, Augustenburger Platz 1, D-13353 Berlin, Germany. E-mail: andre.reis@charite.de

0002-9297/2000/6601-0034$02.00

Am. J. Hum. Genet. 66:330–332, 2000

Transmission-Ratio Distortion at Xp11.4-p21.1 in Type 1 Diabetes

To the Editor:

Naumova et al. (1998) reported a deviation from the expected Mendelian 1:1 ratio of grandpaternal/grandmaternal alleles at loci in Xp11.4-p21.1 in the children of 47 families not selected on the basis of the disease status of the children. The transmission-ratio distortion (TRD) was found only among male offspring and was manifested as a bias in favor of the inheritance of the alleles of the maternal grandfather. The critical region, containing the putative TRD locus, named "*DMS1,*" was mapped to an interval bounded by DXS538 and DXS7 and peaking at DXS1068.

These observations might have an impact on the results of a study in which we have provided evidence of linkage to type 1 diabetes mellitus (MIM 222100), in the same region of chromosome X (Cucca et al. 1998). The possibility of TRD at chromosome Xp gives rise to the question of whether the diabetes-linkage results are indeed disease specific. The following evidence suggests that it is highly unlikely that the *DMS1* locus is responsible for the chromosome Xp linkage to type 1 diabetes.

We genotyped DXS1068 (a marker that was at the peak of our linkage curve) in two sample sets of control families not ascertained on the basis of the disease status of the children. These control families were from the Centre d'Etude du Polymorphisme Humaine (Fondation Jean Dausset/CEPH) and from a population-based sam-

ple from the town of Busselton, Australia, and included 61 large pedigrees with 603 children and 220 families with 525 children, respectively (Hill et al. 1995). Single-point allele sharing in these families with multiple sibships was corrected by means of the method proposed by Hodge (1984), as implemented in the software GENOME ANALYSIS SYSTEM, version 2.0. We obtained 51% sharing, by sib pairs, of one allele identical by descent (252.6 sharing one allele and 243.4 sharing no alleles; $P = .68$). This compares with 60.5% sharing for DXS1068 in 580 type 1 diabetic sib pair families (193 sharing one allele identical by descent and 126 sharing no alleles in a single-point analysis; $P = 2 \times 10^{-4}$). Hence, there is no evidence of TRD in these nondiabetic families that were analyzed in the same way that we analyzed the diabetic families.

We typed both 255 discordant (affected/unaffected) independent sib pairs from the United Kingdom and Sardinian families for DXS1068. We obtained identity-by-descent values of 77 and 86 for sharing one and no alleles, respectively, for all discordant pairs (47.6%); the corresponding values were 25 and 29 for male/male pairs only (46.3%). The trend toward <50% allele sharing in discordant sib pairs is consistent with a type 1 diabetes–specific effect.

The putative *DMS1* TRD locus near the DXS1068 locus affects only male progeny, whereas the strongest linkage that we observed for DXS1068 was in male/female pairs (Cucca et al 1998). The linkage of the DXS1068 region to type 1 diabetes was strongly concentrated in the 97 of 580 families that had human leukocyte antigen (HLA) DR3/X (X is not DR4) sib pairs (multipoint MLS = 3.5 at DXS1068). Some weak evidence of linkage was present also in the 195 DR3/4 sib pairs (multipoint MLS = .75 at DXS1068), and no evidence of linkage was obtained in the other 288 families with affected sib pairs (Cucca et al. 1998). We evaluated linkage on chromosome X, conditioning the data according to the genotype at the HLA *IDDM1* major locus on chromosome 6p21, because, in a large data set from Sardinia, the United Kingdom, and the United States, there was a strong increase in the male:female (M:F) ratio, which was almost exclusively restricted to patients with the DR3/X genotype (M:F ratio = 1.7; $P = 4.7 \times 10^{-7}$), compared with a ratio of 1.0 in the DR4/Y category (Y is not DR3), with a small effect in DR3/4 patients (M:F ratio = 1.2; $P = .03$) (Cucca et al. 1998). Hence, both the evidence of linkage and the bias in the M:F ratio were concentrated in the families with patients positive for the HLA-DR3, which is one of the two main predisposing haplotypes at the HLA/*IDDM1* major locus. These data suggest an interaction between HLA-DR3 haplotypes and the diabetes locus on chromosome X, which is unlikely to be caused by an effect of the putative *DMS1* locus. Furthermore, both in the unaffected parents and siblings of the United Kingdom and in the Sardinian families who were DR3/3 homozygotes, the ratio was reversed: 56 males and 80 females (0.7; $P = .04$) compared with an M:F ratio of 2.2 in DR3/3 patients ($P = 1.3 \times 10^{-6}$), thereby providing another control for our results (Cucca et al. 1998).

