Letters to the Editor

Am. J. Hum. Genet. 63:1216–1220, 1998

Maternal Uniparental Disomy of Chromosome 1 with No Apparent Phenotypic Effects

To the Editor:

Uniparental disomy (UPD) arises when an individual inherits two copies of a specific chromosome from one parent and no copy from the other parent. This unusual non-Mendelian transmission of parental genes may lead to rare recessive disorders, or to developmental disturbances due to aberrant imprinting effects, in the zygote (Ledbetter and Engel 1995). However, UPD may also occur (at some unknown frequency) with no apparent phenotypic consequences. Recently, the *Journal* reported the first case of maternal chromosome 1 UPD (Pulkkinen et al. 1997) and the first case of paternal chromosome 1 UPD (Gelb et al. 1998), both ascertained through a rare recessive condition. We report here the third case of chromosome 1 UPD, and the first UPD to be ascertained inadvertently during a genome-screen linkage study. All three reports suggest that there are no imprinted genes on chromosome 1 with a major effect on phenotype.

The origin of UPD lies in meiotic nondisjunction events. UPD can result from nondisjunction during meiosis I or II in one parent, leading to a disomic gamete, followed by fertilization with a gamete nullisomic for that chromosome from the other parent (gamete complementation) or by postzygotic loss of the other parent's chromosome (trisomy rescue) (Engel 1993; Ledbetter and Engel 1995). If the nondisjunction occurs at meiosis I, the uniparental pair of chromosomes will contain the centromeric regions of both of the parent's homologues (primary heterodisomy), whereas if the nondisjunction occurs at meiosis II, the uniparental pair will contain the replicated centromeric region of one of the parent's homologues (primary isodisomy). Exchanges during meiosis I can introduce regions of homozygosity (secondary isodisomy) into a primary heterodisomy situation and, conversely, regions of heterozygosity (secondary heterodisomy) into a primary isodisomy situation. In addition to meiosis I and II errors, a third mechanism leading to UPD occurs when a normal monosomic gamete is fertilized by a nullisomic gamete, followed by postzygotic duplication of the single monosomic homologue (monosomy duplication)—this results in complete chromosome isodisomy, including the centromere, with no regions of heterozygosity (Engel 1993). Thus, centromeric heterodisomy (heterozygous markers) indicates a meiosis I error, whereas centromeric isodisomy (homozygous markers) indicates either a meiosis II error if there are other regions showing heterozygosity or postzygotic duplication if all other regions are homozygous. Since the homozygosity associated with UPD, generated either by primary or secondary isodisomy, consists of duplicate copies of alleles from a single chromosome, it carries an increased risk of homozygosity for deleterious recessive genes. Indeed, the presence of a recessive disease in the offspring has been the mode of ascertainment of many examples of UPD (reviewed in Pulkkinen et al. 1997). Similarly, if a chromosome carries imprinted genes, so that one active allele at the imprinted locus is necessary for normal growth and development of the embryo, UPD may be associated with intrauterine growth retardation and other developmental abnormalities (reviewed in Hall 1990; Ledbetter and Engel 1995). However, since the advent of comprehensive genome-wide genotyping for purposes of genetic linkage analysis, the possibility now exists that phenotypically "invisible" cases of UPD, not ascertained through recessive disease or through imprinting-associated abnormalities, will be discovered.

We have been performing genome screening of families having at least two children affected with type 1 (insulin-dependent) diabetes, in order to identify by linkage analysis genes predisposing to this disorder (Field et al. 1994, 1996). A subset of 77 families including 203 children and all their parents has been typed for 187 markers across all chromosomes. During the course of these studies, family BD94 (DNA obtained from the British Diabetes Association Warren Repository [Bain et al. 1990]) was noted to produce numerous marker-typing incompatibilities between the second diabetic child and her father. Closer inspection revealed that the incompatibilities between the father and the second child only involved some of the 14 marker loci typed on chromosome 1, whereas genotyping at 173 microsatellite loci on chromosomes 2 through X (multiple markers on all chromosomes) produced no incompatibilities, proving conclusively that the putative father was the biological

father. An additional 15 markers on chromosome 1 were then genotyped for all family members, and further clinical details about the family, particularly the second child, were obtained following a separate informed consent. Table 1 shows the results of typing 29 chromosome 1 markers and the human leukocyte antigen (HLA) types provided by the BDA. For simplicity, genotypes are shown as recoded alleles, with the mother's alleles and then the father's alleles numbered from smallest to largest and with alleles of identical size receiving the same number code (for example, at D1S159, the mother is 145/147, the father 147/149, the first child 147/149, and the second child 145/145). Markers are listed from pter to qter, with positions on the female genetic map indicated in centimorgans according to information from the Marshfield Center for Medical Genetics Website.

Of the 29 chromosome 1 markers, 16 markers, distributed across the entire chromosome, show incompatibility (indicated in table 1) between the father and the second diabetic child, labeled "Child2." For all 29 markers, the second child's genotype is either identical to the mother's genotype or (in a small region on the short arm) shows only a single allele found in the mother. For the latter cases, if the mother is heterozygous but the child is homozygous, then maternal isodisomy is present (indicated in table 1). The centromeric region is heterodisomic. This pattern is consistent with maternal uniparental primary heterodisomy (arising from nondisjunction during meiosis I), with an embedded region of homozygosity (secondary isodisomy) on the short arm created by a double exchange event. The isodisomic region within the double exchange includes markers

Table 1

Results of Typing 29 Chromosome 1 Microsatellites and Chromosome 6 HLA Loci

Marker or Status	Cytogenetic Location	Genetic Location (Female cM)	Mother	Father	Child1	Child2
D1S468	...	4.5	1,2	1,1	1,2	1,2
D1S1612	...	17.8	1,2	3,4	2,4	1,2[a]
D1S1368	...	...	1,2	1,3	1,2	1,2
D1S1622	...	68.5	1,1	2,3	1,2	1,1[a]
D1S186	...	84.6	1,2	3,4	1,4	1,2[a]
D1S2134	...	100	1,2	2,2	1,2	1,2
D1S405	...	117	1,1	1,1	1,1	1,1
D1S3728	...	122	1,1	2,2	1,2	1,1[a]
D1S198	p32-p33	132	1,2	3,4	2,4	1,2[a]
D1S159	p32	...	1,2	2,3	2,3	1,1[a,b]
D1S410	...	135	1,1	1,2	1,2	1,1
D1S1665	...	137	1,2	1,3	1,1	2,2[a,b]
D1S550	...	...	1,2	2,3	2,2	1,1[a,b]
D1S1728	...	144	1,2	2,3	1,3	2,2[b]
D1S551	...	151	1,1	1,2	1,1	1,1
D1S1159	...	151	1,2	2,3	1,3	2,2[b]
D1S116	p21-p31	...	1,1	1,2	1,1	1,1
D1S1588	...	167	1,2	3,4	2,4	1,2[a]
AMY2B	p21	...	1,2	1,3	1,1	1,2
D1S1631	...	177	1,2	2,3	1,2	1,2
D1S305	...	210	1,1	2,3	1,3	1,1[a]
APOA2	q21-q23	227	1,2	3,4	1,4	1,2[a]
D1S1589	...	245	1,2	1,3	2,3	1,2
D1S117	q23-q25	...	1,2	3,3	1,3	1,2[a]
D1S1660	...	271	1,2	3,4	2,3	1,2[a]
GATA124F08	...	...	1,2	1,1	1,2	1,2
D1S213	q32-q44	312	1,2	3,4	2,4	1,2[a]
D1S103	q32-q44	317	1,2	3,4	2,4	1,2[a]
D1S547	...	351	1,2	3,4	2,4	1,2[a]
HLA-A			1,2	3,31	1,31	1,31
HLA-B			8,62	65,60	8,60	8,60
HLA-C			7,3	8,3	7,3	7,3
HLA-DRB			3,4	13,4	3,4	3,4
HLA-DQB			2,3	1,8	2,8	2,8
+ = high risk HLA haplotype			+,+	−,+	+,+	+,+
Type 1 diabetes present			Yes	No	Yes	Yes

[a] Incompatibility with father.

[b] Demonstrable maternal isodisomy.

D1S159, D1S410, D1S1665, D1S550, D1S1728, D1S551, D1S1159, and possibly D1S116 (the mother is uninformative for the latter), which have all been cytogenetically localized between 1p21 and 1p32. Advanced maternal age is often associated with increased risk of nondisjunction, but this is not relevant in the present study, since the mother was 21 years old at the time of the birth of her second child.

The region of homozygosity encompassed by the two recombination events appears to be quite small: the estimated genetic distance between D1S159 and D1S1159 is 16–35 cM (see table 1: 151 − 135 = 16, and 167 − 132 = 35) in a total female-chromosome length of ~365 cM, according to the Marshfield maps. The other case of maternal chromosome 1 UPD primary heterodisomy also shows only a single region of secondary isodisomy (~35 cM on the long arm), created by a double meiotic exchange event (Pulkkinen et al. 1997). It is possible that unusual recombination patterns (e.g., decreased number of chiasmata or closely adjacent chiasmata) predispose to nondisjunction in meiosis I and thus increase the probability of UPD (Koehler et al. 1996). Alternatively, possession of larger regions of homozygosity in heterodisomic UPD zygotes would increase the risk of recessive lethal conditions, so that these zygotes may be selected against early in development. However, it also is possible that the actual number of detected exchanges (i.e., two) may not be particularly unusual. The expected number of chiasmata occurring between chromatids of paired homologues for a chromosome 365 cM long, which is the size of chromosome 1, is on average seven. We have calculated (on the basis of probabilities from table 2 in Robinson et al. 1993) that the chance of observing ≤2 transitions in a UPD zygote, when seven chiasmata have occurred during meiosis, is 8.6%. (The term "exchange" refers to a chiasma that has occurred in the meiosis I tetrad, whereas "transition" refers to a transition from heterodisomy to isodisomy, or vice versa, in a disomic gamete.) The probability of observing ≤2 transitions would be even higher if there was incomplete marker coverage such that a transition event could be missed (which is possible in the present study) and/or if 365 cM is an overestimate of the true map length due to typing errors (genetic maps are commonly inflated for this reason), so that the expected number of chiasmata is actually less than seven. The reason that so few transitions might be observed, even if as many as seven chiasmata have taken place, is that for a transition to be observable by extensive marker typing in a UPD zygote, the exchange event must occur between a transmitted and a nontransmitted chromatid (i.e., about half of exchanges result in potentially observable transitions, when random involvement of chromatids in chiasmata formation is assumed). Furthermore, for a transition to be observable, the mother must be heterozygous for one or more markers proximal to the exchange. Thus, although it may seem that few exchanges have occurred during the meiosis I event leading to this zygote with chromosome 1 UPD, the actual number of transitions is not significantly different from the expected number.

Trisomy 1 conceptuses have not been observed in spontaneous abortions (Hassold et al. 1996), except for one report of a lost pregnancy with no fetal development (Hanna et al. 1997), or among cases of prenatally diagnosed placental or fetal mosaicism (Ledbetter et al. 1992; Teshima et al. 1992; Hahnemann and Vejerslev 1997). To our knowledge, there are only two reports of trisomy 1 mosaicism in humans (outside of cancer cells) (Neu et al. 1988; Howard et al. 1995). However, molecular studies to determine the origin of the trisomy were not performed in either case, and in at least one case both monosomy and trisomy 1 cells were present, indicating that the trisomy arose as a somatic event during development (Neu et al. 1988). On the other hand, sperm or oocytes aneuploid for chromosome 1 are not uncommon (Martin et al. 1991, 1995; Spriggs et al. 1996). This suggests that trisomy 1 conceptuses occur but die prior to implantation. Thus, the finding of chromosome 1 UPD of maternal meiotic origin is most likely due to a gamete complementation mechanism (fertilization of a disomic egg with a sperm nullisomic for chromosome 1) rather than a trisomy-rescue mechanism (postzygotic loss of the father's chromosome 1), unless the trisomy rescue occurred in the first one or two cell divisions with complete selection against the trisomic cells.

The mother and both of the two children in this family have type 1 diabetes, and all three individuals have HLA genotypes associated with a high risk of developing diabetes (see table 1). It is well established that the HLA region contains the strongest susceptibility genes for this disease (for a review of insulin-dependent diabetes mellitus [IDDM] genetics, see Field and Tobias 1997). Thus, we assume that the presence of chromosome 1 UPD in one of the diabetic children is unrelated to her IDDM. Apart from her diabetes, she has no other unusual conditions. There was no evidence of dysmorphic features at birth. She had a full-term birth weight of 2,930 g (consistent with that of her mother and older brother, whose full-term birth weights were 2,840 g and 2,870 g, respectively), with no indication of intrauterine growth retardation. Subsequently (she is now 23 years old), she showed no signs of mental or developmental retardation or precocious puberty.

In the two other cases of chromosome 1 UPD (Pulkkinen et al. 1997; Gelb et al. 1998), ascertainment was through a rare recessive disorder, but there were no features suggestive of imprinting, such as growth or developmental abnormalities. However, since the infant with maternal chromosome 1 UPD died at 2 mo of age

(Pulkkinen et al. 1997), the present case of maternal chromosome 1 UPD in a developmentally normal adult provides valuable additional evidence that there are no imprinted genes on chromosome 1 with major phenotypic effects. This has potential implications for prenatal diagnosis if chorionic villus sampling (CVS) reveals trisomy mosaicism and later amniotic fluid sampling shows fetal disomy (apparent trisomy rescue), since these cases theoretically have a one in three risk of UPD for the relevant chromosome and any associated imprinting effects (Ledbetter and Engel 1995). However, as discussed above, it is probable that conceptuses trisomic for chromosome 1 die before implantation and therefore are unlikely to be detected by CVS.

The data presented here, combined with that from other reports of UPD (Jones et al. 1995; Ledbetter and Engel 1995), suggest that, in the absence of isodisomy for recessive deleterious genes, uniparental disomy for chromosomes that do not harbor imprinted loci may be quite harmless. If so, it would be of interest to know the frequency of this phenomenon in the normal general population. In our laboratory, we have typed >200 children (and their parents) for markers relatively densely distributed across the genome, and this is the first case of UPD that we have recognized. Other laboratories performing large-scale linkage-mapping projects may encounter UPD but may attribute it to lab typing errors, null alleles, or nonpaternity. The possibility of UPD should be considered when typing incompatibilities occur repeatedly for the same family in genome-screen projects, since such studies represent an important source for discovery of additional cases of UPD with no apparent phenotypic effects.

Acknowledgments

We thank the members of family BD94 for their generous participation. BD94 was made available, by the British Diabetic Association (BDA), from the BDA–Warren Repository of multiplex families with type 1 diabetes. We also thank E. Swiergala for her skillful laboratory assistance. This research was funded by grants to L.L.F. from the Medical Research Council of Canada (MT-7910) and the Network of Centres of Excellence Programme of the Canadian government. L.L.F. is an Alberta Heritage Medical Scientist.

L. Leigh Field,[1] Rose Tobias,[1] Wendy P. Robinson,[2] Richard Paisey,[3] and Stephen Bain[4]

[1]*Department of Medical Genetics, University of Calgary, Calgary;* [2]*Department of Medical Genetics, University of British Columbia, Vancouver;* [3]*Torbay Hospital, Torquay, United Kingdom; and* [4]*Department of Medicine, University of Birmingham, Birmingham, United Kingdom*

Electronic-Database Information

URL for data in this article is as follows:

Marshfield Center for Medical Genetics, http://www.marshmed.org/genetics (for marker mapping information)

References

Bain SC, Todd JA, Barnett AH (1990) The British Diabetes Association–Warren Repository. Autoimmunity 7:83–85

Engel E (1993) Uniparental disomy revisited: the first twelve years. Am J Med Genet 46:670–674

Field LL, Tobias R (1997) Unravelling a complex trait: the genetics of insulin-dependent diabetes mellitus. Clin Invest Med 20:41–49

Field LL, Tobias R, Magnus T (1994) A locus on chromosome 15q26 (*IDDM3*) produces susceptibility to insulin-dependent diabetes mellitus. Nat Genet 8:189–194

Field LL, Tobias R, Thomson G, Plon S (1996) Susceptibility to insulin-dependent diabetes mellitus maps to a locus (*IDDM11*) on human chromosome 14q24.3-q31. Genomics 33:1–8

Gelb BD, Willner JP, Dunn TM, Kardon NB, Verloes A, Poncin J, Desnick RJ (1998) Paternal uniparental disomy for chromosome 1 revealed by molecular analysis of a patient with pycnodysostosis. Am J Hum Genet 62:848–854

Hahnemann JM, Vejerslev LO (1997) European Collaborative Research on Mosaicism in CVS (EUCROMIC): fetal and extrafetal cell lineages in 192 gestations with CVS mosaicism involving single autosomal trisomy. Am J Med Genet 70: 179–187

Hall JG (1990) Genomic imprinting: review and relevance to human diseases. Am J Hum Genet 46:857–873

Hanna JS, Shires P, Matile G (1997) Trisomy-1 in a clinically recognized pregnancy. Am J Med Genet 68:98

Hassold T, Abruzzo M, Adkins K, Griffin D, Merrill M, Millie E, Saker D, et al (1996) Human aneuploidy: incidence, origin, and etiology. Environ Mol Mutagen 28:167–175

Howard PJ, Cramp CE, Fryer AE (1995) Trisomy 1 mosaicism only detected on a direct chromosome preparation in a neonate. Clin Genet 48:313–316

Jones C, Booth C, Rita D, Jazmines L, Spiro R, McCulloch B, McCaskill C, et al (1995) Identification of a case of maternal uniparental disomy of chromosome 10 associated with confined placental mosaicism. Prenat Diagn 15:843–848

Koehler KE, Hawley RS, Sherman S, Hassold T (1996) Recombination and nondisjunction in humans and flies. Hum Mol Genet 5:1495–1504

Ledbetter DH, Engel D (1995) Uniparental disomy in humans: development of an imprinting map and its implications for prenatal diagnosis. Hum Mol Genet 4:1757–1764

Ledbetter DH, Zachary JM, Simpson JL, Golbus MS, Pergament E, Jackson L, Mahoney MJ, et al (1992) Cytogenetic results from the US Collaborative Study on CVS. Prenat Diagn 12:317–345

Martin RH, Ko E, Rademaker AW (1991) Distribution of aneuploidy in human gametes: comparison between human sperm and oocytes. Am J Med Genet 39:321–331

Martin RH, Spriggs E, Ko E, Rademaker AW (1995) The re-

lationship between paternal age, sex ratios, and aneuploidy frequencies in human sperm, as assessed by multicolor FISH. Am J Hum Genet 57:1395–1399

Neu RL, Kouseff BG, Madan S, Essig Y-P, Miller K, Tedesco TA (1988) Monosomy, trisomy, fragile sites, and rearrangements of chromosome 1 in a mentally retarded male with multiple congenital anomalies. Clin Genet 33:73–77

Pulkkinen L, Bullrich F, Czarnecki P, Weiss L, Uitto J (1997) Maternal uniparental disomy of chromosome 1 with reduction to homozygosity of the LAMB3 locus in a patient with Herlitz junctional epidermolysis bullosa. Am J Hum Genet 61:611–619

Robinson WP, Bernasconi F, Mutirangura A, Ledbetter DH, Langlois S, Malcolm S, Morris MA, et al (1993) Nondisjunction of chromosome 15: origin and recombination. Am J Hum Genet 53:740–751

Spriggs EL, Rademaker AW, Martin RH (1996) Aneuploidy in human sperm: the use of multicolor FISH to test various theories of nondisjunction. Am J Hum Genet 58:356–362

Teshima IE, Kalousek DK, Vekemans MJ, Markovic V, Cox DM, Dallaire L, Gagne R, et al (1992) Canadian multicenter randomized clinical trial of chorion villus sampling and amniocentesis: chromosome mosaicism in CVS and amniocentesis samples. Prenat Diagn 12:443–466

Address for correspondence and reprints: Dr. L. Leigh Field, Health Sciences Centre, 3330 Hospital Drive NW, Calgary, Alberta T2N 4N1, Canada. E-mail: field@acs.ucalgary.ca

0002-9297/98/6304-0036$02.00

Am. J. Hum. Genet. 63:1220–1224, 1998

Low-Penetrance Branches in Matrilineal Pedigrees with Leber Hereditary Optic Neuropathy

To the Editor:

Leber hereditary optic neuropathy (LHON; MIM 535000) is an inherited form of bilateral optic atrophy in which the primary etiologic event is a mutation in the mitochondrial genome (reviewed by Riordan-Eva et al. 1995; Nikoskelainen et al. 1996; Howell 1997*a*, 1997*b*). It has been recognized, since the earliest studies of LHON (Leber 1871), that the penetrance is incomplete. It is now understood that this incomplete penetrance reflects a complex etiology and that multiple secondary factors modify or determine the manifestation of the optic neuropathy in LHON (reviewed by Howell 1997*a*, 1997*b*).