These results indicate that it is unlikely that the type 1 diabetes linkage that we have observed is explained by the *DMS1* locus. TRD could, however, have a significant effect on the interpretation of linkage studies of polygenes in common diseases in which increases in allele sharing at the disease locus may be very small. The possibility of TRD should be ruled out in disease studies (Eaves et al. 1999).

Acknowledgments

We thank the Italian Telethon, for financial support, and W. Cookson, for the use of the Busselton family DNA samples. F.C. and J.A.T. are recipients of a Wellcome Trust Biomedical Research Collaboration Grant.

PATRIZIA ZAVATTARI,[1,2] LAURA ESPOSITO,[1] SARAH NUTLAND,[1] JOHN A. TODD,[1] AND FRANCESCO CUCCA[1,2]

[1]*Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge; and* [2]*Istituto di Clinica e Biologia dell'Eta' Evolutiva, University of Cagliari, Cagliari, Italy*

Electronic-Database Information

The accession number and URLs for data in this article are as follows:

Fondation Jean Dausset/CEPH, http://www.cephb.fr
Genome Analysis System, http://info.ox.ac.uk/~ayoung/gas.html
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for type 1 diabetes mellitus [MIM 222100])

References

Cucca F, Goy JV, Kawaguchi Y, Esposito L, Merriman ME, Wilson AJ, Cordell HJ, et al (1998) A male-female bias in type 1 diabetes and linkage to chromosome Xp in MHC HLA-DR3-positive patients. Nat Genet 19:301–302

Eaves IA, Bennett ST, Forster P, Ferber KM, Ehrmann D, Wilson AJ, Bhattacharyya S, et al (1999) Transmission ratio distortion at the *INS-IGF2* VNTR. Nat Genet 22:324–325

Hill MR, James AL, Faux JA, Ryan G, Hopkin JM, le Souef P, Musk AW, et al (1995) FcεRI-β polymorphism and risk of atopy in a general population sample. Br Med J 311: 776–779

Hodge SE (1984) The information contained in multiple sibling pairs. Genet Epidemiol 1:109–122

Naumova AK, Leppert M, Barker DF, Morgan K, Sapienza C (1998) Parental origin-dependent, male offspring-specific transmission-ratio distortion at loci on the human X chromosome. Am J Hum Genet 62:1493–1499

Address for correspondence and reprints: Francesco Cucca, Istituto di Clinica e Biologia dell'Eta' Evolutiva, Via Jenner, University of Cagliari, Cagliari, Sardinia, Italy. E-mail: fcucca@mcweb.unica.it

0002-9297/2000/6601-0035$02.00

Am. J. Hum. Genet. 66:332–335, 2000

Predominance of the T14484C Mutation in French-Canadian Families with Leber Hereditary Optic Neuropathy Is Due to a Founder Effect

To the Editor:

The Leber hereditary optic neuropathy (LHON) phenotype was first defined, more than 125 years ago, as a maternally inherited optic neuropathy that primarily affects young adult men (Leber 1871). Loss of central vision may be acute or subacute, and peripheral vision is preserved. Slow delayed recovery can occur, but relative central scotomata usually remain (Johns et al. 1993). The chronic stage of LHON is characterized by optic atrophy (MIM 535000).

LHON is associated with three primary mtDNA mutations, all of which occur in genes coding for subunits of complex I of the mitochondrial respiratory chain: G3460A in *ND1,* G11778A in *ND4,* and T14484C in *ND6* (MIM 516006.0001) (Howell et al. 1995). All three of these mutations alter evolutionarily conserved amino acids and are not found in control individuals. The relative frequency of the three primary LHON mtDNA mutations varies considerably in different populations, although G11778A is the most common worldwide. We have recently found that T14484C is by far the most common mutation in French-Canadian families with LHON (Macmillan et al. 1998). The results of previous studies of mutation profiles of several metabolic disorders have demonstrated founder effects, including phenylketonuria (Rozen et al. 1994) and familial hyperchylomicronemia (De Braekeleer et al. 1991), in the French-Canadian population.

To test the hypothesis that the predominance of the T14484C mutation in French-Canadian families with LHON is caused by a founder effect, we sequenced a segment of the mtDNA displacement (D) loop and a segment of the control region from French-Canadian families with the T14484C mutation. Variation in these noncoding regions has been used extensively to study the evolution of modern populations. The regions sequenced included hypervariable regions (HVR) I and II (Vigilant et al. 1989). The D-loop region extends from nucleotide (nt) 189 backward through 0 to nt 16024 of the 16,569-bp mtDNA. The control region extends from nt 189 forward, toward the tRNAs. These two regions contain the fastest-evolving regions of mtDNA (Upholt and Dawid 1977), which have an estimated rate of evolution that is 2.8–5 times that of the remainder of the mitochondrial genome (Aquadro and Greenberg 1983; Cann et al. 1984). The average nucleotide-sequence variation in this region has been calculated to be 1.7% (Aquadro and Greenberg 1983). We reviewed the records of patients with suspected LHON who were independently referred, for molecular diagnosis, to the Montreal Neurological Hospital DNA Diagnostic Laboratory. By cycle sequencing with the use of a New England Biolabs kit (manual) or dye-labeled dideoxy terminators on an ABI system (automated), we sequenced regions that included HVR I and HVR II in the following individuals: one member of each French-Canadian family with the T14484C mutation, one member of a French-Canadian family with the G11778A mutation, and one member of a French-Canadian family without a family history of LHON and with none of the three primary LHON mutations. Sequencing was done in two reactions, by use of the following primer pairs: L15996 and H16401 (which includes HVR I) and L29 and H408 (which includes HVR II; Vigilant et al. 1989). "L" refers to "light" strand and "H" refers to "heavy" strand, and the numbers refer to the Cambridge sequence (Anderson et al. 1981). The institutional review boards of the Montreal Neurological Hospital and the University of Illinois at Chicago approved this study.

We analyzed 27 independently referred French-Canadian families with LHON and the T14484C mutation, all of which were homoplasmic for the T14484C mutation. We found eight homoplasmic transition mutations (C16069T, T16126C, G16213A, A73G, G185A, G228A, A263G, and C295T; table 1), compared with the Cambridge sequence, in the families with the T14484C mutation. Of the 27 families analyzed, 26 shared identical substitutions at all eight sites that were different from the Cambridge sequence. Six of these mutations were found only in families with the T14484C mutation and not in either the family with the G11778A family or the family without LHON (table 1). The mutated sites were distributed throughout the D-loop and control region: one was located in the large central conserved sequence block (CSB; A73G), one (G228A) was in CSB 1, two (T16126C and G16213A) were in HVR I, three (G185A, A263G, and C295T) were in HVR II, and one (C16069T) was just outside HVR I.

In addition, 22/27 families with LHON and the T14484C mutation had a C insertion in the homopolymeric stretch of C's before the T at position 310. All 27

Table 1

Variable Sites in the D-Loop and Control Regions of mtDNA from French-Canadian Families with LHON and the T14484C Mutation (27 Pedigrees), with LHON and the G11778A Mutation, and without LHON

	Presence of Mutation[a]							
Pedigree	C16069T	T16126C	G16213A	A73G	G185A	G228A	A263G	C295T
1	+	+	+	+	+	+	+	+
2	+	+	+	+	+	+	+	+
3	+	+	+	+	+	+	+	+
4	+	+	*	+	+	+	+	A
5	+	+	+	+	+	+	+	+
6	+	+	+	+	+	=	+	+
7	+	+	+	+	+	+	+	+
8	+	+	+	+	+	+	+	+
9	+	+	+	+	+	+	+	+
10	+	+	+	+	+	+	+	+
11	+	+	+	+	+	+	+	+
12	+	+	+	+	+	+	+	+
13	+	+	+	+	+	+	+	+
14	+	+	+	+	+	+	+	+
15	+	+	+	+	+	+	+	+
16	+	+	+	+	+	+	+	+
17	+	+	+	+	+	+	+	+
18	+	+	+	+	+	+	+	+
19	+	+	+	+	+	+	+	+
20	+	+	+	+	+	+	+	+
21	+	+	+	+	+	+	+	+
22	+	+	+	+	+	+	+	+
23	+	+	+	+	+	+	+	+
24	+	+	+	+	+	+	+	+
25	+	+	+	+	+	+	+	+
26	+	+	+	+	+	+	+	+
27	+	+	+	+	+	+	+	+
G11778A	+	*	*	*	*	*	+	*
Without LHON	*	*	*	*	*	*	+	*

[a] A plus sign (+) indicates that the transition mutation indicated was present at that site. An asterisk (*) indicates the Cambridge consensus sequence. "A" indicates adenine.

families with the T14484C mutation, the family with the G11778A mutation, and the family without LHON had a C insertion in the homopolymeric stretch of C's after the T at position 310. Finally, there were some sites at which the sequence appeared to be heteroplasmic; however, these results require independent confirmation.