The identification of these secondary etiologic factors has been difficult, but heavy smoking and alcohol consumption have received epidemiological support (e.g., see Johns 1994). It appears, however, that there are numerous, but poorly defined, physiological, environmental, societal, and demographic "life style" factors that modify the risk of optic neuropathy. For example, there has been a relatively recent (i.e., during the second half of this century) parallel decline in penetrance among Australian LHON families and in the incidence of a pathologically similar, *acquired* optic-nerve disorder, tobacco-nutritional amblyopia. This trend suggests that there is a common factor in their etiology (Mackey and Howell 1994). In addition, penetrance in LHON families from different northern European countries varies more than twofold (e.g., see Mackey et al. 1996). Even within a single country, such as Australia, there are substantial penetrance differences among 11778 LHON families (Howell et al. 1993).

We have been analyzing penetrance in large, multigeneration Australian and British LHON families, as one approach to the elucidation of these secondary etiologic factors. A previously undescribed pattern of results was obtained during this survey, and we describe here the occurrence of distinct low- (and high-) penetrance branches in LHON pedigrees.

The TAS1 LHON family is the largest matrilineal pedigree that has been assembled. It spanned 11 generations by the early 1990s, and it now comprises >1,600 maternally related individuals, all of whom are descended from a woman who was born in 1777 (III-12 in fig. 1; also see Mackey and Buttery 1992). This LHON family carries the primary mutation at nucleotide 11778 of the mitochondrial ND4 gene (Mackey 1994). Because this family has been located within a relatively small geographical area, because of the good clinical record keeping, and because of the high level of compliance and cooperation on the part of the family, we are confident about the identification of affected and unaffected family members. However, there is an inherent uncertainty in all studies of LHON penetrance, which results from the variable and unpredictable age at onset, spanning the 1st through 8th decades, with a mean in the mid 20s (e.g., see Riordan-Eva et al. 1995; Nikoskelainen et al. 1996). Therefore, LHON carriers (especially males) are always at risk, and there is no age at which one can state with absolute confidence that a family member will remain unaffected.

To control, as much as possible, for the confounding factors in the analysis of penetrance, we have applied the following guidelines. In the first place, we limited our penetrance calculations to *males* who were >30 years of age, to include only those individuals who were past the age of maximum risk. The number of affected females is generally too low, even in the largest LHON families, to provide robust information on penetrance, and they were excluded from the present study. Second, we define here "affected" and "unaffected" in terms of a significant vision loss whose characteristics are compatible with LHON. There are subtle, subclinical

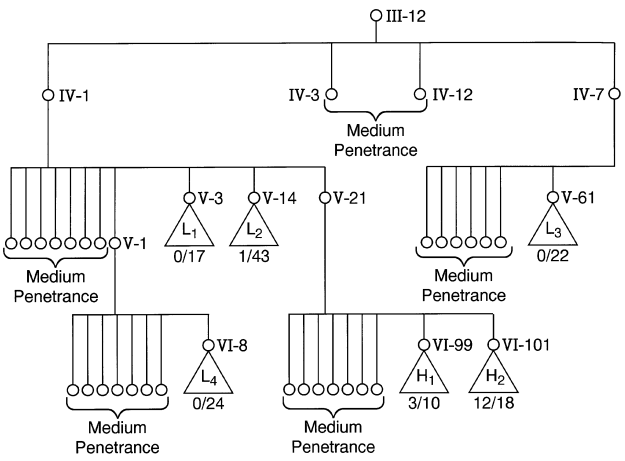

Figure 1 High- and low-penetrance branches in the TAS1 11778-mutation LHON family. This partial pedigree has been drawn to show the genealogical origin of the branches in which there is an unusually low (L1–L4) or high (H1 and H2) penetrance of the optic neuropathy in male family members. Some of the female origins of family branches are shown, with pedigree designations that follow standard numbering schemes (in which the generation designation is denoted by a Roman numeral). The fractions beneath the low- and high-penetrance branches refer to the number of affected (numerator) and total (denominator) males within that particular branch.

changes in the eye (most prevalently, a microangiopathy; see the discussion in Riordan-Eva et al. 1995; Nikoskelainen et al. 1996) that are found at high frequency in LHON family members, but these are not considered here. Ongoing clinical studies of the TAS1 LHON family give no indication that the present results are biased by a high frequency of atypical or unreported ophthalmological abnormalities. Finally, significant recovery of vision is very rare in 11778-mutation LHON patients (reviewed in Howell 1997*a,* 1997*b*), and there is no indication that the penetrance frequencies in the TAS1 LHON family have been biased by this phenomenon.

Analysis of the TAS1 pedigree revealed that there are low-penetrance four branches (designated "L1"–"L4"), in which the penetrance of the optic neuropathy has essentially dropped to zero (fig. 1). A branch is defined here as the descendants of any female in a matrilineal pedigree; the descendants span at least four generations, to provide sufficient information for the determination of penetrance. There is only a single affected male among the L1, L2, L3, and L4 branches, which include a total of 17, 43, 22, and 24 males, respectively. This individual lost vision soon after suffering head trauma in an automobile accident, a severe precipitating factor. For comparison, we ascertained the penetrance in the more typical (designated here as "medium-penetrance") branches of the pedigree. Whereas the L2 branch (which starts with female VI-18) contained 1 affected male among a total of 43, there were 9 affected males, among a total of 53, in the branch that descended from female V-7 and that spans generations VI–IX (this female is not designated in fig. 1). This difference in penetrance frequencies is statistically significant ($P < .05$; 2×2 χ^2 test, adjusted for continuity). These statistical tests must be treated with caution, however, because it is difficult to rule out post hoc bias in the identification of low-penetrance branches. We attempted to address this concern by further analysis of penetrance in the TAS1 pedigree. Thus, the L4 branch is one of several branches that descend from female V-1, and the penetrance is ~12% among males in the other branches that descend from her. In a similar fashion, the penetrance among the descendants of females V-18 and V-57 is 15% and 16%, respectively (these females are not designated in fig. 1). Female V-21 gave rise to two branches if one distinguishes the descendants from her two marriages, and the approximate penetrance values are 33% (which includes the H1 and H2 branches; see fig. 1 and the results given below) and 20%. Therefore, in the comparison of branches of similar size, the low-penetrance branches stand out clearly, a result that argues against severe bias.

The evidence for low-penetrance branching is further supported when the results for all four branches are pooled and the results are compared with the overall penetrance in the matrilineal pedigree. Thus, there is 1 affected male among the total of 106 males in the four branches (a penetrance of ~1%), whereas there are ~200 affected males among a total of ~800 in the medium- and high-penetrance branches of the TAS1 pedigree (an overall penetrance of ~25%). The actual difference in penetrance values is larger, because the estimate of 25% is not adjusted upward to account for those males who are <30 years of age.

There may also be high-penetrance branches, although, because of the small number of family members in these branches, this possibility is less robust. There are in the TAS1 family two small branches (designated "H1" and "H2") in which the penetrance was unusually high. Thus, in the small H2 branch, 12 (67%) of 18 males were affected. Only 3 (30%) of 10 males were affected in branch H1, but 5 (25%) of 20 females were also affected.

The most obvious explanation for the low-penetrance branches in the TAS1 pedigree is heteroplasmy of the 11778 mutation in the early generations. The pathogenic mutation could have segregated into both homoplasmic mutant and homoplasmic wild-type branches (this situation has occurred in the QLD2 11778 LHON family, as described in Howell et al. 1995, p. 298). To test this possibility, we analyzed DNA from seven members of

low-penetrance branches and from two members of high-penetrance branches. In brief, our approach involves both PCR amplification of short (300–400 bp) spans of the mitochondrial genome and subsequent sequencing analysis of multiple, independent M13 clones that contain the mtDNA insert (e.g., see Howell et al. 1991, 1995). For these nine TAS1 LHON family members, the DNA sequences of >400 independent mtDNA inserts that contained a short segment of the ND4 gene were determined. It was found that *all* of them carried the 11778 mutant allele. Furthermore, restriction-site assays of another 40 TAS1 LHON family members have confirmed that the 11778 primary mutation is homoplasmic in all family members (data not shown). Furthermore, tissue-distribution studies indicate that mutation load in blood either reflects the levels in other tissues (Juvonen et al. 1997) or is *lower* than those in other tissues (Howell et al. 1994). Thus, the cumulative results show that the low penetrance in some branches of the TAS1 family is not due to segregational loss (or reversion) of the 11778 mutant allele.

We then extended the sequencing analysis to search for a second site, or suppressor, mitochondrial gene mutation. Family members from the low-penetrance branches may carry a secondary mutation that phenotypically suppresses the pathogenic effects of the 11778 mutation. For example, a suppressor mutation might have arisen in a common maternal ancestor, persisted in the heteroplasmic state for several generations, and eventually become fixed in some branches of the matrilineal pedigree, but not in others, as a result of segregation in the germ line. There are results that suggest the occurrence of mitochondrial suppressor mutations. Thus, the QLD1 LHON family carries, at nucleotide 4160 of the ND1 gene, a mutation that is associated with the severe neurological abnormalities (Howell 1994). A putative intragenic suppressor mutation at nucleotide 4136 has arisen in one small branch (Howell et al. 1991). In addition, Hammans et al. (1995) and El Meziane et al. (1998) have reported suppressor mutations of a pathogenic tRNA mutation.

Six overlapping, PCR-amplified fragments of the mitochondrial genome, which cumulatively spanned nucleotides 10435–12373 (numbered according to Anderson et al. 1981), were analyzed for each of the five TAS1 LHON family members. This 1.9-kb span of the mtDNA included the 3′ half of the $tRNA^{Arg}$ gene, the ND4L gene (nucleotides 10470–10763), the ND4 gene (nucleotides 10760–12137), a cluster of three butt-joined tRNA genes ($tRNA^{His}$, $tRNA^{Ser[AGY]}$, and $tRNA^{Leu[CUN]}$), and the first 36 nucleotides of the ND5 gene. Multiple ($\geqslant$10) independent clones were sequenced for each of the six mtDNA fragments and for each family member, in an effort to detect heteroplasmic mutations. No new sequence changes were detected in any of the five family members. The sequence of this span of the mitochondrial genome was identical for all family members, including the presence of a rare, silent polymorphism at nucleotide 11788. Among the >200 pedigrees (control and LHON) that we have screened, this polymorphism thus far is unique to the TAS1 LHON family, and we have thus verified that the members of the low-penetrance branches are indeed of the correct maternal lineage.

Finally, we have begun a wider search for an *intergenic* mitochondrial suppressor mutation. The first fragment that we analyzed, which spanned nucleotides 3286–3564, included the site of the primary LHON mutation, at nucleotide 3460; the second fragment that we analyzed, which spanned nucleotides 4027–4294, included the sites of both the pathogenic mutation, at nucleotide 4160, and the putative suppresser mutation, at nucleotide 4136, as well as that of the putative secondary LHON mutation, at nucleotide 4216 (Johns and Berman 1991); the third fragment that we analyzed, which spanned nucleotides 14381–14699, included the site of the primary LHON mutation, at nucleotide 14484, and several other sites at which pathogenic mutations have been identified (see the discussion in Howell et al. 1998). The TAS1 mtDNA does not carry any of the aforementioned "accessory" LHON mutations, and there were *no* new mutations in these regions of the mitochondrial genome, among any of the low- and high-penetrance family members who were analyzed.

In addition to the results for the TAS1 LHON family, there are other examples of low-penetrance branches in LHON families. We have also observed that low-penetrance branches apparently occur in the large 14484-mutation TAS2 LHON family (D. A. Mackey, unpublished data), which comprises ~700 maternally related individuals (Mackey and Buttery 1992). As one example, none of the 28 males ($\geqslant$30 years of age) who have descended from female VII-22 have lost vision (authors' unpublished data). We are continuing our analysis of the TAS2 pedigree, because penetrance in 14484-mutation LHON families is more difficult to quantitate with acceptable certainty, because of the high frequency of vision recovery. It becomes more difficult to distinguish a true lack of vision loss from a mild vision loss and rapid recovery, particularly when one must rely, in part, on second-hand information about vision status in relatives. Inspection of pedigree data in the literature also suggests the presence of low-penetrance branches that have been unremarked until now (see, especially, pedigrees XX and XXVIII in van Senus 1963).

Overall, therefore, it appears that low-penetrance branching in LHON matrilineal pedigrees is a biologically "real" phenomenon. One explanation is that the low-penetrance branches are real but that there are dif-

ferent epigenetic and/or environmental factors that lower the penetrance in each branch. Alternatively, low-penetrance branching may be due to the introduction of a *nuclear* genetic suppressor locus. This explanation, however, is problematic, because each low-penetrance branch involves a number of outbreeding events (i.e., marriages), which should act to "localize" any effects of a dominantly acting nuclear locus to one or two generations. Third, low-penetrance branching may be caused by a *mitochondrial* suppressor locus, but one that lies in a mitochondrial genome region that was not sequenced in the experiments that are reported here. Thus far, we have sequenced (*a*) only approximately one-third of the mitochondrial genome that encodes the seven subunits of complex I (NADH-ubiquinone oxidoreductase) or (*b*) only approximately one-fifth of the entire coding region.

The suggestion of a mitochondrial mutation that decreases penetrance in the TAS1 LHON family converges with the related issue of phylogenetic clustering. Both the 11778 mutation and, especially, the 14484 LHON mutation occur more often in European haplogroup J mtDNA backgrounds than would be expected on a random basis (although the TAS1 mtDNA haplotype does not belong to this haplogroup). There is debate over the basis of this clustering phenomenon (see the discussion in Howell et al. 1995 and Mackey et al. 1998), but Brown et al. (1997) and Torroni et al. (1997) have concluded that LHON penetrance is influenced by the mtDNA background in which the pathogenic mutations arise. Thus, the apparent underrepresentation of some mtDNA haplotypes among LHON patients is caused by low penetrance, because of one or more sequence changes within these mtDNAs. As a consequence of the lower penetrance, fewer pedigrees come to the attention of clinicians. Within the haplotype J mtDNA, the site(s) that influences penetrance has not been identified, but the basic premise is similar to that proposed here to explain the presence of low-penetrance branches within a single LHON pedigree. In summary, the present results underscore both the complex etiology of LHON and the fact that the identification of the secondary etiologic factors is a prerequisite for a further understanding of this disorder.

Acknowledgments

We gratefully acknowledge the cooperation and assistance of the members of the TAS1 LHON family. Technical assistance was provided by Iwona Kubacka and Steven Halvorson. This research was funded by National Eye Institute grant EY10758 and a John Sealy Endowment Fund grant (both to N.H.). D.A.M. acknowledges the support of the Clifford Craig Memorial Research Trust.

NEIL HOWELL[1] AND DAVID A. MACKEY[2]
[1]*Departments of Radiation Oncology and Human Biological Chemistry and Genetics, The University of Texas Medical Branch, Galveston; and* [2]*Departments of Ophthalmology and Paediatrics, The University of Melbourne, Melbourne, and Menzies Centre for Population Health Research, The University of Tasmania, Hobart*

Electronic-Database Information

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

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim (for LHON [MIM 535000])

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

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

El Meziane A, Lehtinen SK, Hance N, Nijtmans LGJ, Dunbar D, Holt IJ, Jacobs HT (1998) A tRNA suppressor mutation in human mitochondria. Nat Genet 18:350–353

Hammans SR, Sweeney MG, Hanna MG, Brockington M, Morgan-Hughes JA, Harding AE (1995) The mitochondrial DNA transfer RNALeu(UUR) A→G(3243) mutation. A clinical and genetic study. Brain 118:721–734

Howell N (1994) Primary LHON mutations: trying to separate "fruyt" from "chaf." Clin Neurosci 2:130–137

——— (1997*a*) Leber hereditary optic neuropathy: how do mitochondrial DNA mutations cause degeneration of the optic nerve? J Bioenerg Biomembr 29:165–173

——— (1997*b*) Leber hereditary optic neuropathy: mitochondrial mutations and degeneration of the optic nerve. Vision Res 37:3495–3507

Howell N, Bogolin C, Jamieson R, Marenda DR, Mackey DA (1998) mtDNA mutations that cause optic neuropathy: how do we know? Am J Hum Genet 62:196–202

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

Howell N, Kubacka I, Halvorson S, Mackey D (1993) Leber's hereditary optic neuropathy: the etiological role of a mutation in the mitochondrial cytochrome *b* gene. Genetics 133:133–136

Howell N, Kubacka I, Xu M, McCullough DA (1991) Leber

hereditary optic neuropathy: involvement of the ND1 gene and evidence for an intragenic suppressor mutation. Am J Hum Genet 48:935–942

Howell N, Xu M, Halvorson S, Bodis-Wollner I, Sherman J (1994) A heteroplasmic LHON family: tissue distribution and transmission of the 11778 mutation. Am J Hum Genet 55:203–206

Johns DR (1994) Genotype-specific phenotypes in Leber's hereditary optic neuropathy. Clin Neurosci 2:146–150

Johns DR, Berman J (1991) Alternative simultaneous complex I mitochondrial DNA mutations in Leber's hereditary optic neuropathy. Biochem Biophys Res Commun 174: 1324–1330

Juvonen V, Nikoskelainen E, Lamminen T, Penttinen M, Aula P, Savontaus M-L (1997) Tissue distribution of the ND4/ 11778 mutation in heteroplasmic lineages with Leber hereditary optic neuropathy. Hum Mutat 9:412–417

Leber T (1871) Über hereditäre und congenital-angelegte Sehnervenleiden. Graefes Arch Clin Exp Ophthalmol 17 (part 2): 249–291

Mackey DA (1994) Epidemiology of Leber's hereditary optic neuropathy in Australia. Clin Neurosci 2:162–164

Mackey DA, Buttery RG (1992) Leber hereditary optic neuropathy in Australia. Aust NZ J Ophthalmol 20:177–184

Mackey DA, Howell N (1994) Tobacco amblyopia. Am J Ophthalmol 117:817–818

Mackey DA, Oostra R-J, Rosenberg T, Nikoskelainen E, Bronte-Stewart J, Poulton J, Harding AE, et al (1996) Primary pathogenic mtDNA mutations in multigeneration pedigrees with Leber hereditary optic neuropathy. Am J Hum Genet 59:481–485

Mackey D, Oostra R-J, Rosenberg T, Nikoskelainen E, Poulton J, Barratt T, Bolhuis P, et al (1998) Reply to Hofmann et al. Am J Hum Genet 62:492–495

Nikoskelainen EK, Huoponen K, Juvonen V, Lamminen T, Nummelin K, Savontaus M-L (1996) Ophthalmologic findings in Leber hereditary optic neuropathy, with special reference to mtDNA mutations. Ophthalmology 103:504–514

Riordan-Eva P, Sanders MD, Govan GG, Sweeney MG, Da Costa J, Harding AE (1995) The clinical features of Leber's hereditary optic neuropathy defined by the presence of a pathogenic mitochondrial DNA mutation. Brain 118: 319–338

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

van Senus AHC (1963) Leber's disease in the Netherlands. Doc Ophthalmol 17:1–162

Address for correspondence and reprints: Dr. Neil Howell, Biology Division 0656, Department of Radiation Oncology, The University of Texas Medical Branch, Galveston, TX 77555-0656. E-mail: nhowell@utmb.edu

0002-9297/98/6304-0038$02.00

Am. J. Hum. Genet. 63:1224–1227, 1998

Double Heterozygotes for the Ashkenazi Founder Mutations in BRCA1 and BRCA2 Genes

To the Editor:

Three Jewish founder mutations, 185delAG and 5382insC in BRCA1 and 6174delT in BRCA2 genes, have been identified in breast cancer (BC) and ovarian cancer (OC) Ashkenazi patients. In the Ashkenazi general population, the carrier frequencies of these founder mutations are 1% for 185delAG (Struewing et al. 1996), 0.13% for 5382insC, and 1.35% for 6174delT (Roa et al. 1996; Oddoux et al. 1996). Given these high population frequencies, one would expect to find individuals homozygous for the mutations 185delAG/185delAG, 6174delT/6174delT, and 5382insC/5382insC, compound heterozygous for 185delAG/5382insC, or double heterozygous for 185delAG/6174delT or 5382insC/ 6174delT, provided the individuals are viable. The effect of two mutations in a single individual is important both for an understanding of the mode of action and interaction between the BRCA1 and BRCA2 genes and for appropriate genetic counseling. To date, two double heterozygous patients (185delAG/6174delT; Ramus et al. 1997; Gershoni-Baruch et al. 1997) and one patient homozygous for a mutation in exon 11 of the BRCA1 gene (Boyd et al. 1995) have been reported.

By pooling results from four cancer/genetics centers in Israel, we have analyzed ~1,500 BC/OC Ashkenazi patients. All subjects received genetic counseling and signed informed consent forms in compliance with institutional ethics committees (institutional review boards). Each patient was tested for the three Ashkenazi founder mutations: in BRCA1, the mutations 185delAG and 5382insC, and in BRCA2, the mutation 6174delT (Abeliovich et al. 1997; Levy-Lahad et al. 1997; Bruchim Bar-Sade et al. 1998). Four patients were found to be double heterozygotes. Summaries of their clinical status and pedigrees are presented in table 1 and figure 1.

Patient 1 is an Ashkenazi mother of two children who was diagnosed with unilateral breast cancer at the age of 38 years. Her family history was positive for both OC, with which her mother was diagnosed at the age of 50 years, and breast cancer, with which her paternal aunt was diagnosed at the age of 60 years and her daughter at the age of 35 years. Her paternal grandfather had lung cancer at the age of 45 years. A test for 185delAG/ 6174delT in her father revealed neither mutation; DNA could not be retrieved from the paraffin block of her mother. Analysis of the polymorphic markers D17S855, D17S1322, D17S1323, D9S55, and D11S1337 in the father and in Patient 1 confirmed paternity. It was thus

Table 1

Genotypes and Clinical Status of the Patients

Individual	Genotype	Clinical Status	Age at Diagnosis (years)
Patient 1	185delAG/6174delT	BC	38
Mother of Patient 1	185delAG/6174delT[a]	OC	50
Patient 2	185delAG/6174delT	OC	57
Patient 3	185delAG/6174delT	Healthy	50
Patient 4	5382insC/6174delT	BC	45

[a] Inferred genotype.

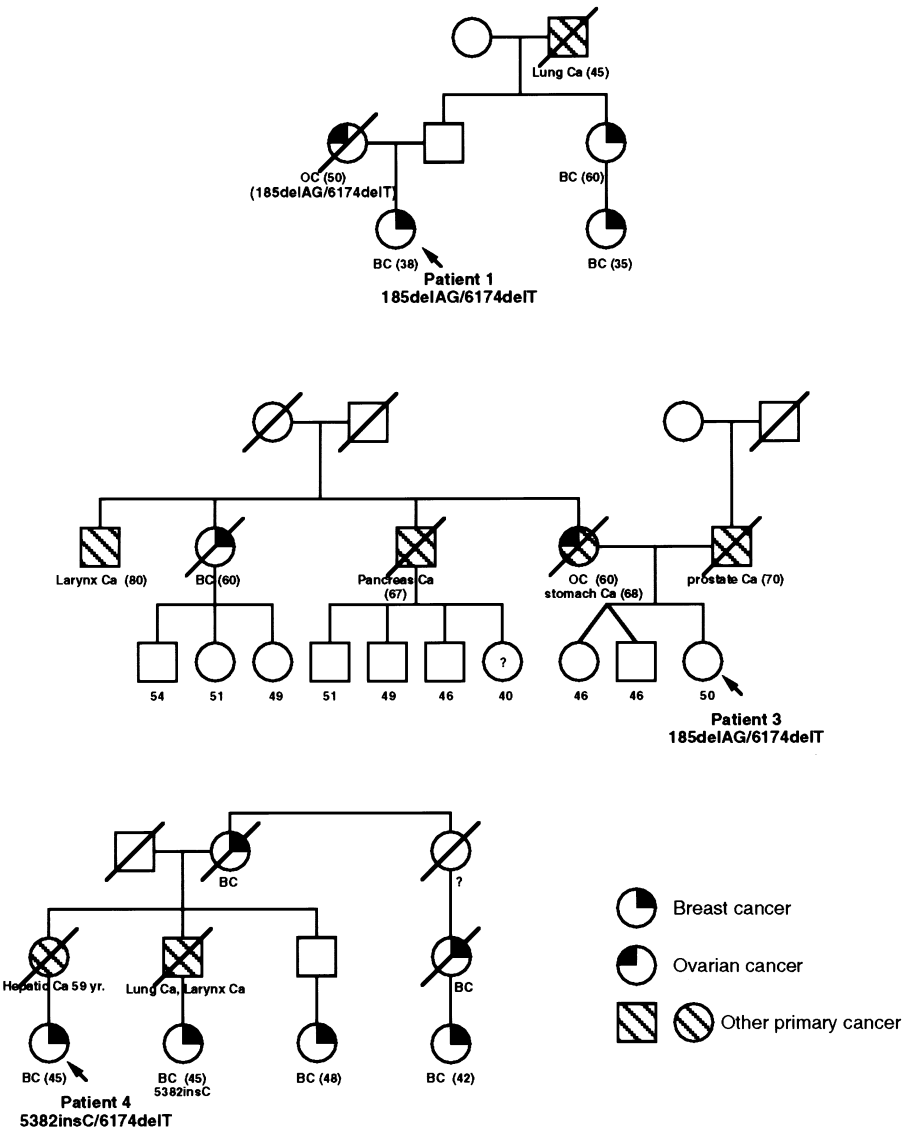

Figure 1 Pedigrees of Patients 1, 3, and 4. In parentheses is the inferred genotype and the ages at diagnosis.

assumed that she had inherited both mutations from her double heterozygous mother.

Patient 2 is a 57-year-old Ashkenazi woman who presented with stage IV OC. The patient is alive with no evidence of disease 5 years after treatment. Her family history includes breast cancer in her mother (age at diagnosis unknown). No further information was available. Patient 2 had irregular menses and primary sterility, which were treated with low doses of steroids.

Patient 3 is a 50-year-old asymptomatic Ashkenazi woman who was referred for evaluation of her breast cancer risk before starting hormonal replacement therapy for increasing loss in bone density. The maternal family history was positive for ovarian, breast, pancreas, stomach, and laryngeal cancers. Her father had prostate cancer. The patient had idiopathic premature menopause at the age of 37 years after bearing three children.

Patient 4 is a 46-year-old Ashkenazi woman who was diagnosed with breast-infiltrating ductal carcinoma. The family history was positive for cancer: hepatic carcinoma at the age of 59 years in her mother and breast cancer in her maternal grandmother. Two of her maternal cousins and two more distant relatives had breast cancer at the ages of 45, 48, and 42 years (the age at diagnosis of one of the relatives is unknown). One of them is a carrier of the mutation 5382insC. The others were not available for mutation analysis.

As compared with carriers of single mutations, the four double heterozygotes we observed did not have a particularly severe phenotype, based on the tumor types and age at diagnosis: one was unaffected at the age of 50 years; two were affected with unilateral breast cancer, one at the age of 38 years and one at the age of 46 years; and one was diagnosed with OC at the age of 57 years. An inferred double heterozygote (the mother of Patient 1) had OC at the age of 50 years. None had more than one primary tumor, and tumor histology and clinical course were unremarkable. Two other 185delAG/6174delT carriers were reported: one had BC and OC diagnosed at the ages of 48 and 50 years, respectively (Ramus et al. 1997); the other had bilateral BC at the ages of 41 and 50 years, respectively (Gersoni-Baruch et al. 1997).

Although the small number of cases precludes definite conclusions, our results suggest that the phenotypic effects of double heterozygosity for BRCA1 and BRCA2 germ-line mutations are not cumulative. This is in agreement with the observation that the phenotype of mice that were homozygote knockouts for the BRCA1 and BRCA2 genes was similar to that of mice that were BRCA1 knockouts. This suggests that the BRCA1 mutation is epistatic over the BRCA2 mutation (Ludwig et al. 1997).

Interestingly, two of the double heterozygotes described have had reproductive problems: one (Patient 2) had primary sterility and irregular menses, and another (Patient 3) had premature menopause at the age of 37 years. This latter patient was asymptomatic at the age of 50 years. These preliminary observations raise the possibility of hormonal effects in double heterozygotes, including the possibility that the lack of estrogen may have a protective effect.

At the population level, given the known heterozygote frequencies in Ashkenazi Jews, the expected frequencies of double heterozygotes would be the multiplication of the heterozygote frequencies 185delAG/6174delT (1.35×10^{-4}) and 5382insC/6174delT (1.75×10^{-5}). The expected frequencies of BRCA1 and BRCA2 homozygotes will be the multiplication of the mutation frequencies (approximately one-half of the heterozygote frequency), which are 2.5×10^{-5} for 185delAG homozygotes and 4.6×10^{-5} for 6174delT homozygotes. Therefore the ratio of 185delAG/6174delT double heterozygotes and 6174delT and 185delGA homozygotes is 3:1:0.5, respectively. Namely, the double heterozygotes should be about three to six times more common than the homozygotes 185delAG or 6174delT. In this respect, we might have expected to observe 185delAG or 6174delT homozygotes. The fact that we did not observe these or any other homozygotes may be due to chance, and more patients should be tested before a homozygous patient is found or, alternatively, before homozygosity for 185delAG or 6174delT decreases viability or causes different phenotypic consequences.

The clinical implication of this study is that mutation analysis in Ashkenazi Jews should include all known founder mutations. Identification of additional carriers of more than one mutation will increase our understanding of the interaction between various mutations and will improve genetic counseling.

Eitan Friedman,[1] Revital Bar-sade Bruchim,[1] Anna Kruglikova,[1] Shulamit Risel,[1] Ephrat Levy-Lahad,[2] David Halle,[3] Elchanan Bar-On,[4] Ruth Gershoni-Baruch,[8] Ephrat Dagan,[8] Ilana Kepten,[8] Tamar Peretz,[5] Israela Lerer,[6] Naomi Wienberg,[6] Asher Shushan,[7] and Dvorah Abeliovich[6]

[1]*The Oncogenetics Unit and Clinical Epidemiology, Chaim Sheba Medical Center, Tel Hashomer; Departments of* [2]*Medicine,* [3]*Oncology, and* [4]*Gynecology, Shaare Zedek Medical Center, and* [5]*Sharett Institute of Oncology, and Departments of* [6]*Human Genetics and* [7]*Obstetrics and Gynecology, Hadassah Hebrew University Hospital, Jerusalem; and* [8]*Genetics Institute, Rambam Medical Center and Bruce Rappoport Faculty of Medicine, Haifa*

References

Abeliovich D, Kaduri L, Lerer I, Weinberg N, Amir G, Sagi M, Zlotogora J, et al (1997) The founder mutations

185delAG and 5382insC in BRCA1 and 6174delT in BRCA2 appear in 60% of ovarian cancer and 30% of early-onset breast cancer patients among Ashkenazi women. Am J Hum Genet 60:505–514

Boyd M, Harris F, McFarlene R, Davidson RH, Black DM (1995) A human BRCA1 gene knockout. Nature 375: 541–542

Bruchim Bar-Sade R, Kruglikova A, Modan B, Gak E, Hirsh-Yechezkel G, Theodor L, Novikov I, et al (1998) The 185delAG BRCA1 mutation originated before the dispersion of Jews in the Diaspora and is not limited to Ashkenazim. Hum Mol Genet 7:801–805

Gershoni-Baruch R, Dagan E, Kepten I, Fried G (1997) Co-segregation of BRCA1 185delAG mutation and BRCA2 6174delT in one single family. Eur J Cancer 33:2283–2284

Levy-Lahad E, Catane R, Eisenberg S, Kaufman B, Hornreich G, Lishinsky E, Shohat M, et al (1997) Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and breast-ovarian cancer families. Am J Hum Genet 60: 1059–1067

Ludwig T, Chapman D, Papaioannou, VE, Efstratiadis A (1997) Targeted mutations of breast cancer susceptibility gene homologes in mice: lethal phenotypes of BRCA1, BRCA2, BRCA1/BRCA2, BRCA1/p53, and BRCA2/p53 nullizygous embryos. Genes Dev 11:1226–1241

Oddux C, Strewing JP, Clayton MC, Neuhausen S, Brody LC, Kaback M, Haas B, et al (1996) The carrier frequency of the BRCA2 6174delT mutation among Ashkenazi Jewish individuals is approximately 1%. Nat Genet 14:188–190

Ramus SJ, Friedman LS, Gayther SA, Ponder AJ, Bobrow LG, van der Looji, Papp J, et al (1997) A breast/ovarian cancer patient with germline mutations in both BRCA1 and BRCA2. Nat Genet 15:14–15

Roa BB, Boyd AA, Volcik K, Richards CS (1996) Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. Nat Genet 14:185–187

Struewing JP, Abeliovich D, Peretz T, Avishai N, Kaback MK, Collins FS, Brody LC (1995) The carrier frequency of the BRCA1 185delAG is approximately 1 percent in Ashkenazi Jewish individuals. Nat Genet 11:198–200

Address for correspondence and reprints: Dr. Dvorah Abeliovich, Department Human Genetics, Hadassah University Hospital, P.O. Box 12000, Jerusalem, Israel 91120. E-mail: dvoraha@cc.huji.ac.il

All authors are members of the Israeli Consortium of Breast Cancer Genetics.

0002-9297/98/6304-0039$02.00

Am. J. Hum. Genet. 63:1227–1232, 1998

Partial Triplication of mtDNA in Maternally Transmitted Diabetes Mellitus and Deafness

To the Editor:

Maternally inherited diabetes and deafness (MIDD) is a recently recognized subtype of diabetes mellitus (DM) that is associated with mtDNA mutations (Maassen et al. 1997). The first mtDNA defect described for MIDD was a deletion associated with a duplication of the mtDNA in a family presenting DM and deafness over three generations (Ballinger et al. 1992, 1994). Subsequent to this observation, a mutation in nucleotide (nt) 3243 was reported in several pedigrees presenting DM and deafness (Reardon et al. 1992; van den Ouweland et al. 1992; Kadowaki et al. 1993). We report a partial tandem triplication of 9.2 kb in one member of a family presenting MIDD associated with a tandem duplication of 4.6 kb.