These data demonstrate that French-Canadian families with LHON and the T14484C mutation likely share the same maternal lineage and suggest that they may all have been derived from a single founder woman. All the observed sequence variants—with the exception of the insertions in the homopolymeric tract of C—were transitions, as compared with the Cambridge sequence; the transition-to-transversion ratio of nucleotide substitutions has been reported to be very high in this region (Aquadro and Greenberg 1983). All our patients were positive for mutations at positions 3394, 4216, and 13708, relative to the standard sequence, and it is likely that they belong to the same haplogroup as does the patient described by Brown et al. (1992). This haplogroup is clearly related to white haplogroup J, which is characterized by mutations at positions 4216 and 13708 and which is where T14484C mutations cluster in the European population (Brown et al. 1997; Torroni et al. 1997). The most-similar published haplotype is that of the haplogroup J T14484C "TAS2" individual, in which seven of the mutations (C16069T, T16126C, A73G, G185A, G228A, A263G, and C295T) were common (Howell et al. 1995). There were four additional TAS2 mutations that were not found and three additional TAS2 mutations that were not evaluated in the French-Canadian families. The French-Canadian mutation array is even less similar to those reported for five other haplogroup J T14484C individuals, seven G11778A individuals, and three G3460A individuals (Howell et al. 1995; Hofmann et al. 1997). Furthermore, the French-Canadian array is different from that reported for haplogroup J individuals without LHON (Hofmann et al. 1997).

Although the T16213A mutation has been reported in several populations (Horai and Hayasaka 1990; Lum et al. 1994; Mountain et al. 1995), this is the first report

of its association with a family with LHON. The single family (pedigree 4) that did not share this mutation may have undergone a reversion mutation to the original sequence; the representative of this family also had a C→A transversion at site 295. The small changes in the D-loop–sequence homopolymeric tracts of C starting at position 303 are variants of the published sequence first reported by Greenberg et al. (1983). The C insertion after the T at position 310 was present in all French-Canadian families, and it was also found in the family with TAS2. However, the C insertion before the T at position 310 was not universally present in the French-Canadian families with the T14484C mutation. These insertions are very unstable and may have arisen after the migration of the common female ancestor to New France. These French-Canadian families with LHON and the T14484C mutation represent a unique genetic resource in which to evaluate the rate of accumulation of sequence variants and the resolution of heteroplasmy in mtDNA on a time scale of a few hundred years.

Acknowledgments

We thank the families for their participation in this study. This study was supported by a University of Illinois at Chicago Campus Research Board grant (to C.M.), grants from the Medical Research Council of Canada, and an FRSQ-Hydro-Quebec technology transfer grant (to E.A.S.). E.A.S. is a Montreal Neurological Institute Killam Scholar.

C. Macmillan,[1] T. A. Johns,[2] K. Fu,[2,3] and E. A. Shoubridge[2,3,4]

[1]*Departments of Neurology and Pediatrics, University of Illinois at Chicago, Chicago; and* [2]*Montreal Neurological Institute and Departments of* [3]*Human Genetics and* [4]*Neurology and Neurosurgery, McGill University, Montreal*

Electronic-Database Information

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

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for Leber optic atrophy [MIM 535000] and MTND6*LHON14484A [MIM 516006.0001]).

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457–465

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103:287–312

Brown MD, Sun F, Wallace DC (1997) Clustering of Caucasian Leber hereditary optic neuropathy patients containing the 11778 or 14484 mutations on an mtDNA lineage. Am J Hum Genet 60:381–387

Brown MD, Voljavec AS, Lott MT, MacDonald I, Wallace DC (1992) Leber's hereditary optic neuropathy: a model for mitochondrial neurogenetic diseases. FASEB J 6:2791–2799

Cann RL, Brown WM, Wilson AC (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 106:479–499

De Braekeleer M, Dionne C, Gagne C, Julien P, Brun D, Ven Murthy MR, Lupien PJ, et al (1991) Founder effect in familial hyperchylomicronemia among French Canadians of Quebec. Hum Hered 41:168–173

Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21:33–49

Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, Paprotta A, Gerbitz KD (1997) Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. Hum Mol Genet 6: 1835–1846

Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am J Hum Genet 46:828–842

Howell N, Kubacka I, Halvorson S, Howell B, McCullough DA, Mackey D (1995) Phylogenetic analysis of the mitochondrial genomes from Leber hereditary optic neuropathy pedigrees. Genetics 140:285–302

Johns DR, Heher KL, Miller NR, Smith KH (1993) Leber's hereditary optic neuropathy: clinical manifestations of the 14484 mutation. Arch Ophthalmol 111:495–498

Leber T (1871) Ueber hereditaere und congenital-angelegte Sehnervenleiden. Arch Ophthalmol 17:249–291

Lum JK, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum Biol 66:567–590

Macmillan C, Kirkham T, Fu K, Allison V, Andermann E, Chitayat D, Fortier D, et al (1998) Pedigree analysis of French Canadian families with T14484C Leber's hereditary optic neuropathy. Neurology 50:417–422

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979–992

Rozen R, Mascisch A, Lambert M, Laframboise R, Scriver CR (1994) Mutation profiles of phenylketonuria in Quebec populations: evidence of stratification and novel mutations. Am J Hum Genet 55:321–326