In 1966, a 44-year-old man (II-7) of Italian origin was hospitalized for insulin-dependent DM and hearing loss. In 1973, his nephew (III-2), who was born in 1932, was hospitalized for non–insulin-dependent DM and deafness. At that time, the morbid association led to a study of the pedigree (fig. 1), which showed transmission of DM and deafness over four generations, with a total of 13 affected individuals (Kressmann 1976). Seven individuals from the pedigree (III-3, III-4, IV-1, IV-2, IV-3, IV-4, and IV-5) were examined by clinicians. The clinical history was the same for all affected patients: the first manifestation was deafness, beginning at 20–30 years of age, with a rapid and severe increase in bilateral sensory hearing loss. DM developed later in the 3d decade, and insulin was required either immediately or at a later date. At that time, the individuals from the fourth generation, who were <20 years of age, presented no deafness or DM. No pedigree member had ptosis, ophthalmoplegia, or muscle weakness. Recently, the maternal inheritance pattern of DM and deafness in this family

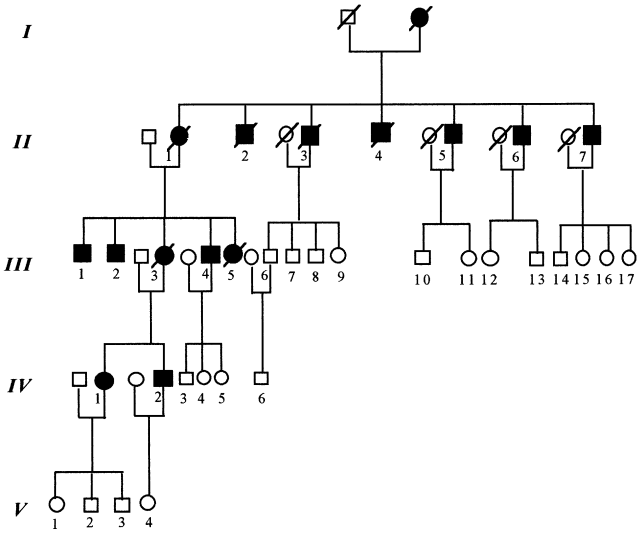

Figure 1 Pedigree of family analyzed in this study. Unblackened symbols indicate unaffected individuals, and blackened symbols indicate affected individuals. Nine family members (II-7, III-2, III-3, III-4, IV-1, IV-2, IV-3, IV-4, and IV-5) were examined in 1973.

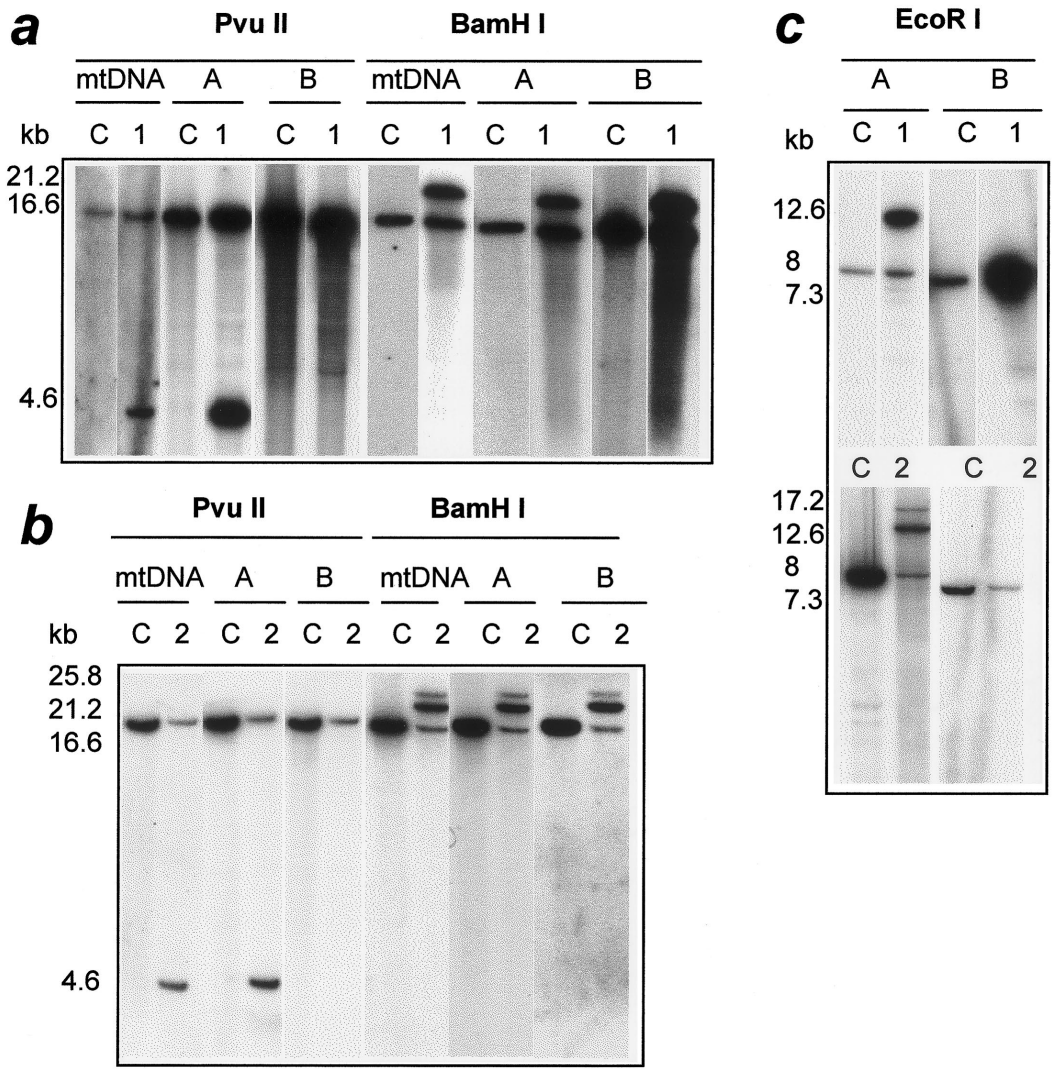

Figure 2 *a,* mtDNA of patient 1, digested with *Pvu*II or *Bam*HI and probed with an mtDNA probe, probe A and probe B. The mtDNA showed additional fragments of 4.6 kb (*Pvu*II digest) and 21.2 kb (*Bam*HI digest), respectively, that are consistent with a partial duplication of 4.6 kb. *b,* mtDNA of patient 2. A supplementary band of 25.8 kb was visualized with *Bam*HI digestion. This band was detected after hybridization with probe A included in the duplicated region (nts 2630–3353) and with probe B not retained in the duplicated segment (nts 7392–8351), thus ruling out circular deleted monomers or dimers. *c,* Digestion with *Eco*RI. For the control DNA, the expected fragments of 8, 7.3, and 1 kb are shown (the 1-kb band is not visualized). The mtDNA of patient 1 shows a supplementary band of 12.6 kb labeled with probe A but not with probe B, corresponding to the partially duplicated molecule. For patient 2, one additional band of 17.2 kb was evidenced with probe A but not with probe B. This is interpreted as a new mtDNA species harboring a tandem repetition of the 4.6-kb duplicated sequence of patient 1.

was noticed, and three patients were examined again by clinicians. Patient 1 (IV-1), 40 years old, and patient 2 (III-1), 65 years old, presented severe deafness and DM that, with time, required insulin. Patient 3 (IV-2), 36 years old, had moderate bilateral sensory hearing loss and subnormal glucose tolerance.

Histopathological studies of the skeletal muscle biopsy specimens from patients 1 and 2 showed no ragged red fibers, a complex IV enzymatic deficiency in a few fibers, and very limited lipid storage on electron microscopy. Neither mitochondrial hyperplasia nor inclusions were observed. No abnormalities were observed for patient 3.

Total DNA was extracted from the muscle biopsy specimens and blood of the three patients. The search for the mtDNA mutation in the $tRNA^{leu(UUR)}$ gene at nt 3243 was performed in accordance with a protocol described elsewhere (Ciafaloni et al. 1991). For Southern blotting, 5 μg of total muscle DNA and 10 μg of blood DNA were digested with restriction enzyme *Pvu*II (nt

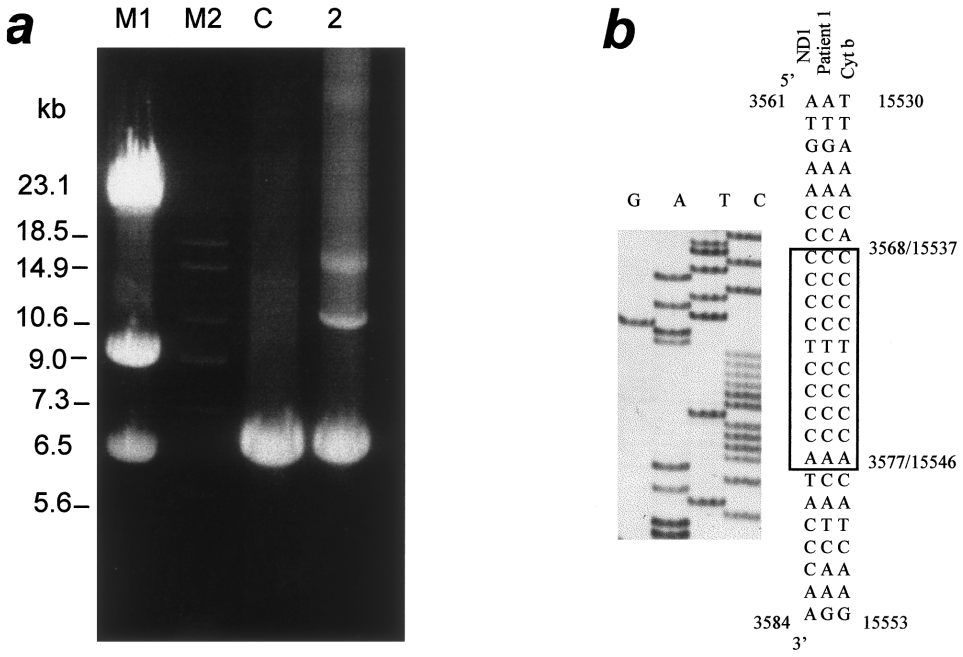

Figure 3 *a,* PCR products obtained after amplification with primers 5 and 6. A 6.8-kb band was detected in control DNA. For patient 2, two supplementary bands of 11.4 kb and 16 kb were present, thus confirming the results of the Southern blot analysis, with regard to the existence of mtDNA molecules linked to one (duplicated species) or two (triplicated species) rearranged molecules of 4.6 kb. "M1" and "M2" indicate the molecular-weight markers. Lane M1, Phage λ, digested with *Hind*III. Lane M2, Raoul. Lane C, Control DNA. Lane 2, Patient 2. *b,* Sequence across the duplication junction of patient 1. The sequencing of the cloned 369-bp PCR products obtained after amplification with primers 3 and 4 showed a normal sequence for region 3274–3577. Subsequent nts corresponded exactly to region 15547–15600. The duplication junction is a perfect direct repeat of 10 nts, located in regions 3568–3577 (ND1 gene) and 15537–15546 (Cyt b gene). The boxed region indicates the 10-bp perfect direct repeat. The normal sequences of ND1 and Cyt b correspond to the left and right sequences, respectively. The same 369-bp PCR products were sequenced in patients 2 and 3, and identical results were obtained.

2650), *Bam*HI (nt 14258), or *Eco*RI (nts 4121, 5274, and 12640), in accordance with the manufacturer's recommendations; were separated by gel electrophoresis; and were blotted onto nylon membrane (Hybond N^+, Amersham). Hybridization was performed with a random-primed ^{32}P-labeled mtDNA probe (Lutfalla et al. 1985) and with two random-primed ^{32}P-labeled mtDNA probes derived from PCR products spanning nts 2630–3353 (probe A) and nts 7392–8351 (probe B). Quantification was performed with a Phosphor Imager (Molecular Dynamics) by scanning of the nylon filters of the *Bam*HI digests hybridized with probe B. PCR analyses of the duplicated region were performed on muscle and blood samples by use of two different couples of primers (primer 1, nts 2630–2650, 5′-GAA TGG CTC CAC GAG GGT TC-3′, and primer 2, nts 16255–16274, 5′-CCT AGT GGG TGA GGG GTG GC-3′; primer 3, nts 3274–3293, 5′-ACA GTC AGA GGT TCA ATT CC-3′, and primer 4, nts 15581–15600, 5′-GGG ACG GAT CGG AGA ATT GT-3′). Amplification conditions were 30 cycles of 1 min at 93°C, 1 min at 62°C (primers 1 and 2) or at 55°C (primers 3 and 4), and 2 min at 72°C, with 2.5 U of *Taq* polymerase (Promega). The PCR products obtained with primers 1 and 2 were analyzed with restriction enzymes *Bcl*I (nts 3658, 7657, 8591, and 11921), *Eco*RI (nts 4121, 5274, and 12640), *Eco*RV (nts 3179, 6734, and 12871), *Kpn*I (nts 2573, 16048, and 16121), and *Xho*I (nt 14955). The 369-bp PCR fragment obtained after amplification with primers 3 and 4 was cloned into the pGEM-T vector (Promega) and was used as a template for dideoxy sequencing using the T7 sequencing kit (Pharmacia), in accordance with the manufacturer's specifications, to reveal the duplication junction. To amplify all length variants of the mtDNA molecules (normal, duplicated, and triplicated) in patient 2, a long PCR was performed with a DNA thermal cycler (Robocycler, Stratagene) and the Expand Long PCR Template PCR system (Boehringer Mannheim), by use of the manufacturer's recommendations modified as described elsewhere (Fromenty et al. 1996). The amplification conditions were 35 cycles for 30 s at 93°C, 30 s at 66°C, and 17 min at 68°C. The primer pair comprised primer 5 (forward primer), nts 13949–13972, 5′-CCT ATC TAG GCC TTC TTA CGA

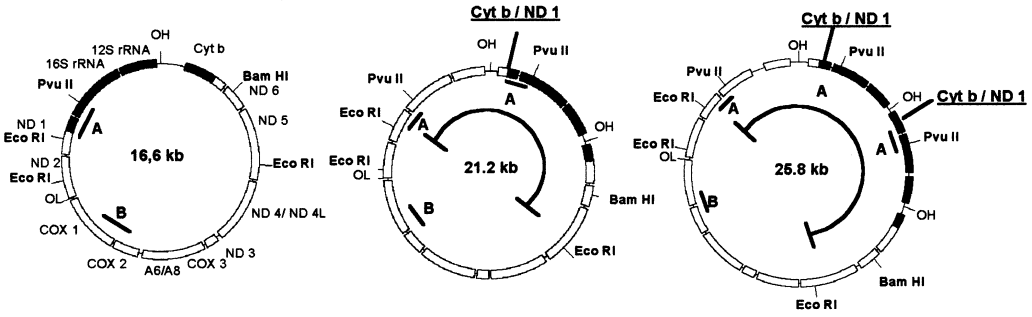

Figure 4 Schematic representation showing the normal mitochondrial genome (16.6 kb), the partially duplicated mtDNA molecule (21.2 kb) found in the three patients, and the abnormal molecule harboring the triplication (25.8 kb) in patient 2. The *Pvu*II, *Bam*HI, and *Eco*RI sites and the locations of probe A (nts 2630–3353) and probe B (nts 7392–8351) are indicated. The curved lines indicate the regions corresponding to the *Eco*RI digests of 12.6 kb and 17.2 kb, for the Southern blot analysis of patient 2. Black boxes indicate the genes involved in the rearrangement in the normal molecule and the duplicated or triplicated fragments in the rearranged genomes.

GCC-3′, and primer 6 (reverse primer), nts 4207–4186, 5′-GTA ATG CTA GGG TGA GTG GTA G-3′.

None of the patients carried the pathogenic point mutation at nt 3243. On the other hand, the results of Southern blot analysis of muscle DNA from patients 1 and 3, digested with restriction enzymes *Pvu*II and *Bam*HI, were consistent with a partial duplication of a 4.6-kb region of mtDNA that included the *Pvu*II restriction site (nt 2650) but not the *Bam*HI site (nt 14258) (fig. 2*a*). Southern blot analysis of skeletal muscle DNA from patient 2 unexpectedly revealed an additional 25.8-kb band on *Bam*HI digestion (fig. 2*b*), which could correspond to either (1) an undigested circular deletion monomer or dimer, (2) a second, larger duplicated molecule, or (3) an additional insert of 4.6 kb corresponding to a partially triplicated molecule. Hybridization of the 25.8-kb band with a probe not included in the duplication (probe B) ruled out a circular deletion monomer or dimer. The possibility of a second species duplication also was ruled out, because an abnormal band >4.6 kb was not detected with the *Pvu*II digest, and only one band was obtained by PCR when primers 3 and 4 were used. The possibility of an mtDNA triplication in patient 2 was confirmed by digestion of the DNA, with *Eco*RI, which gave two additional fragments, compared with that of the control (fig. 2*c*): one fragment, of 12.6 kb, corresponded to the partial duplication also found in patient 1, and the other, of 17.2 kb, was consistent with an mtDNA molecule linked to two partially duplicated molecules. The triplication was confirmed further by means of long PCR using primers 5 and 6 (fig. 3*a*). PCR analysis and sequencing showed that the breakpoint junction was located between the ND1 gene and the cytochrome (Cyt) b gene at a 10-bp perfect direct repeat (fig. 3*b*). These results indicate the presence of three species of mtDNA molecules in patient 2: a normal molecule (16.6 kb), a rearranged molecule (21.2 kb) that contains an additional 4.6-kb fragment corresponding to a partial tandem duplication, and a rearranged molecule (25.8 kb) that contains two copies of the 4.6-kb fragment corresponding to a partial triplication (fig. 4). The proportion of duplicated mtDNA in muscle was 42% for patient 1 and 61% for patient 2. The proportion of triplicated molecules was only 6% for patient 2. In blood, the proportion of duplicated molecules was 52% for patient 1 and 67% for patient 2. No triplicated molecules were detected in the blood.

Partial triplication of human mtDNA is an extremely

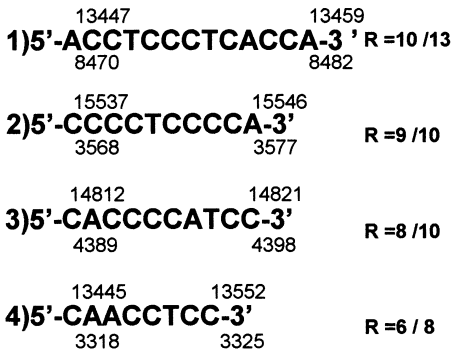

Figure 5 DNA sequences of perfect direct repeats located across breakpoint junctions of the mtDNA reported in the literature and showing a polypyrimidine tract (*1*) in the common deletion (Schon et al. 1989); (*2*) in the family described here and in two other cases of duplication/deletion associated with myopathy (Fromenty et al. 1996; Manfredi et al. 1997); (*3*) in a case of duplication/deletion associated with DM (Ballinger et al. 1992, 1994); and (*4*) in a duplication associated with DM and myopathy. "R," reported as 6/8 in (*4*), corresponds to a ratio of 13/18 if the entire imperfect direct repeat of 18 nts is considered (Dunbar et al. 1993). "R" indicates the number of pyrimidines in the direct repeat of the light-strand DNA template.

rare event. Only two cases have been reported previously: one in cell culture (Holt et al. 1997) and a second identified from autopsy material from a clinically asymptomatic individual (Tengan and Moraes 1998). The molecular mechanisms leading to large-scale rearrangements have not been well characterized yet, and models of slippage mispairing or illegitimate recombination events have been proposed (Shoffner et al. 1989; Poulton et al. 1993). Nevertheless, the origin of slippage mispairing still remains elusive. Like other reported examples of large-scale rearrangements (fig. 5), our direct repeat harbors a long polypyrimidine (L strand)/polypurine (H strand) sequence. We suggest that the second direct repeat (polypyrimidine/polypurine tract) could interact with the first direct repeat to form a triple helix (H DNA) and leads thereafter to the first tandem duplication. The repetition of this mechanism then could lead to the triplication.

A major cause of diabetes in DM and deafness seems to be a decrease in ATP production in pancreatic β cells that could be responsible for the decrease in insulin secretion (Dukes et al. 1994; Gerbitz et al. 1996). Under normal physiological conditions, the increase in blood glucose concentration results in an increase in ATP production in pancreatic β cells, which in turn leads to the closure of K^+ channels located in the cell membrane. This closure induces a membrane depolarization and the opening of voltage-dependent Ca^{2+} channels. The influx of Ca^{2+} into β cells then stimulates insulin exocytosis. In DM and deafness, gene defects lead to an oxydative phosphorylation disturbance and eventually to decreased ATP production. The pathogenic role of duplicated or triplicated mtDNA molecules in this context is difficult to assess, because all of the mtDNA information content is present in these rearranged molecules. In addition, some experiments have indicated a pathogenic role only for mtDNA deletions (Manfredi et al. 1997). Nevertheless, like others (Dunbar et al. 1993), we have not detected any deleted molecules, and we cannot exclude a possible respiratory-chain impairment secondary to duplicated or triplicated molecules. Indeed, an increase in lactate production in cell cultures that harbor duplicated and triplicated mtDNA has been demonstrated (Holt et al. 1997).

Acknowledgments

We thank J. P. Mazat for his helpful discussion, M. Perrot for the DNA quantification, and C. Mehaye for technical assistance. This work was supported by a grant from the Ministère des Affaires Sociales de la Santé et de la Ville, Projet Hospitalier de Recherche Clinique, in 1994.

Marie-Laure Martin Negrier,[1] Michelle Coquet,[1] Brigitte Teissier Moretto,[1] Jean-Yves Lacut,[2] Michel Dupon,[2] Bertrand Bloch,[1] Patrick Lestienne,[3] and Claude Vital[1]

[1]*Laboratoire d'Anatomie Pathologique and* [2]*Service des Maladies Infectieuses, Centre Hospitalier Régional Pellegrin, and* [3]*Contrat Jeune Formation 97-05, Institut National de la Santé et de la Recherche Médicale, Université de Bordeaux II, Bordeaux*

References

Ballinger SW, Shoffner JM, Gebhart S, Koontz DA, Wallace DC (1994) Mitochondrial diabetes revisited. Nat Genet 7: 458–459

Ballinger SW, Shoffner JM, Hedaya EV, Trounce I, Polak MA, Koontz DA, Wallace DC (1992) Maternally transmitted diabetes and deafness associated with a 10.4 kb mitochondrial DNA deletion. Nat Genet 1:11–15

Ciafaloni E, Ricci E, Servidei S, Shanske S, Silvestri G, Manfredi G, Schon EA, et al (1991) Widespread tissue distribution of a tRNALeu(UUR) mutation in the mitochondrial DNA of a patient with MELAS syndrome. Neurology 41: 1663–1664

Dukes ID, McIntyre MS, Mertz RJ, Philipson LH, Roe MW, Spencer B, Worley JF III (1994) Dependence on NADH produced during glycolysis for beta-cell glucose signaling. J Biol Chem 269:10979–10982

Dunbar DR, Moonie PA, Swingler RJ, Davidson D, Roberts R, Holt IJ (1993) Maternally transmitted partial direct tandem duplication of mitochondrial DNA associated with diabetes mellitus. Hum Mol Genet 2:1619–1624

Fromenty B, Manfredi G, Sadlock J, Zhang L, King MP, Schon EA (1996) Efficient and specific amplification of identified partial duplications of human mitochondrial DNA by long PCR. Biochim Biophys Acta 1308:222–230

Gerbitz KD, Gempel K, Brdiczka D (1996) Mitochondria and diabetes: genetic, biochemical, and clinical implications of the cellular energy circuit. Diabetes 45:113–126

Holt IJ, Dunbar DR, Jacobs HT (1997) Behaviour of a population of partially duplicated mitochondrial DNA molecules in cell culture: segregation, maintenance and recombination dependent upon nuclear background. Hum Mol Genet 6:1251–1260

Kadowaki H, Tobe K, Mori Y, Sakura H, Sakuta R, Nonaka I, Hagura R, et al (1993) Mitochondrial gene mutation and insulin-deficient type of diabetes mellitus. Lancet 341: 893–894

Kressmann F (1976) Association diabète et surdité: a propos d'une famille atteinte de cette double tare. MD thesis, University of Bordeaux II, Bordeaux

Lutfalla G, Blanc H, Bertolotti R (1985) Shuttling of integrated vectors from mammalian cells to *E. coli* is mediated by head-to-tail multimeric inserts. Somat Cell Mol Genet 11: 223–238

Maassen JA, van den Ouweland JM, 't Hart LM, Lemkes HH (1997) Maternally inherited diabetes and deafness: a dia-

betic subtype associated with a mutation in mitochondrial DNA. Horm Metab Res 29:50–55
Manfredi G, Vu T, Bonilla E, Schon EA, DiMauro S, Arnaudo E, Zhang L, et al (1997) Association of myopathy with large-scale mitochondrial DNA duplications and deletions: which is pathogenic? Ann Neurol 42:180–188
Poulton J, Deadman ME, Bindoff L, Morten K, Land J, Brown G (1993) Families of mtDNA re-arrangements can be detected in patients with mtDNA deletions: duplications may be a transient intermediate form. Hum Mol Genet 2:23–30
Reardon W, Ross RJ, Sweeney MG, Luxon LM, Pembrey ME, Harding AE, Trembath RC (1992) Diabetes mellitus associated with a pathogenic point mutation in mitochondrial DNA. Lancet 340:1376–1379
Schon EA, Rizzuto R, Moraes CT, Nakase H, Zeviani M, DiMauro S (1989) A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. Science 244: 346–349
Shoffner JM, Lott MT, Voljavec AS, Soueidan SA, Costigan DA, Wallace DC (1989) Spontaneous Kearns-Sayre/chronic external ophthalmoplegia plus syndrome associated with a mitochondrial DNA deletion: a slip-replication model and metabolic therapy. Proc Natl Acad Sci USA 86:7952–7956
Tengan CH, Moraes CT (1998) Duplication and triplication with staggered breakpoints in human mitochondrial DNA. Biochim Biophys Acta 1406:73–80
van den Ouweland JM, Lemkes HH, Ruitenbeek W, Sandkuijl LA, de Vijlder MF, Struyvenberg PA, van de Kamp JJ, et al (1992) Mutation in mitochondrial tRNA(Leu)(UUR) gene in a large pedigree with maternally transmitted type II diabetes mellitus and deafness. Nat Genet 1:368–371

Address for correspondence and reprints: Marie-Laure Martin Negrier, Laboratoire d'Anatomie Pathologique, CHR Pellegrin, 33076 Bordeaux, Cedex, France. E-mail: mlmartin@hippocrate.u-bordeaux2.fr

0002-9297/98/6304-0041$02.00

Am. J. Hum. Genet. 63:1232–1234, 1998

Reply to Inglehearn

To the Editor:

In our article "Localization of a Novel X-Linked Progressive Cone Dystrophy Gene to Xq27: Evidence for Genetic Heterogeneity" (Bergen and Pinckers 1997), we presented evidence favoring a location, on Xq27, for a cone dystrophy gene. This localization is questioned by Dr. Inglehearn (1998) in his letter "LOD Scores, Location Scores, and X-Linked Cone Dystrophy." Although Dr. Inglehearn makes a good (methodological) point, we feel that the majority of his criticism is not justified.

Clearly, as Dr. Inglehearn states correctly, figure 2 in our previous article (Bergen and Pinckers 1997) shows a picture of the multipoint location scores rather than of the multipoint LOD scores. Although the presentation of location scores instead of multipoint LOD scores is not wrong in itself, it is rather unconventional and therefore confusing. Thus, we agree that, with regard to a multipoint location score of 10.8, the calculated multipoint LOD score is indeed 2.35. Obviously, for X-chromosomal disorders, the latter score is still considered to be significant.

Subsequently, Dr. Inglehearn calculates, on the basis of the data presented, multipoint (maximum?) LOD scores of 3.38 and 2.46 at DXS998, using different LOD-score strategies. Unfortunately, additional calculations for other markers are not given. Both these LOD scores for DXS998 are higher than the true multipoint LOD scores calculated by us (maximum LOD score [Z_{max}] of 2.35). Thus, in our article (Bergen and Pinckers 1997), our calculation of LOD scores and our choice of parameters were in fact very conservative. Therefore, the assertion by Dr. Inglehearn (1998) that "these data do indeed suggest a locus for X-linked cone dystrophy in this region but with rather less significance than Bergen and Pinckers have stated" (p. 900) is not justified. Most likely, the true findings for the Z_{max} score at DXS998 are somewhere within the range 2.35–3.38.

Dr. Inglehearn states that a second weakness of the article is the order and placement of markers used in the multipoint linkage analysis. However, this assertion is based on out-of-date and incomplete genetic maps of the region, as indicated by the references to literature published in 1992 and 1994 (NIH/CEPH Collaborative Mapping Group 1992; Gyapay et al. 1994), and therefore is not justified. Much more recent and up-to-date consensus maps (Dib et al. 1996) place DXS998 ~15 cM from the distal tip of the X chromosome and at least 7 cM proximal to the red cone pigment (RCP)/green cone pigment (GCP) gene cluster.

In addition, in our article (Bergen and Pinckers 1997), data on two additional markers, DXS297 and DXS1123, are presented. Both DXS297 and DXS1123 reveal higher (maximum two-point) LOD scores of 2.54 and 2.60, respectively, without recombination with COD2, but these markers are ignored in the comments by Dr. Inglehearn. Most likely, on the basis of recombination counting, haplotype analysis, and marker-to-marker analysis, both DXS297 and DXS1123 are part of a cosegregating haplotype, together with DXS998 and COD2. Although DXS297 and DXS1123 are not present on the CEPH/Généthon consensus maps, at least two independent reports in the literature (Richards et al. 1991; Donnelly et al. 1994) place DXS297 proximal to the fragile X site, which is located on Xq27.3 (Dib et al. 1996). Similar, although somewhat weaker, evidence can be found for DXS1123. In contrast, the RCP/GCP gene cluster is located on Xq28. In conclusion, there is

convincing evidence that the marker order used by us in our previous study is the (most) correct one.

On the basis of the likelihood data only (also see above), sufficient evidence for the "most likely" presence of a COD2 locus on Xq27 already existed; however, special *additional* attention was given to markers surrounding the RCP/GCP locus on Xq28, in view of the relatively close genetic distance between COD2 and the RCP/GCP cluster. Thus, the markers that very closely flank (0.5 cM each) the RCP/GCP gene cluster—namely, DXS8103 and DXS8069—were used. Again, this information can be obtained easily by detailed study of recent genetic databases.

On the basis of haplotype analysis only, the involvement of RCP/GCP in this pedigree is very unlikely. Markers DXS8103 and DXS8069 are only 1 cM apart and cosegregate with markers DXS52 and DXS1113, without recombination in the pedigree, when the fewest number of recombination events are assumed (see fig. 1 in Bergen and Pinckers 1997). If the RCP/GCP gene cluster is involved in the X-linked progressive cone dystrophy (XLPCD) in this pedigree, a double recombination event would be assumed to have occurred between DXS8103 and DXS8069 (potentially revealed by the haplotype of individual III-13/16). From multiple studies reported in the literature and from our own segregation data of hundreds of families, we know that such double recombination events on such a short stretch of DNA are extremely rare and occur in <0.1% of cases. Theoretically, without consideration of genetic interference down-regulating recombination of closely linked loci, the "risk for a double recombination" could be calculated as follows: (the chance of the first recombination occurring in 1 cM) × (the chance of the second recombination occurring in 1 cM) × (the number of meiosis in which these recombinations potentially could occur). If we assume that, in our pedigree, these recombinations could have taken place in ~5 meioses, which is the number of female meiosis between the two larger branches of the pedigree, then the overall risk for a double recombination not detected by our DNA analysis would be $.01 \times 5 \times .01 \times 5 = .0025$, or 0.25%. If we consider "genetic interference," this figure most likely drops to $\leqslant 0.1\%$, which is the figure given above. In conclusion, on the basis of haplotype data and risk calculations, the chance that RCP/RCG is involved in XLPCD in our pedigree is $\leqslant$1:1,000.

Given the complete cosegregation of both DXS8103 and DXS8069 with both DXS1113 and DXS52, two-point LOD scores for the first two markers and XLPCD are similar to those for the last two markers, which are given in our previous article (Bergen and Pinckers 1997). At a recombination fraction (θ) of .00, the LOD scores were −3.40 (DXS8103) or −4.26 (DXS1113). For the same markers, Z_{max} is reached at $\theta = .05$ (LOD score 1.22) and at $\theta = .10$ (LOD score 0.57), respectively. If equal distances between RCP/GCP and both DXS8103 and DXS8069 (0.005 cM each) are assumed, the LOD scores would be 0.46 (DXS8103) and −0.377 (DXS8069) more than those for RCP/GCP.

Similar low(er) or negative LOD values are obtained with multipoint linkage analysis, with different combinations of various markers and RCP/GCP, different parameters, and different LOD-score strategies. Given the fact that markers at the Xq27 cluster (DXS998, DXS1123, and DXS297) reach a multipoint Z_{max} of ~2.5, the markers at RCP/GCP should reach multipoint LOD scores of >1.5 in order to be significant, according to the so-called $Z_{max} - 1$ LOD unit rule. Instead, Z_{max} scores at the RCP/GCP cluster remained <0.5. Thus, by statistical means, the involvement of the RCP/GCP cluster was excluded in this pedigree.

On the other hand, the involvement of a rare and spontaneous Xq27/Xq28 dislocation or abnormal duplication(s) of the RCP/GCP gene cluster (as have been described elsewhere) or of other rearrangements further away from the RCP/GCP cluster or even other genetic mechanisms involved in the XLPCD in this pedigree could not and cannot yet be excluded. To obtain initial evidence for the exclusion of these hypotheses, however, Southern analysis with RCP/GCP cDNA was performed, and no structural abnormalities were found. Although this data alone does not *exclude* the involvement of RCP/GCP, they *do* suggest that involvement of RCP/GCP is even less likely, when considered in the context of the evidence that we obtained earlier.

The authors welcome the suggestion by Dr. Inglehearn that mutations or rearrangements upstream of the RCP/GCP locus possibly could be implicated in this XLPCD family, although our data suggest that such a genomic abnormality must be very much further away than the 43 kb mentioned. In conclusion, although the possible involvement of (regulatory elements of) the RCP/GCP gene cluster in the described XLPCD pedigree certainly is worth further investigation, the evidence accumulated thus far suggests the presence of a separate and distinct XLPCD locus, on Xq27.

A. A. B. Bergen[1] and A. J. L. G. Pinckers[2]
[1]*The Netherlands Ophthalmic Research Institute, Amsterdam; and* [2]*Department of Ophthalmology, University of Nijmegen, Nijmegen, The Netherlands*

References

Bergen AAB, Pinckers AJLG (1997) Localization of a novel X-linked progressive cone dystrophy gene to Xq27: evidence for genetic heterogeneity. Am J Hum Genet 60:1468–1473

Dib C, Fauré 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

Donnelly A, Kozman H, Gedeon AK, Webb S, Lynch M, Sutherland GR, Richards RI, et al (1994) A linkage map of microsatellite markers on the human X chromosome. Genomics 20:363–370

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, et al (1994) The 1993–94 Généthon human genetic linkage map. Nat Genet 7:246–339

Inglehearn CF (1998) LOD scores, location scores, and X-linked cone dystrophy. Am J Hum Genet 63:900–901

NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. Science 258:67–86

Richards RI, Shen Y, Holman K, Kozman H, Hyland VJ, Mulley JC, Sutherland GR (1991) Fragile X syndrome: diagnosis using highly polymorphic microsatellite markers. Am J Hum Genet 48:1051–1057

Address for correspondence and reprints: Dr. A. A. B. Bergen, P. O. Box 12141, 1100 AC Amsterdam, The Netherlands. E-mail: A.Bergen@ioi.knaw.nl

0002-9297/98/6304-0042$02.00

Am. J. Hum. Genet. 63:1234–1236, 1998

mtDNA Suggests Polynesian Origins in Eastern Indonesia

To the Editor:

mtDNA evidence has previously been interpreted as providing strong support for a model of rapid expansion of the Polynesian peoples from a homeland in Taiwan or southern China ~6,000 years ago into the remote Pacific. Here, we argue that the evidence is consistent with an alternative view, namely, that the Polynesian expansion originated within the Indonesian archipelago.