Torroni A, Petrozzi M, D'Urbano L, Sellitto D, Zeviani M, Carrara F, Carducci C, et al (1997) Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. Am J Hum Genet 60: 1107–1121

Upholt WB, Dawid WB (1977) Mapping of mitochondrial DNA of individual sheep and goats: rapid evolution in the D loop region. Cell 11:571–583

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86:9350–9354

Address for correspondence and reprints: Dr. Carol Macmillan, Department of Neurology, M/C 796, University of Illinois at Chicago, 912 South Wood Street, Room 855N, Chicago, IL 60612-7330. E-mail: cmacmill@uic.edu

0002-9297/2000/6601-0036$02.00

Am. J. Hum. Genet. 66:335–338, 2000

Testing for Linkage Disequilibrium, Maternal Effects, and Imprinting with (In)complete Case-Parent Triads, by Use of the Computer Program LEM

To the Editor:

The traditional transmission/disequilibrium test (TDT) and related tests (see Thomson 1995) require complete triads of genotyped cases plus both parents, in order to test for linkage disequilibrium in the presence of population admixture. A problem in empirical research is that some of the genotype measurements will usually be missing. These incomplete triads must be discarded to ensure the validity of the TDT (Curtis and Sham 1995). Recently, Weinberg (1999*a*) developed likelihood-ratio tests (LRTs) that used the expectation-maximization (EM) algorithm (Dempster et al. 1977), to use incomplete triads as well. Weinberg's tests capitalize on the fact that parent-child dyads may be informative about the genotype of the missing parent. For instance, if a child and a parent are both homozygous for the variant allele, the genotype of the missing parent should comprise at least one copy. Simulations showed that the EM-LRTs were more powerful than the traditional tests that exclude incomplete triads and that they recaptured much of the loss in information caused by missing parental genotypes.

The widespread use of this valuable approach, however, seems hampered by a lack of accessible software. Weinberg, for instance, used the commercial package GLIM, which is good and flexible software but not very user friendly (see remarks on their Internet site), and it requires programming in order to perform the EM-LRTs. To suggest an alternative, we discuss the script to perform Weinberg's tests (1999*b*) for linkage disequilibrium, maternal effects, or parent-of-origin effects in LEM, which is a program for log-linear analysis with missing data that uses the EM algorithm (Vermunt 1997*a*, 1997*b*). An important advantage of LEM is that, with this script, all the tests discussed by Weinberg (1999*b*) can readily be performed in the presence of all possible patterns of missing data, without programming work or the need to learn more LEM syntax. Furthermore, the program is optimized for rapid convergence with EM algorithm, and standard errors of the estimates, fit indices, and a number of appropriate tests are automatically reported in the output so that they do not have to be programmed separately. A final advantage is that the program (which has a DOS and a Windows version) and the manual can be downloaded free of charge on the Internet at the Web site for Methoden en Technieken van Onderzoek (mto).

With a biallelic locus assumed, the genotypes of the mother (M), father (P), and child (C) contain no copy, one copy, or two copies of the variant allele. If the D's are dummy variables (e.g., $D_{(C=1)}$ means that the variable is 1 in all triads in which C = 1 and is 0 otherwise), then the log of the expected cell counts $E(n_{MPC})$ of Weinberg's (1999*b*, see table 1) full model can be written as

$$\ln[E(n_{MPC})] = \gamma_j + \beta_p D_{(C=1)} + \beta_2 D_{(C=2)} + \alpha_1 D_{(M=1)} + \alpha_2 D_{(M=2)} + \ln(w_{MPC}) \ ,$$

where $e^{\gamma_j} = \mu_j$ are the mating-type–stratum effects (e is the natural exponent), $e^{\beta_p} = R_p$ is the ratio of the risk of disease for genotypes with one copy versus no copies of the variant allele, $e^{\beta_2} = R_2$ is the risk ratio when the genotype comprises two versus no copies of the variant allele, $e^{\alpha_1} = S_1$ is the risk ratio or maternal effect when the mother has one copy versus no copies of the variant allele, and $e^{\alpha_2} = S_2$ is the risk ratio when the mother has two copies versus no copies of the variant allele. The w_{MPC} are cell weights (this becomes clearer when the component is moved to the left-hand side of the equation, so that we obtain $\ln[E(n_{MPC})] - \ln[E(w_{MPC})] = \ln[E(n_{MPC}/w_{MPC})]$), or, in GLIM terminology, $\ln(w_{MPC})$ is called the "offset." The weights can have four different values. First, they can be 0. Because the expected counts in these cells have to be multiplied with $e^{\ln(0)} = 0$, the implication is that the cell frequencies are fixed at 0. This weight is therefore assigned to combinations—such as M = 2, P = 2, and C < 2—that, for theoretical reasons, cannot occur. They are also useful in the context of recovery of information from incomplete triads. For example, if, in the situation described above, the genotype of the child is missing, the 0 weights for C < 2 imply that the missing genotype must comprise two copies of the variant allele. Second, the weights can be 1, so that the expected cell counts are multiplied with $e^{\ln(1)} = 1$, implying that the frequencies as predicted by R_p, R_2, S_1, and S_2 remain unaltered. Third, in the triads M = 2, P = 1, C = 1; M = 2, P = 0, C = 1; and M = 1, P = 0, C = 1 (M > F), where the child receives the copy of the variant allele from the mother, the weights equal the "parent of origin" or "imprinting" effect I_m. Because

the models also specify a "main" effect $e^{\beta_p} = R_p$, the total effect of C = 1 on the expected count becomes $I_m R_p$. It is a bit unusual to use parameters as weights. The cause is the triads consisting entirely of heterozygotes (M = P = C = 1) for whom only the total cell count is observed, and it is unclear how many children receive the variant allele from the mother and how many from the father. As a result, the effect of C = 1 on the cell count involves the sum of $R_p + I_m R_p$, which cannot be modeled in the usual way as products of effects. The effect of C = 1 is therefore written as $(1 + I_m)R_p$, where $(1 + I_m)$ is the cell weight that can be modeled as a sum of effects.

A LEM script that estimates this model in the presence of all possible patterns of missing genotypes is shown in the Appendix. The data are analyzed as a 3 × 3 × 3 table (indicated in the script by the last three numbers after the statement *dim*), defined by the three manifest (*man* 3) or measured genotypes of the mother, father, and child, labeled "M," "P," and "C," respectively (see *lab* statement). The cell indices correspond to the number of copies of the variant allele plus one. Thus, the count of the triads M = 0, P = 2, C = 1 falls into cell 1,3,2. The cells are numbered in increasing order, where the last indices change first (1,1,1; 1,1,2; 1,1,3; 1,2,1; 1,2,2; etc.). The statements *mod* and *des* are used to specify the model and parameters. The *mod* statement indicates the number of parameters and the margin of the table that is affected. For instance, *fac*(*C,2*) means that two parameters or main effects are estimated for the effects of the genotype of the child. The margin of C consists of three cells, and the *des* statement specifies how the parameters affect these cells. In this case, *"0 1 2"* means that (1) the effect in all cells where C = 0 is 0, so that this category is used as the baseline; (2) the first parameter represents the effect in all cells where C = 1 (β_p)*;* and (3) the second parameter represents the effect in all cells where C = 2 (β_2). The mating-type stratum effects are defined by the specific combination of the maternal and paternal genotype and, therefore, pertain to the margin MP. Although there are 3 × 3 = 9 possible combinations, because of the assumed symmetry across parents within each mating type (e.g., M = 1, P = 2 and M = 2, P = 1 have equal effects) only six effects are estimated. LEM knows such a symmetric margin as the prespecified design 3a, so that with the use of the statement *spe*(*MP,3a*) there is no need for further specification in the *des* statement. The weights are combinations of constants and the imprinting parameter β_m and are specified with the help of a latent variable *X* (statement *lat 1*), which has two discrete classes (the second number after the command *dim*). The effects of the first class are 0, implying an impact of $e^0 = 1$ on the cell counts, and the effects of the second class are β_m, corresponding with the imprinting parameter $e^{\beta_m} = I_m$. Because only one parameter is estimated, and because this parameter is modeled as an effect of the second latent class, *fac*(*X,1*) is used in the model statement, and *0 1* is used in the design statement. The command *wei*(*XMPC*) means that the effects of the latent classes on the cell counts are mediated by the weight vector. The values for $X = 1$ after the statement *sta wei*(*XMPC*) specify which of the 27 cells are affected by the first latent class ("0" means not affected, and "1" means affected), and the values for $X = 2$ indicate the cells that are affected by the second latent class. For the combinations that cannot occur, two 0's are specified, so that the expected cell counts are multiplied with $e^{\ln(w\text{MPC})} = e^{\ln(0 \times 1 + 0 \times I_m)} = 0 \times 1 + 0 \times I_m = 0$. For the triads in which M > F, a value of 0 is specified for the first latent class, and a value of 1 is specified for the second latent class. This implies an effect of $0 \times 1 + 1 \times I_m = I_m$ on the cell count. For the triads M = P = C = 1, 1's are specified for both latent classes, so that the total impact becomes $1 \times 1 + 1 \times I_m = (1 + I_m)$. Note that, if the effect of the second latent class is fixed to 0 as well (no imprinting $\beta_m = 0$ and $e^{\beta_m} = 1$), the weight becomes 1 for all combinations that can occur and becomes 2 for triads consisting entirely of heterozygotes.