Several studies have been published concerning the settlement of the remote Pacific that use the phylogeographic analysis of mtDNA, either large-scale sampling and control-region sequence analysis (Lum et al. 1994; Redd et al. 1995; Sykes et al. 1995) or sequence-specific oligonucleotide analysis (Melton et al. 1995). These have distinguished two main hypotheses concerning Polynesian origins. The first hypothesis, often referred to somewhat incongruously as the "express train to Polynesia" (Diamond 1988), was proposed by Bellwood (1991, 1997). This suggests that the Polynesians originated in a demic expansion of Austronesian-speaking agriculturalists from the southern China mainland, ~6,000 years ago, and spread successively to Taiwan, the Philippines, eastern Indonesia, and then Melanesia, reaching Fiji by ~3,500 years ago and radiating across the Pacific to fill the Polynesian triangle by ~1,000 years ago. They would have absorbed and replaced the local hunter-gatherer populations in Southeast Asia, who would have been of Australo-Melanesian ancestry. The principal alternative view, argued by Terrell (1986), is that the Polynesians evolved locally in Melanesia or, at least, within the voyaging corridor between the mainland and the Solomon Islands, defined by Irwin (1992).

Melton et al. (1995) and Redd et al. (1995) analyzed the history of a COII/tRNA^{Lys} intergenic 9-bp deletion by means of a suite of characteristic control-region transitions at positions 16189, 16217, 16247, and 16261 of the first hypervariable segment (according to the Cambridge Reference Sequence; Anderson et al. 1981). They referred to this as the "Polynesian motif," because of its high frequencies in Polynesia, despite its occurrence farther west (Hagelberg and Clegg 1993; Redd et al. 1995). They traced the origin of this motif to Taiwan and proposed that this represented the Polynesian homeland, in line with the Bellwood (1997) hypothesis, while acknowledging that the motif itself probably arose in eastern Indonesia. Sykes et al. (1995) agreed in tracing the origin of the motif to Taiwan but also pointed out that the lack of the motif in Taiwan, Borneo, and the Philippines might complicate the issue. In addition, they pointed out, along with Lum et al. (1994), that somewhat <5% of Polynesians had control-region sequences derived from Melanesia. Furthermore, Sykes et al. (1995) distinguished a third hypothesis, proposed by Heyerdahl (1950), suggesting that Polynesian ancestry may have been from South America, a view that received little or no support from the mitochondrial evidence (Sykes et al. 1995; Bonatto et al. 1996).

Although the evidence is therefore strong that Polynesians derive most of their maternal lineages from Southeast Asia, a fourth hypothesis has received little attention. This view, in contrast to the "express train" model of an agricultural expansion from Taiwan, suggests that the Austronesian speakers originated neither in southern China nor in Taiwan but toward the center of island Southeast Asia, in the vicinity of the Sulawesi-Mindanao region of the Philippines and Indonesia (Solheim 1994) or perhaps over the entire region of island Southeast Asia in which Austronesian languages are now spoken (Meacham 1984–85). This would suggest that the extant inhabitants of island Southeast Asia were the descendants of earlier Pleistocene settlers rather than of Neolithic people from the mainland. Meacham (1984–85) cites the paucity of extant Austronesian speakers on the southern Chinese mainland—or, indeed, any historical evidence for their existence there—in support of this view. There is also anthropometric evidence that Polynesians closely resemble island Southeast Asian populations but not aboriginal Taiwanese or southern Chinese populations (Pietrusewsky 1997).

Combining the published mitochondrial evidence al-

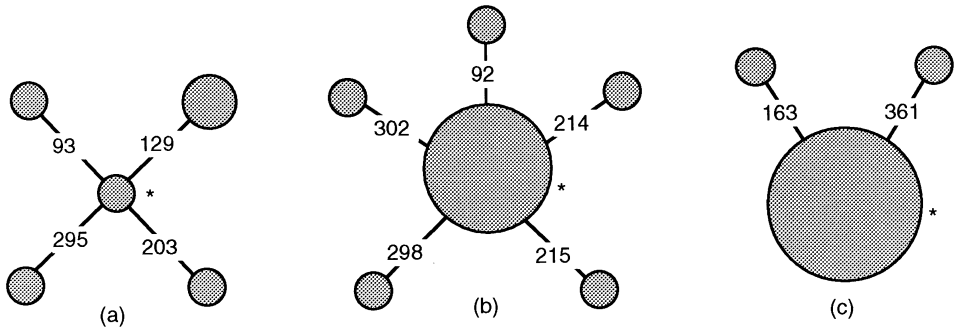

Figure 1 Phylogenetic tree of mitochondrial sequence haplotypes containing the "Polynesian motif" in (*a*) eastern Indonesia, (*b*) Papua New Guinea, and (*c*) American Samoa (data of Redd et al. 1995), in the part of the first hypervariable segment of the control region encompassing bp 16090–16365. The circles represent sequence haplotypes, with area proportional to frequency. The links represent transitional mutations (less 16,000) from the central motif sequence, which deviates from the Cambridge Reference Sequence by transitions at 16189, 16217, 16247, and 16261 (labeled with an asterisk [*]).

lows us to assess this model and to refine our model of predominantly Southeast Asian origins of the Polynesians. Although elevated to very high frequencies throughout Polynesia, probably as a result of severe population bottlenecks and expansions, the Polynesian motif is not exclusively Polynesian but also occurs at moderate frequencies in island Melanesia, coastal New Guinea, eastern Indonesia, and even Madagascar (Melton et al. 1995; Redd et al. 1995; Soodyall et al. 1995; Sykes et al. 1995). The motif evolved, via a transition at position 16247, from a sequence haplotype characterized by transitions at positions 16189, 16217, and 16261. Whereas the full motif itself is rather restricted geographically, the ancestral haplotype and others derived from it are found throughout island Southeast Asia, China, and even, at low frequencies, as far afield as Mongolia and India (Melton et al. 1995; Kolman et al. 1996). Its diversity in Taiwan, calculated by use of the statistic ρ (Forster et al. 1996), suggests a divergence time of ~30,000 years, although with a wide 95% credible region.

On the other hand, the Polynesian motif itself is much more restricted geographically, with the highest diversity in eastern Indonesia, a considerable decrease on the New Guinea coast, and the lowest diversity in Polynesia. This suggests that it arose in eastern Indonesia (Melton et al. 1995; Redd et al. 1995). Phylogenetic trees of the sequences characterized by the motif in the data of Redd et al. (1995), from eastern Indonesia, Papua New Guinea, and Samoa, are shown in figure 1. With these data and those of Sykes et al. (1995), we can use the

Table 1

Divergence Time Estimates for the "Polynesian Motif" in Eastern Indonesia, Coastal Papua New Guinea, Samoa, and the Cook Islands, and Its Ancestor Haplotype in Taiwan

Ancestral Sequence Haplotype	Sampling Location	N	ρ	Mean Divergence Time t (years)[a]	Central 95% Credible Region (years)[a]
16189–16217–16261	Taiwan[b]	14	1.14	30,500	17,500–47,000
16189–16217–16261–16247	Eastern Indonesia[c]	6	.83	17,000	5,500–34,500
16189–16217–16261–16247	Coastal Papua New Guinea[b,c]	22	.23	5,000	1,500–10,000
16189–16217–16261–16247	Samoa[b,c]	38	.13	3,000	1,000–6,000
16189–16217–16261–16247	Cook Islands[b]	48	.04	1,000	0–3,000

[a] To the nearest 500 years. For divergence times based on samples sequenced over different extents of hypervariable segment I (HVS I), a weighted mutation rate was used: $\mu = (N_1\mu_1 + N_2\mu_2)/(N_1 + N_2)$, where N_1 and N_2 are the numbers of samples sequenced over the two ranges and μ_1 and μ_2 are the rates appropriate to those ranges. The credible regions (Berger 1985) encompass the central 95% of the posterior density of t, under the assumption of a Jeffreys' prior for t and a likelihood appropriate for a perfectly starlike coalescent tree. It should be noted that the credible regions quoted on t do not take into account uncertainties in the mutation rate.

[b] Data are from Sykes et al. (1995), using a transition rate of 1 in 26,600 years for the truncated HVS I sequences from positions 16189–16375.

[c] Data are from Redd et al. (1995), using a transition rate of 1 in 20,180 years (Forster et al. 1996) for HVS I sequences from positions 16090–16365.

statistic ρ to calculate divergence times for the motif in various regions (table 1). Whereas the ages estimated for the populations of New Guinea, Samoa, and central Polynesia are ~5,000, ~3,000, and ~1,000 years, respectively, indicating successive recent bottlenecks predicted by the hypothesis of expansion from the west, the age for the population of eastern Indonesia (the Moluccas and Nusa Tenggara) is much greater, ~17,000 years.

Given the wide 95% credible regions associated with these age estimates, one cannot, on the basis of these data, confidently rule out either a Taiwanese or even a Melanesian origin for the Polynesians, especially given that much of island Melanesia has yet to be sampled. Nevertheless, they lend little support to the "express train" model. The most likely explanation for these data is that, although the ancestry of the motif goes back to the Southeast Asian Pleistocene era, the Polynesian expansion itself did not originate in either Taiwan or southern China but within tropical island Southeast Asia—most probably in eastern Indonesia, somewhere between southeastern Borneo and the Moluccas, given the almost complete absence of the full motif in western Indonesia and the Philippines (Melton et al. 1995; Sykes et al. 1995). This might also explain the appearance of the motif in Madagascar, in a population speaking an Austronesian language more closely related to Indonesian than to Polynesian languages (Soodyall et al. 1995). It is consistent with the hypothesis that the Austronesian languages originated within island Southeast Asia during the Pleistocene era and spread through Melanesia and into the remote Pacific within the past 6,000 years.

Acknowledgments

We are grateful to Peter Bellwood for stimulating discussions and critical advice and to Vincent Macaulay for statistical assistance. This work was supported by the Wellcome Trust.

MARTIN RICHARDS, STEPHEN OPPENHEIMER, AND BRYAN SYKES

Institute of Molecular Medicine
John Radcliffe Hospital
University of Oxford
Oxford

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

Bellwood P (1991) The Austronesian dispersal and the origin of languages. Sci Am 265:70–75

——— (1997) Prehistory of the Indo-Malaysian archipelago. University of Hawaii Press, Honolulu

Berger JO (1985) Statistical decision theory and Bayesian analysis. Springer-Verlag, New York

Bonatto SL, Redd AJ, Salzano FM, Stoneking M (1996) Lack of ancient Polynesian-Amerindian contact. Am J Hum Genet 59:253–256

Diamond JM (1988) Express train to Polynesia. Nature 336: 307–308

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Hagelberg E, Clegg JB (1993) Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. Proc R Soc Lond B Biol Sci 252:163–170

Heyerdahl T (1950) Kontiki: across the Pacific by raft. Rand McNally, Chicago

Irwin G (1992) The prehistoric exploration and colonisation of the Pacific. Cambridge University Press, Cambridge

Kolman C, Sambuughin N, Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics 142: 1321–1334

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

Meacham W (1984–85) On the improbability of Austronesian origins in South China. Asian Perspect 26:89–106

Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, Stoneking M (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. Am J Hum Genet 57:403–414

Pietrusewsky M (1997) The people of Ban Chiang: an early Bronze Age site in northeast Thailand. In: Bellwood P (ed) Indo-Pacific Prehistory Association Bulletin 16: the Chiang Mai papers. Vol 3. Indo-Pacific Prehistory Association, Canberra, pp 119–147

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/tRNA(Lys) intergenic 9-base-pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615

Solheim WG II (1994) South-east Asia and Korea from the beginnings of food production to the first states. In: De Laet SJ (ed) Prehistory and the beginnings of civilization. Vol 1 in: The history of humanity. Routledge, London, pp 468–481

Soodyall H, Jenkins T, Stoneking M (1995) "Polynesian" mtDNA in the Malagasy. Nat Genet 10:377–378

Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M (1995) The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. Am J Hum Genet 57: 1463–1475

Terrell JE (1986) Prehistory in the Pacific Islands. Cambridge University Press, Cambridge

Address for correspondence and reprints: Martin Richards, Department of Cellular Science, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, United Kingdom. E-mail: mrichard@worf.molbiol.ox.ac.uk

0002-9297/98/6304-0043$02.00

Am. J. Hum. Genet. 63:1237–1240, 1998

On the Probability of Neanderthal Ancestry

To the Editor:

The controversial relationship between Neanderthals and modern humans recently received much attention, owing to the recovery of a Neanderthal mtDNA fragment, the analysis of which indicated that the most-recent common ancestor (MRCA) of Neanderthal and modern-human mitochondria was several times more ancient than that of modern humans only (Krings et al. 1997; fig. 1). This finding was considered to be strong evidence that Neanderthals and anatomically modern humans are separate species, the latter having replaced the former without interbreeding ("In our genes?" 1997; Kahn and Gibbons 1997; Lindahl 1997; Wade 1997; Ward and Stringer 1997). Here, I investigate the strength of this evidence by considering the probability of erroneous rejection of interbreeding (i.e., the probability of a type I error). I demonstrate that, although completely random mating clearly can be rejected, more-relevant models of interbreeding cannot.

The question of whether Neanderthals and anatomically modern humans interbred is a question of ancient levels of gene flow. Thus, although the relevant features of the data can be conveniently summarized as in figure 1, this figure is not, a priori, a phylogenetic tree for Neanderthals and humans: indeed, the question is whether such a tree exists. Figure 1 is simply a genealogical tree representing the history of the sampled mtDNA. In the following discussion, I ignore the considerable uncertainty in the estimation of this history and focus on the question of whether, given perfect knowledge of mtDNA genealogy, we would be able to conclude that anatomically modern humans and Neanderthals did not interbreed.

First, I consider whether Neanderthals and anatomically modern humans could have mated randomly. Two features of the data summarized in figure 1 provide evidence against such a scenario: The first is the topology, with the modern sample being monophyletic, and the second is the more than fourfold difference between T_r, the age of the MRCA of the modern humans and the Neaderthal, and T_e, the age of the MRCA of the modern humans only. If anatomically modern humans and Neanderthals mated randomly, the probability of such a result can be calculated as follows. Let $A_n(t) \in \{1,...,n\}$ be the random number of ancestors, at time t, of a sample of n mtDNAs at $t = 0$; its distribution is known under a variety of neutral models (Tavaré 1984). Conditional on $A_{986}(t_s) = k$, the number of ancestors of the modern sample who are contemporary with the sampled Neanderthal, the probability sought can be written as the product of the probability that a compatible topology is observed and the probability that sufficiently extreme coalescence times are observed. The former probability is easily shown to be $P[\text{topology} \mid A_{986}(t_s) = k] = 2/[k(1 + k)]$ (this also may be obtained as a special case of more-general results [Watterson 1982; Saunders et al. 1984]). An exact expression for the latter probability also can be obtained (T. Nagylaki and M. Nordborg, unpublished data) but is cumbersome and in some cases difficult to evaluate numerically. Estimation of the probability through standard Monte Carlo–simulation techniques is more convenient (e.g., Marjoram and Donnelly 1997).

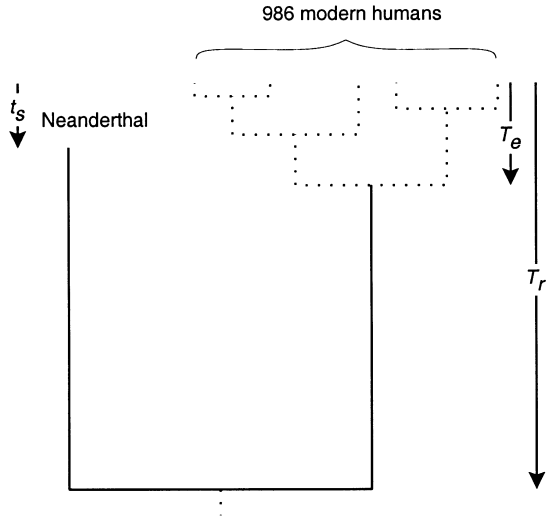

Figure 1 Schematic genealogy of the 986 modern-human mtDNAs and a single Neanderthal mtDNA (the carrier of which lived at time t_s before the present). The MRCA of the entire sample was inferred to be at least four times more ancient than the MRCA of the modern sample—that is, $T_r \geqslant 4T_e$ (Krings et al. 1997).

Two simple scenarios for human demography were used—namely, constant population size and constant ancient-population size followed by exponential growth 50,000 years ago. For both cases, the effective number of females in the constant population was assumed to be 3,400, growing exponentially to 5×10^8 for the latter case. These parameters were chosen so that the probability would be high that T_e lies within the range 100,000–200,000 years, when a generation time of 20 years is assumed. The age of the sampled Neanderthal, t_s, was assumed to be 30,000–100,000 years (the recovery of DNA more ancient than 100,000 years seems highly doubtful [Krings et al. 1997]). I argue below that the absolute values of all these parameters are of considerably lesser importance than their relative values.

Table 1 gives the results for models of random mating. As expected, the probability that both a compatible topology and an extreme difference between T_e and T_r would be observed is low, and, therefore, the hypothesis

Table 1

Results for Models of Random Mating

	Constant Population Size and t_s (in Years) =		Recent Population Growth and t_s (in years) =	
Parameter	30,000	100,000	30,000	100,000
$E[A_{986}(t_s)]$	4.86	1.75	782	2.86
P(topology)	.085	.56	3.3×10^{-6}	.24
P(topology and $T_r \geqslant 4T_e$)	.0063	.035	3.7×10^{-8}	.002

Note.—$E[A_{986}(t_s)]$ is the expected number of ancestors of the modern sample who are contemporary with the sampled Neanderthal. P(topology) is the probability that the topology in figure 1 would be observed, and P(topology and $T_r \geqslant 4T_e$) is the probability that both unlikely features of the data would be observed. All values were estimated through Monte Carlo simulation, as well as by calculation from the analytical results, except for those in the third column, for which the latter approach proved to be computationally too difficult. The 95% confidence intervals for the simulated values do not alter the decimals given. In the constant–population-size model, the expected T_e was ~136,000 years, with an SD of ~70,000 years; for recent exponential growth, the expected T_e was ~180,000 years, with, again, an SD of ~70,000 years.

that modern humans and Neanderthals were a randomly mating population may be rejected. However, closer inspection reveals the more interesting fact that the topology alone may not be unlikely. The reason for this is that, unless the sampled Neanderthal lived long after human populations had started to grow exponentially, most of the modern mtDNA lineages would have coalesced at t_s: if, for example, the modern sample only had two ancestors who were contemporary with the sampled Neanderthal, it would not be surprising if they were monophyletic (probability of 1/3). A large difference between T_e and T_r, on the other hand, is always unlikely under random mating.

Thus, the data constitute considerable evidence against the hypothesis that all sequences were drawn from a single population. This perhaps should not be surprising: the recovered Neanderthal sequence clearly was not sampled from a random individual at time t_s but was sampled specifically from an individual who was morphologically distinct from anatomically modern humans. Furthermore, fossil data strongly suggest that Neanderthals and anatomically modern humans were not a randomly mating population. To ask questions about interbreeding, more-interesting null hypotheses are needed. One pleasingly simple scenario is the following. Assume that Neanderthals were an isolated population for a long time, until they encountered anatomically modern humans at time t_m and *merged* with them to form a single, randomly mating population, with a fraction, c, of the population being Neanderthal. Then, the so-called replacement hypothesis is simply that $c = 0$. The data in figure 1 are perfectly consistent with this scenario; that is, the probability of the data is 1, without interbreeding. However, this provides support for replacement only to the extent that alternative scenarios can be shown to have a much lower probability. Therefore, the probability of the data must be found for different values of $c > 0$.

Under the assumption that the sampled Neanderthal lived before t_m (i.e., a "pure" Neanderthal), the probability sought is simply the probability that none of the ancestors at time t_m came from the Neanderthal fraction of the population. This probability can be written as $\sum_{k=1}^{986}(1-c)^k P[A_{986}(t_m) = k]$, which is the probability-generating function for $A_{986}(t_m)$. Figure 2 shows a plot for the two demographic scenarios described above, with t_m = 30,000 or 100,000 years. Clearly, for the scenarios in which the expected number of ancestors at t_m is low (table 1), the data tell us little about interbreeding, except perhaps that the Neanderthals did not make up the majority. The situation is completely different if the expected number of ancestors at t_m is high. In this case, all but very small values of c may be rejected.

In cases for which we expect few ancestors at t_m, the probability that none of the 986 *sampled* mtDNAs came from the Neanderthal fraction of the population does not differ much from the probability that none of the *currently existing* mtDNAs did so. This latter probability is equal to the well-known probability that an allele starting at frequency c is lost, through drift, by time t_m (Kimura 1955). Under this assumption, another question of interest can be addressed: Given that extant humans do not carry Neanderthal mtDNA, what does this sug-

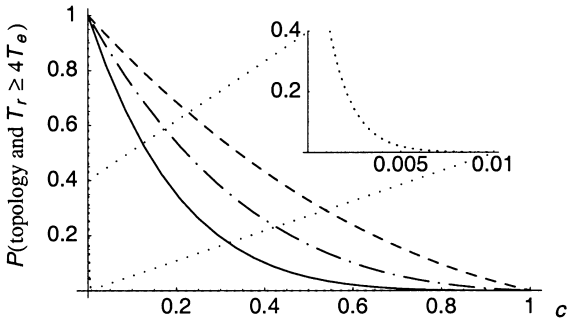

Figure 2 Probability of the data, if Neanderthals and anatomically modern humans merged at time t_m, with Neanderthals composing a fraction, c, of the new population. The four curves are for different demographic assumptions (see text) and values of t_m: constant population size, t_m = 30,000 years (solid line); constant population size, t_m = 100,000 years (dashed line); recent exponential growth, t_m = 30,000 years (dotted line [magnified in insert]); and recent exponential growth, t_m = 100,000 years (dotted-dashed line). The plots were calculated numerically by use of the known probability-generating function (Tavaré 1984), except for the third scenario, for which Monte Carlo simulation was used because of computational difficulties.

gest about the rest of the genome? For the constant–population-size model, for example, assume that Neanderthals and anatomically modern humans merged 1 coalescent-time unit ago (equivalent to $t_m = 68{,}000$ years, for the population size used above) and that Neanderthals composed 25% of the new population. Then, the probability that all Neanderthal mtDNA was lost through drift is .52 (the probability that Neanderthal mtDNA was not in the sample [calculated as above] is the same, to two decimal places). At the same time, each nuclear locus, for which the coalescence-time scale is four times slower, would have lost all Neanderthal alleles with probability .10 and would have become fixed for them with probability 9.8×10^{-5}. Thus, 90% would still be segregating for Neanderthal alleles.

In conclusion, data such as those shown in figure 1 shed little light on the issue of replacement versus interbreeding, unless the number of ancestors of the sample was large throughout the periods of interest. This is part of a general problem: in order to estimate gene flow, a large sample is needed, and, in order to estimate ancient-gene flow, a large ancient sample is needed. According to coalescent theory, large ancient samples usually cannot be obtained by the sampling of modern populations. The rate of coalescence is quadratic in the number of ancestors and linear in the inverse of the population size. Thus, the expected number of ancestors of a sample usually decreases rapidly as earlier time periods are studied. Exceptions include exponentially growing populations, in which the number of ancestors may be large shortly after the onset of growth (reviewed in Donnelly and Tavaré 1995; Marjoram and Donnelly 1997). In the present case, it seems clear that the statistical power to detect interbreeding that took place before the human population started to grow exponentially is close to zero.

I also have considered the mtDNA genealogy as known. The extreme uncertainty of the reconstruction of ancient DNA and the genealogy shown in figure 1 presumably suggests that conclusions from the data should be made with even more caution. Additional Neanderthal mtDNA sequence data would reduce these sources of uncertainty, but the main problem discussed above can be alleviated only by the study of data from several unlinked loci. The fact remains that an inference about population properties that is based on a single locus (or a nonrecombining genome) is an inference from a single data point. This does not mean that single loci contain no information: I have shown that random mating can be rejected, and the existence of a single Neanderthal mtDNA that differed little from modern mtDNA would allow rejection of the hypothesis that there was no interbreeding. Such an observation probably could never be made, however, since contamination would be impossible to rule out.

Finally, the above analysis depends on the selective neutrality of mtDNA variation. It is well known that human mtDNA variation suggests a genealogy that is "star shaped": this has been interpreted as the result of a historical population expansion (Di Rienzo and Wilson 1991; Merriwether et al. 1991; Vigilant et al. 1991; Rogers and Harpending 1992). However, data from several nuclear loci do not show this pattern (Harding et al. 1997; Hey 1997). Together, these observations may constitute evidence against neutrality, with a plausible alternative being a recent selective sweep in human mtDNA (Hey 1997). The conclusions in this paper clearly are not robust to this type of violation of assumptions: if there has been a recent selective sweep in human mtDNA, even random mating cannot be rejected.

Acknowledgments

I thank B. Bengtsson, A. Di Rienzo, P. Donnelly, R. Harding, the reviewers, and especially T. Nagylaki, for their comments on the manuscript. This work was supported by the Erik Philip-Sörensen Foundation.

Magnus Nordborg

Department of Genetics
Lund University
Lund
Sweden

References

Di Rienzo A, Wilson AC (1991) The pattern of mitochondrial DNA variation is consistent with an early expansion of the human population. Proc Natl Acad Sci USA 88:1597–1601

Donnelly P, Tavaré S (1995) Coalescents and genealogical structure under neutrality. Annu Rev Genet 29:401–421

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, et al (1997) Archaic African *and* Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:772–789

Hey J (1997) Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol Biol Evol 14: 166–172

In our genes? (1997) The Economist 344(8025), July 12th, pp 71–72

Kahn P, Gibbons A (1997) DNA from an extinct human. Science 277:176–178

Kimura M (1955) Solution of a process of random genetic drift with a continuous model. Proc Natl Acad Sci USA 41: 144–150

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997) Neanderthal DNA sequences and the origin of modern humans. Cell 90:19–30

Lindahl T (1997) Facts and artifacts of ancient DNA. Cell 90: 1–3

Marjoram P, Donnelly P (1997) Human demography and the time since mitochondrial Eve. In: Donnelly P, Tavaré S (eds)

Progress in population genetics and human evolution. Springer-Verlag, New York, pp 107–131
Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, et al (1991) The structure of human mitochondrial DNA variation. J Mol Evol 33: 543–555
Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–569
Saunders IW, Tavaré S, Watterson GA (1984) On the genealogy of nested subsamples from a haploid population. Adv Appl Prob 16:471–491
Tavaré S (1984) Line-of-descent and genealogical processes, and their applications in population genetic models. Theor Popul Biol 26:119–164
Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507
Wade N (1997) Neanderthal DNA sheds new light on human origins. New York Times, July 11, sec A
Ward R, Stringer C (1997) A molecular handle on the Neanderthals. Nature 388:225–226
Watterson GA (1982) Mutant substitutions at linked nucleotide sites. Adv Appl Prob 14:206–224

Address for correspondence and reprints: Dr. Magnus Nordborg, Department of Genetics, Lund University, Sölvegatan 29, 223 62 Lund, Sweden. E-mail: magnus.nordborg@gen.lu.se

0002-9297/98/6304-0043$02.00

Am. J. Hum. Genet. 63:1240–1242, 1998

Do Human Chromosomal Bands 16p13 and 22q11-13 Share Ancestral Origins?

To the Editor:

Ancient duplications and rearrangements within a genome are believed to be important mechanisms of evolution. Although most duplications are of gene segments, single genes, or chromosomal segments, molecular evidence has been gathered suggesting that whole-genome duplication has facilitated evolution in yeast (Wolfe and Shields 1997). Identifying these duplicated genomic areas can be valuable not only for understanding the timing and nature of evolutionary events; additionally, this information can greatly facilitate the pinpointing of novel (disease-related) genes by positional cloning techniques.

While mapping and cloning the human gene encoding the CREB-binding protein (CBP, encoded by the *CREBBP* gene) on chromosome band 16p13.3 (Giles et al. 1997*b*), we noticed an emerging pattern concerning the genomic relationship between this chromosome band and a region of chromosome 22q. CBP exhibits extensive homology to the adenovirus E1A–associated protein p300, whose gene has been mapped to human chromosome band 22q13 (Eckner et al. 1994; Lundblad et al. 1995). At that time we noted with interest that the heme oxygenase-1 (*HMOX1*) gene, just centromeric of *CREBBP* on 16p13.3, has a paralogue mapping to chromosome band 22q12, heme oxygenase-2 (*HMOX2*; Kutty et al. 1993). Our interest was further piqued when the molecular defect in families with carbohydrate-deficient glycoprotein type I syndrome (CDG1) was determined to be caused by mutations in the phosphomannomutase 2 gene (*PMM2*) on 16p13 (Matthijs et al. 1997*a*); the same investigators had previously mapped the first phosphomannomutase gene (*PMM1*) to 22q13 (Matthijs et al. 1997*b*). Sequence comparison at the amino acid level revealed that homologies between these paralogous proteins are high: homology between CBP and p300 is 63% (Arany et al. 1995), that between PMM1 and PMM2 is 66% (Matthijs et al. 1997*a*), and that between HMOX1 and HMOX2 is 74% (authors' observation). Subsequent examination of genome databases (e.g., OMIM) resulted in six additional sets of paralogues mapping to chromosomes 16p13 and 22q11-13, although the extent of homology between these paralogue sets is not known (table 1). YAC contigs connecting outlying genes of each paralogous cluster, *CREBBP* to *MYH11* on chromosome 16 and the *CRYB* genes to *PMM1* on chromosome 22, suggest that the extent of the redundant area presented here is ~12–14 Mb. Furthermore, *CREBBP* and *MYH11* are also thought to be near the borders for the conserved synteny group in mouse chromosome 16 (Doggett et al. 1996).

We propose that the existence of these paralogous sets suggests that chromosome bands 16p13 and 22q11-13 share ancestral origins and that at some point a large-scale duplication gave rise to this second set of genes. It is well established that such duplicated regions exist (Lundin 1993; Holland et al. 1994), and a catalogue of putative paralogous regions can be found on-line (Database of Duplicated Human Chromosomal Regions). This database suggests two duplicated regions for areas of 16p: a well-documented gene cluster on chromosome band 16p11.1, which shares high homology with a locus on Xq28 (Eichler et al. 1996), and a region of 16p13, which resembles 19p13, although no specific genes are named.

A hypothesis set forth by Ohno (1993) suggests that at the stage of fish, the mammalian ancestral genome underwent tetraploid duplication. Although certain aspects of this hypothesis are not universally accepted, most scientists agree that the fourfold increase, in the number of genes, between invertebrates and vertebrates implies at least two rounds of genome duplication (Aparicio 1998). Paralogues such as the HOX-

Table 1

Paralogous Genes Mapped to Chromosome Bands 16p13 and 22q11-13

Paralogues		
Gene/Chromosome[a]	Gene /Chromosome	Description
SSTR5/16p13.3	SSTR3/22q13.1	Somatostatin receptors
CREBBP/16p13.3	p300/22q13	Transcriptional cofactors
CSNK2A1/16p13.3	CSNK1E/22q12-13	Casein kinase isoforms
UBE2I/16p13.3	UBE2L3/22q11.2-13.1	Ubiquitin-conjugating enzymes
PMM2/16p13.3	PMM1/22q13.1	Phosphomannomutase isoforms
HMOX2/16p13.3	HMOX1/22q12	Heme oxygenase isoforms
MYH11/16p13.13-13.12	MYH9/22q11.2	Myosin heavy-chain subunits
CRYM/16p13.11-12.3	CRYBB1/22q11.2-12.1	Crystallin isoforms
	CRYB2/22q11.2-12.2	
	CRYB3/22q11.2-12.2	
	CRYBA4/22q11.2-13.1	
IL4R/16p12	IL2RB/22q12	Interleukin receptors

[a] Listed from telomere to centromere

gene clusters, which are situated at four distinct chromosomal loci, bolster this hypothesis. If the gene redundancy observed on chromosomes 16 and 22 is a result of Ohno's proposed ancestral event, then one might expect that two additional loci exist in the human genome that shares at least partial homology. CBP and p300 do, in fact, count two additional protein family members, p270 (Dallas et al. 1997) and p400 (Barbeau et al. 1994), although the genes for these proteins have not yet been mapped. Candidate regions, however, can be inferred from the literature. For example, clues can be taken from the somatic translocation t(8;16)(p11;p13.3), associated with acute myeloid leukemia, which disrupts the *CREBBP* gene and fuses it to a gene on chromosome 8, called "*MOZ*" (Borrow et al. 1996; Giles et al. 1997*a*). Phenotype-identical variants of the t(8;16) have been described: the t(8;22)(p11;q13), postulated to fuse *p300* to *MOZ*, as well as t(6;8)(q27;p11) (Tanzer et al. 1988), t(8;19)(p11;q13.2) (Tanzer et al. 1988; Stark et al. 1995), t(8;14)(p11;q11.1) (Slovak et al. 1991), and t(3;8;17)(q27;p11;q12) (Bertheas et al. 1989). If it is assumed that these phenotypically similar leukemias all fuse *MOZ* to genes situated at the breakpoints on chromosome bands 3q27, 6q27, 14q11.1, 17q12, or 19q13.2, then these loci become good candidates for the *p270/p400* genes—and, thus, for additional redundant clusters. Interestingly, two of these loci do harbor additional gene-family members paralogous to those mapping to 16p13 and 22q11-q13 (table 1): the *SSTR1, UBE2L1, MYH6,* and *MYH7* genes map to chromosome bands 14q11-q13, whereas the *SSTR2, CSNK1D,* and *CRYBA1* genes map to chromosome 17q11-q25. The gene-mapping data coupled with the leukemia breakpoint locations strongly suggest that these gene families have arisen by tetrapoidization with members on chromosomes 14q, 16p, 17q, and 22q.

Genetic redundancy is potentially of great relevance to organismal evolution, since it may protect organisms from potentially harmful mutations and may provide a pool of diverse yet functionally similar proteins for further evolution. Transcription factors such as CBP and p300 are thought particularly to "profit" from redundancy, as demonstrated by recent knockout mouse studies, which show that the combined dose of CBP and p300 is essential for survival (reviewed by Giles 1998). The existence of these duplicated gene clusters is not just a matter of redundancy; in the cases of CBP/p300 and PMM1/PMM2, the proteins have been shown to be functionally divergent. Where in vitro experiments suggest almost complete functional redundancy, CBP and p300 are clearly not physiologically interchangeable (reviewed by Giles et al. 1998); inactivating germ-line mutations of one copy of the *CREBBP* gene cause the Rubinstein-Taybi syndrome (Petrij et al. 1995). Likewise, mutations in *PMM2,* but not those in *PMM1,* result in CDG1 (Matthijs et al. 1997*a;* Schollen et al. 1998).