Tests can be performed by merely changing the number of parameters in the *mod* statement plus the parameter specification in the *des* statement. For instance, to fit a model without imprinting, we would use *fac*(*X,0*) instead of *fac*(*X,1*) and *0 0* instead of *0 1*. The output of LEM reports the log likelihoods plus a variety of other fit indices, parameter estimates, standard errors of the estimates, and comparisons between estimated and observed cell frequencies. To perform an LRT, one needs to take two times the difference between the log likelihoods of the full model and the model without imprinting. Because one parameter is fixed to 0, this statistic will be χ^2 distributed with 1 df. A number of submodels are worth mentioning. If we assume that there are no imprinting and no maternal effects (*fac*(*M,0*) and that *des* = *0 0 0*), then Schaid and Sommer's (1993) genotype relative-risk method is obtained, in which $e^{\beta_p} = P_1$ and $e^{\beta_2} = P_2$. Recessive models $\beta_p = 0$, $\beta_2 > 0$ can be specified by *fac*(*C,1*) and *des* [*0 0 1*], dominance models $\beta_p = \beta_2$ by *fac*(*C,1*) and *des* [*0 1 1*]. Although for polygenic traits it may be a somewhat coincidental situation (Van den Oord 1999), a gene-dosage model is obtained by imposing the constraint $\beta_2 = 2 \times \beta_p$ by use of *cov*(*C,1*) and *des* [*0 1 2*]. Note that the command *cov* instead of *fac* must be used. The reason is that C is now treated as a covariate rather than as a nominal factor, because the expected cell frequencies are linear in C (if C = 0, the effect is $0 \times \beta_p$; if C = 1, the effect is $1 \times \beta_p$; and, if C = 2, the effect is $2 \times \beta_p$). This latter test is asymptotically equivalent to the traditional TDT (Spielman et al. 1993), so that LEM also enables one to perform a variant of the TDT with incomplete triads.

The name after the command *dat* in the LEM script means that the data are in the file TEST.DAT. The number after *rec* shows that there are 100 triads. The data are in free format, with one record for each triad. The first two records are 3 3 3 and 1 0 1. The numbers indicate the cell to which the triad belongs, and 0's are used for missing genotypes. Thus, 3 3 3 pertains to a triad in which all three members have two copies of the variant allele (M = F = C = 2), and 1 0 1 pertains to a triad in which the genotype of the father is missing and in which the mother, as well as the child, has 0 copies of the variant allele. There are seven possible data patterns. This is indicated by the first number after the command *dim*. To inform LEM about the nature of patterns, the statement *sub* is used. For example, MPC pertains to triads with nothing missing, MC to triads with the genotype of the father missing. Maximum-likelihood estimates are obtained by means of the EM algorithm. The E step of this iterative method is of the form

$$n^e_{\mathrm{MPC}} = n_{\mathrm{MPC}} + n_{\mathrm{MP0}}\pi^e_{\mathrm{C}\,|\,\mathrm{MP}} + n_{\mathrm{M0C}}\pi^e_{\mathrm{P}\,|\,\mathrm{MC}} + n_{\mathrm{0PC}}\pi^e_{\mathrm{M}\,|\,\mathrm{PC}} + n_{\mathrm{M00}}\pi^e_{\mathrm{PC}\,|\,\mathrm{M}} + n_{\mathrm{0P0}}\pi^e_{\mathrm{MC}\,|\,\mathrm{P}} + n_{\mathrm{00C}}\pi^e_{\mathrm{MP}\,|\,\mathrm{C}} \ .$$

The 0's indicate that the genotype is missing, and superscript *e* indicates that the statistic is estimated and not observed. Thus, estimates of observed cell entries are computed with the use of the observed data plus the current estimates of the predicted cell frequencies that are made on the basis of the information from incomplete triads as well. In the M step of the EM algorithm, the predicted cell counts n^e_{MPC} are treated as if they were really observed, to obtain new estimates of the log-linear parameters and of the cell frequencies. To speed up the estimation, the program is instructed to switch to Newton-Raphson after 10 iterations (command *new*). Convergence is usually reached in <1 s on an ordinary computer.

To examine whether the script worked properly, we first computed expected cell frequencies, using the full model. Fitting the script to these frequencies gave a perfect fit, and the correct parameters were recovered. Next, we simulated 1,000 samples of 100 triads in six different conditions for which missing paternal genotypes of 0%, 10%, 20%, 30%, 40%, and 50% were assumed. The data were simulated with the assumption of two completely segregated strata that were mixed, so that the sample comprised approximately equal proportions of triads from each stratum. Within the first stratum, the frequency of the disease allele was .10 and the disease risk was .01; within the second stratum, the frequency of the disease allele was .9 and the disease risk of .1 was 10 times greater. When data were simulated under the assumption of no genetic effects, the null hypothesis $\beta_p = \beta_2 = 0$ was rejected in 4.3%, 6.1%, 4.4%, 5.5%, 4.8%, and 4.9% of the 1,000 samples. *Z*-tests showed that none of the rejection rates differed significantly from the expected type 1 or alpha error of 5%. This showed that the tests for genetic effects were accurate, even in conditions under which the number of missing paternal genotypes was substantial. The whole simulation was repeated by generating the data with $\beta_2 > 0$ assumed. The rejection rates of the null hypothesis or the power in the six conditions was 52.1%, 53.4%, 48.0%, 49.9%, 42.2%, and 43.8%. This confirmed results, reported by Weinberg (1999*a*), showing that, even with many incomplete triads, the EM LRT recaptures much of the loss in information.

The scripts for all the tests discussed in this article, sample data, and output can be downloaded from the first author's Internet site, Pedagogiek Utrecht. We should mention that Weinberg (1999*b*) proposed an alternative test for parent-of-origin effects that is also valid in situations in which the locus is a marker rather than a candidate gene. A script plus documentation for this parent-of-origin LRT can be found at that site as well.

Edwin J. C. G. van den Oord[1] and
Jeroen K. Vermunt[2]
[1]*Department of Child and Adolescent Psychology, Utrecht University, Utrecht, and* [2]*Department of Methodology, Tilburg University, Tilburg, the Netherlands*

Appendix

The following LEM script estimates the full model reported by Weinberg (1999*b*, table 1). The numbers and text in boldface indicate the only instructions that need to be changed in order to perform significance tests and to adjust the data format to one's own data.

```
* variable and table definition
man 3                    * # manifest variables
lat 1                    * # latent variables
res 1                    * # response variables
dim 7 2 3 3 3            * dimension table: R × X × M × P × C
lab R X M P C            * labels R=patterns, X=lat var, M=moth,
P=fath, C=child
sub MPC MP MC PC M P C   * possible data patterns or subgroups

* model
mod XMPC {spe(MP,3a)     *  mating-type-stratum effects
          fac(C,2)       *  child effects
          fac(M,2)       *  maternal effects
          fac(X,1)       *  imprinting effect
          wei(XMPC)}     *  weight vector

* data format
rec 100                  * # records or triads
dat TEST.DAT             * data file

* design matrix/parameter specification
des [0 1 2               * child effects
     0 1 2               * maternal effects
     0 1  ]              * imprinting effect

* values weight vector
sta wei(XMPC)
*M=0,P=0;M=0,P=1;M=0,P=2;M=1,P=0;M=1,P=1;M=1,P=2;M=2,P=0;M=2,P=1;M=2,P=2

[ 1 0 0   1 1 0   0 1 0   1 0 0   1 1 1   0 1 1   0 0 0   0 0 1   0 0 1  * X=1

  0 0 0   0 0 0   0 0 0   0 1 0   0 1 0   0 0 0   0 1 0   0 1 0   0 0 0] * X=2


* optimization
new 10 1                 * switch to Newton-Raphson after 10 EM
iterations
```

Electronic-Database Information

URLs for data in this article are as follows:

KUB, Departement Methoden en Technieken van Onderzoek (mto), http://cwis.kub.nl/~fsw_1/mto_snw.htm#software
Pedagogiek Utrecht, http://www.fss.uu.nl/ped/welcome.html

References

Curtis D, Sham PC (1995) A note on the application of the transmission disequilibrium test when a parent is missing. Am J Hum Genet 56:811–812

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Serv B 39:1–22

Schaid DJ, Sommer SS (1993) Genotype relative risks: methods for design and analysis of candidate-gene association studies. Am J Hum Genet 53:1114–1126

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

Thomson G (1995) Mapping disease genes: family-based association studies. Am J Hum Genet 57:487–498

van den Oord EJCG (1999) A comparison between different designs and tests to detect QTLs in association studies. Behav Genet 29:245–256

Vermunt JK (1997*a*) LEM: a general program for the analysis of categorical data. Tilburg University, Tilburg, the Netherlands

——— Vermunt JK (1997*b*) Advanced quantitative techniques in the social sciences. Vol 8: Log-linear models for event histories. Sage, Thousand Oakes, CA

Weinberg CR (1999*a*) Allowing for missing parents in genetic studies of case-parent triads. Am J Hum Genet 64: 1186–1193

——— (1999*b*) Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. Am J Hum Genet 65:229–235

Address for correspondence and reprints: Dr. Edwin van den Oord, Department of Child and Adolescent Psychology, Universiteit Utrecht, Heidelberglaan 1, Centrumgebouw Zuid, Postbus 80140, 3508 TC, Utrecht, the Netherlands. E-mail: E.vandenOord@fss.uu.nl

0002-9297/2000/6601-0037$02.00