Rachel H. Giles,[1] Hans G. Dauwerse,[1]
Gert-Jan B. van Ommen,[1] and
Martijn H. Breuning[2]
Departments of [1]Human Genetics and [2]Clinical Genetics, Leiden University Medical Center, Leiden

Electronic-Database Information

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

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim

Database of Duplicated Human Chromosomal Regions, http://www.cib.nig.ac.jp/dda/timanish/dup.html

References

Aparicio S (1998) Exploding vertebrate genomes. Nat Genet 18:301–303

Arany Z, Newsome D, Oldread E, Livingston DM, Eckner R (1995) A family of transcriptional adaptor proteins targeted by the E1A oncoprotein. Nature 374:81–84

Barbeau D, Charbonneau R, Whalen SG, Bayley ST, Branton PE (1994) Functional interactions within adenovirus E1A protein complexes. Oncogene 9:359–373

Bertheas MF, Jaubert J, Vasselon C, Reynaud J, Pomier G, Le Petit JC, Hagemeijer A, et al (1989) A complex t(3;8;17) involving breakpoint 8p11 in a case of M5 acute nonlymphocytic leukemia with erythrophagocytosis. Cancer Genet Cytogenet 42:67–73

Borrow J, Stanton VP Jr, Andresen JM, Becher R, Behm FG, Chaganti RSK, Civin CI, et al (1996) The translocation t(8;16)(p11;p13) of acute myeloid leukemia fuses a putative acetyl transferase to the CREB-binding protein. Nat Genet 14:33–41

Dallas PB, Yaciuk P, Moran E (1997) Characterization of monoclonal antibodies raised against p300: both p300 and CBP are present in intracellular TBP complexes. J Virol 71: 1726–1731

Doggett NA, Breuning MH, Callen DF (1996) Report of the Fourth International Workshop on Human Chromosome 16 Mapping 1995. Cytogenet Cell Genet 72:271–293

Eckner R, Ewen ME, Newsome D, Gerdes M, DeCaprio JA, Lawrence JB, Livingston DM (1994) Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes Dev 8:869–884

Eichler EE, Lu F, Shen Y, Antonacci R, Jurecic V, Doggett NA, Moyzis RK, et al (1996) Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. Hum Mol Genet 5:899–912

Giles RH (1998) CBP/p300 transgenic mice. Trends Genet 14: 214

Giles RH, Dauwerse JG, Higgins C, Petrij F, Wessels JW, Beverstock GC, Döhner H, et al (1997*a*) Detection of *CBP* rearrangements in acute myelogenous leukemia with t(8;16). Leukemia 11:2087–2096

Giles RH, Peters DJM, Breuning MH (1998) Conjunction dysfunction: CBP/p300 in human disease. Trends Genet 14: 178–183

Giles RH, Petrij F, Dauwerse JG, den Hollander AI, Lushnikova T, van Ommen G-JB, Goodman RH, et al (1997*b*) Construction of a 1.2-Mb contig surrounding, and molecular analysis of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome 16p13.3. Genomics 42: 96–114

Holland PW, Garcia-Fernandez J, Williams NA, Sidow A (1994) Gene duplications and the origins of vertebrate development. Dev Suppl 125–133

Kutty RK, Kutty G, Rodriguez IR, Chader GJ, Wiggert B (1994) Chromosomal localization of the human heme oxygenase genes: heme oxygenase-1 (HMOX1) maps to chromosome 22q12 and heme oxygenase-2 (HMOX2) maps to chromosome 16p13.3. Genomics 20:513–516

Lundblad JR , Kwok RPS, Laurance ME, Harter ML, Goodman RH (1995) Adenoviral E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP. Nature 374:85–88

Lundin LG (1993) Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics 16:1–19

Matthijs G, Schollen E, Pardon E, Veiga-Da-Cuhna M, Jaeken J, Cassiman J-J, van Schaftingen E (1997*a*) Mutations in PMM2, a phosphomannomutase gene on chromosome 16p13, in carbohydrate-deficient glycoprotein type I syndrome (Jaeken syndrome). Nat Genet 16:88–92

Matthijs G, Schollen E, Pirard M, Budarf ML, van Schaftingen E, Cassiman J-J (1997*b*) PMM (PMM1), the human homologue of SEC53 or yeast phosphomannomutase, is localized on chromosome 22q13. Genomics 40:41–47

Ohno S (1993) Patterns in genome evolution. Curr Opin Genet Dev 3:911–914

Petrij F, Giles RH, Dauwerse JG, Saris JJ, Hennekam RC, Masuno M, Tommerup N, et al (1995) Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. Nature 376:348–351

Schollen E, Pardon E, Heykants L, Renard J, Doggett NA, Callen DF, Cassiman J-J, et al (1998) Comparative analysis of the phosphomannomutase genes PMM1, PMM2 and PMM2Ψ: the sequence variation in the processed pseudogene is a reflection of the mutations found in the functional gene. Hum Mol Genet 7:157–164

Slovak ML, Nemana L, Traweek ST, Stroh JA (1991) Acute monoblastic leukemia (FAB-M5b) with t(8;14)(p11;q11.1). Cancer Genet Cytogenet 56:237–242

Stark B, Resnitzky P, Jeison M, Luria D, Blau O, Avigad S, Shaft D, et al (1995) A distinct subtype of M4/M5 acute myeloblastic leukemia (AML) associated with t(8;16)(p11;p13), in a patient with the variant t(8;19)(p11;q13): case report and review of the literature. Leuk Res 19:367–379

Tanzer J, Brizard A, Guilhot F, Benz-Lemoine E, Dreyfus B, Lessard M, Herchkovitch C, et al (1988) La leucémie aigüe à translocation (8;16). Nouv Rev Fr Hematol 30:83–87

Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708–713

Address for correspondence and reprints: Dr. Rachel H. Giles, Department of Human Genetics, Leiden University Medical Center, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands. E-mail: rachel@ruly46.medfac.leidenuniv.nl

0002-9297/98/6304-0044$02.00

Am. J. Hum. Genet. 63:1242–1245, 1998

How Sib Pairs Reveal Linkage

To the Editor:

The Haseman-Elston (1972) method, widely used for

studying linkage, has been criticized for incomplete utilization of sib-pair information. As an alternative, Amos (1994) created and advocates the "variance-components" approach; Wright (1997), using a "likelihood argument," found that the phenotypic difference discards sib-pair linkage information; and Fulker and Cherny (1996) came to a similar conclusion after an analysis of sib-pair covariances (Fulker et al. 1995). Here, I propose an extension of the Haseman-Elston (1972) method that puts the sib-trait sum into linkage testing.

Suppose a trait X has a normal distribution with a mean genetically determined and environmental (residual) variance σ_e^2; each sib pair has i alleles identical by descent (IBD) at the trait locus, $i = 0$, 1, or 2; and the sib pair–trait vector $X^T \equiv (X_1, X_2)^T$ has joint normal (binormal) distribution:

$$\mathcal{F}(X) = \frac{1}{2\pi\sqrt{|\Sigma_x|}}\exp\left[-\tfrac{1}{2}(X-\mu)^T\Sigma_x^{-1}(X-\mu)\right] ,$$

where μ is the overall mean and the symbol T stands for "transpose." The matrix Σ_x^{-1} is the inverse of the variance-covariance matrix, which has the form

$$\Sigma_x = \begin{pmatrix} v & c \\ c & v \end{pmatrix} ,$$

where

$$\begin{aligned} v = \mathrm{var}(X) &= V_p + V_c + \tfrac{1}{2}V_e + V_a + V_d \\ &= V_p + V_c + \tfrac{1}{2}V_e + V_g , \end{aligned}$$

$$\begin{aligned} c = \mathrm{cov}(X_1, X_2) &= \tfrac{1}{2}V_p + V_c + \tfrac{1}{2}iV_a + \tfrac{1}{2}i(i-1)V_d \\ &= \tfrac{1}{2}V_p + V_c + \tfrac{1}{2}iV_g + \tfrac{1}{2}i(i-2)V_d , \end{aligned} \quad (1)$$

and the variances are as follows: polygenic, V_p; common environment, V_c; additive genetic, V_a; dominance genetic, V_d; residual, V_e ($= 2\sigma_e^2$); and total genetic, $V_g = V_a + V_d$ (Malécot 1966, p. 320; Amos 1994; Fulker and Cherny 1996).

Let us introduce two new variables: $D = X_1 - X_2$, and $S = X_1 + X_2$. By use of matrix algebra methods, it is easy to show that the variance-covariance matrix of D and S is diagonal:

$$\Sigma = \begin{bmatrix} 2(v-c) & 0 \\ 0 & 2(v+c) \end{bmatrix} ;$$

thus, these new "coordinates" are uncorrelated, each of them having the normal distribution, and their joint distribution is

$$\mathcal{F}(D,S) = \frac{1}{2\pi\sigma_D\sigma_S}\exp\left[-\frac{D^2}{2\sigma_D^2} - \frac{(S-2\mu)^2}{2\sigma_S^2}\right] .$$

The variances are $\sigma_D^2 = 2(v-c)$ and $\sigma_S^2 = 2(v+c)$. Instead of variances, let us consider the squared pair-trait difference $Y \equiv D^2$ and the squared pair sum $Z \equiv S^2$:

$$\begin{aligned} \mathcal{E}(Y|i) &= \sigma_D^2 \\ &= (V_e + V_p + 2V_g) - iV_g + i(2-i)V_d , \end{aligned} \quad (2)$$

and

$$\begin{aligned} \mathcal{E}(Z|i) &= \sigma_S^2 + 4\mu^2 \\ &= (V_e + 3V_p + 4V_c + 2V_g + 4\mu^2) + iV_g \\ &\quad - i(2-i)V_d , \end{aligned} \quad (3)$$

where the symbol $\mathcal{E}$ stands for "expectation." The squared pair-trait difference, Y, has been studied (Haseman and Elston 1972; Blackwelder and Elston 1982).

Each of the variables (2) and (3) is a function of the number of alleles IBD, i, at the trait locus, and their expected values, conditional on the marker information, are of interest:

$$\mathcal{E}(\ldots|M) = \sum_{i=0,1,2}\mathcal{E}(\ldots|i)P(i|M) , \quad (4)$$

where $f_i \equiv P(i|M)$ is the probability of i alleles IBD ($i = 0$, 1, or 2) at the trait locus. The expectations are

$$\begin{aligned} \mathcal{E}(Y|M) &= \sum_{i=0,1,2}[(V_e + V_p + 2V_g) - iV_g \\ &\quad + i(2-i)V_d]f_i \\ &= (V_e + V_p + 2V_g) - 2V_g(\pi + \tfrac{1}{2}) \\ &\quad + V_d(\varphi + \tfrac{1}{2}) \\ &= (V_e + V_p + V_g + \tfrac{1}{2}V_d) - 2V_g\pi + V_d\varphi \end{aligned}$$

and

$$\begin{aligned}\mathcal{E}(Z\,|\,M) &= \sum_{i=0,1,2}[(V_e + 3V_p + 4V_c + 2V_g + 4\mu^2) \\ &\quad + iV_g - i(2-i)V_d]f_i \\ &= (V_e + 3V_p + 4V_c + 2V_g + 4\mu^2) \\ &\quad + 2V_g(\pi + \tfrac{1}{2}) - V_d(\varphi + \tfrac{1}{2}) \\ &= (V_e + 3V_p + 4V_c + 3V_g - \tfrac{1}{2}V_d + 4\mu^2) \\ &\quad + 2V_g\pi - V_d\varphi\ ,\end{aligned}$$

where

$$\pi \equiv \tfrac{1}{2}f_1 + f_2 - \tfrac{1}{2} \quad \text{and} \quad \varphi \equiv f_1 - \tfrac{1}{2} \tag{5}$$

at the trait locus. These definitions of π and φ differ from those introduced by Haseman and Elston (1972) and used by Blackwelder and Elston (1982) by the term $\frac{1}{2}$. So defined, π and φ are proportional to the same functions (5) of $\{f_i\}$, calculated at the marker locus (Drigalenko, in press):

$$\pi = \eta\pi_m\ , \qquad \varphi = \eta^2\varphi_m\ , \tag{6}$$

where π_m and φ_m are calculated on the basis of relatives' marker phenotypes, $\eta = (1-2r)^2$, and r is the recombination coefficient between the trait locus and the marker locus that depends on the (unknown) distance between them. Finally, the regression equations become

$$\begin{aligned}\mathcal{E}(Y\,|\,M) &= (V_e + V_p + V_g + \tfrac{1}{2}V_d) - 2V_g\eta\pi_m \\ &\quad + V_d\eta^2\varphi_m \\ &= \alpha_D - \beta\pi_m + \gamma\varphi_m\end{aligned} \tag{7}$$

and

$$\begin{aligned}-\mathcal{E}(Z\,|\,M) &= -(V_e + 3V_p + 4V_c + 3V_g \\ &\quad - \tfrac{1}{2}V_d + 4\mu^2) - 2V_g\eta\pi_m + V_d\eta^2\varphi_m \\ &= \alpha_S - \beta\pi_m + \gamma\varphi_m\ ,\end{aligned} \tag{8}$$

where $\alpha_D \equiv V_e + V_p + V_g + \frac{1}{2}V_d$, $\alpha_S \equiv -(V_e + 3V_p + 4V_c + 3V_g - \frac{1}{2}V_d + 4\mu^2)$, $\beta \equiv V_g\eta$, and $\gamma \equiv V_d\eta^2$. So, consideration of the squared pair sum of the trait values (taken with the opposite sign) results in a regression line that is parallel to that for the squared pair difference. Since seven parameters are unknown (V_e, V_p, V_c, V_g, V_d, μ, and η) and four regression coefficients are independent (α_D, α_S, β, and γ), all the parameters cannot be estimated. Note that only the slopes, β and γ, are important for testing linkage (Haseman and Elston 1972; Blackwelder and Elston 1982) and that these are the same for the sum and the difference of the sib pair–trait values.

The method described here uses all the information from the sib pair. To demonstrate the gain obtained when the sum and the difference are used together, let us ignore dominance, suppose that the residuals have the same variance in (6) and (7), and use Student's t-statistic to test the hypothesis H_0: $\beta = 0$. Then, joint use of the sum and the difference (rather than the difference alone) doubles the number of points on the regression line and, therefore, doubles the estimated values of both β and its variance, so that the t-statistic is enlarged by a factor of $\sim\sqrt{2}$, increasing the power of the test. Fulker and Cherny (1996, fig. 1) obtained similar results using simulated data and maximum-likelihood estimation.

More explicitly, for N sib pairs, indexed by j ($j = 1, \ldots, N$), the regression equations (7) and (8) include residuals ε_D and ε_S, assumed to be normally distributed and common for each sib pair (the dominance is ignored):

$$Y_j = \alpha_D - \beta\pi_j + \varepsilon_D\ , \quad -Z_j = \alpha_D - \beta\pi_j + \varepsilon_S\ . \tag{9}$$

These regression lines give the least-squares estimates of the slope:

$$\begin{aligned}\hat{\beta}_D &= \frac{N\Sigma Y_j\pi_j - \Sigma Y_j\Sigma\pi_j}{N\Sigma\pi_j^2 - (\Sigma\pi_j)^2}\ , \\ \hat{\beta}_S &= \frac{N\Sigma Z_j\pi_j - \Sigma Z_j\Sigma\pi_j}{N\Sigma\pi_j^2 - (\Sigma\pi_j)^2}\ .\end{aligned} \tag{10}$$

Under the assumption that the residuals have the same variance in (9), $\mathrm{var}(\varepsilon_D) = \mathrm{var}(\varepsilon_S)$, it is easy to prove that the least-squares estimate of the slope based on combined data for D and S (denoted by $D\oplus S$) is

$$\begin{aligned}\hat{\beta}_{D\oplus S} &= \frac{N\Sigma[(Y_j - Z_j)/2]\pi_j - \Sigma[(Y_j - Z_j)/2]\Sigma\pi_j}{N\Sigma\pi_j^2 - (\Sigma\pi_j)^2} \\ &= \tfrac{1}{2}(\hat{\beta}_D + \hat{\beta}_S)\ ,\end{aligned} \tag{11}$$

that is, the "combined" regression line is exactly between the two individual lines. Owing to the properties of variances,

$$\begin{aligned}\mathrm{var}(\hat{\beta}_{D\oplus S}) &= \mathrm{var}\left[\tfrac{1}{2}(\hat{\beta}_D + \hat{\beta}_S)\right] \\ &= \tfrac{1}{4}[\mathrm{var}(\hat{\beta}_D) + \mathrm{var}(\hat{\beta}_S)]\ ,\end{aligned}$$

because $\mathrm{cov}(\hat{\beta}_D, \hat{\beta}_S) = 0$, which is easy to see from (10) under the condition of $\mathrm{cov}(Y_j, Z_j) = 0$, discussed above. Hence, the estimate based on combined data for D and S has the smallest variance, that is, it is the most effective.

Note that, for every pair, (11) is based on the half-difference of Y and Z, which is $\frac{1}{2}(Y - Z) = \frac{1}{2}[(X_1 - X_2)^2 - (X_1 + X_2)^2] = -2X_1X_2$. The half-sum of (7) and (8) gives the equation

$$\mathcal{E}(-2X_1X_2|M) = \tfrac{1}{2}(\alpha_D + \alpha_S) - \beta\pi_{\mathrm{m}} + \gamma\varphi_{\mathrm{m}} ,$$

which may be easily derived from (1) and (4). Thus, the most clear estimate, $\hat{\beta}_{D\oplus S}$, is based on the pair-trait multiplication, because the linkage test depends on the number of alleles IBD (which is a characteristic of a pair rather than an individual); the covariance (1) gives the same information as any combination of the squared pair-trait difference and the squared pair sum. This explains the effectiveness of the variance-components method (Amos 1994).

Acknowledgments

I thank Dr. Robert Elston for pointing out this issue and for his help in the interpretation of the results. This work was supported by research grant F05 TW05285 from the Fogarty International Center and resource grant P41 RR03655 from the National Center for Research Resources, National Institutes of Health.

EUGENE DRIGALENKO
Department of Epidemiology and Biostatistics
Rammelkamp Center for Education and Research
MetroHealth Campus
Case Western Reserve University
Cleveland

References

Amos CI (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 54: 535–543

Blackwelder WC, Elston RC (1982) Power and robustness of sib-pair linkage tests and extension to larger sibships. Commun Stat Theor Methods 11:449–484

Drigalenko E. Matrix representation of the Haseman-Elston method. Theor Popul Biol (in press)

Fulker DW, Cherny SS (1996) An improved multipoint sib-pair analysis of quantitative traits. Behav Genet 26:527–532

Fulker DW, Cherny SS, Cardon LR (1995) Multipoint interval mapping of quantitative trait loci, using sib pairs. Am J Hum Genet 56:1224–1233

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Malécot G (1966) Probabilités et hérédité. Presses universitaires de France, Paris

Wright FA (1997) The phenotypic difference discards sib-pair QTL linkage information. Am J Hum Genet 60:740–742

Address for correspondence and reprints: Dr. Eugene Drigalenko, Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, 2500 MetroHealth Drive, Room R258, Cleveland, OH 44109-1998. E-mail: dei@darwin.mhmc.cwru.edu

0002-9297/98/6304-0045$02.00

Am. J. Hum. Genet. 63:1245–1247, 1998

Allele Identical by Descent Sharing at Any Point of a Chromosome of a Sib Pair

To the Editor:

The distribution of identical by descent (IBD) alleles on a chromosome is a key component of multipoint linkage analysis (Goldgar 1990; Kruglyak and Lander 1995; Whittemore 1996). Goldgar (1990) and Guo (1994) considered a proportion of genetic material shared IBD by sibling pairs. Kruglyak and Lander (1995) used "inheritance vectors" (Lander and Green 1987) to calculate the probability that a sib pair shares 0, 1, or 2 alleles IBD. I propose a simple and straightforward procedure, based on the Haseman and Elston (1972) approach.

Suppose a chromosome has m markers, the distances between them being known. Assuming no crossover interference, the Haldane mapping function is used. Family data on marker phenotypes provide the probability $f_{i_k} \equiv P(i_k|M_k)$ that a sib pair has i_k alleles IBD at the kth marker loci, for $k = 1, 2, \ldots, m$ and $i_k = 0, 1,$ or 2 (Haseman and Elston 1972, table 2). Denote by z the coordinate, on the chromosome, of the point studied that is between markers c ("closest") and $c + 1$.

The probability $P(i_z|M)$ that a sib pair shares i_z alleles IBD ($i_z = 0$, 1, or 2) at a point z, conditional on all the marker data M, is calculated by use of the formulas of total and conditional ("chain") probabilities:

$$\begin{aligned} P(i_z|M) &= \sum_{i_1,\ldots,i_m} P(i_z, i_1, \ldots, i_m|M) \\ &= \sum_{i_1,\ldots,i_m} P(i_1|i_2, \ldots, i_m, M)P(i_2|i_3, \ldots, i_m, M) \\ &\quad \times \cdots \times P(i_c|i_z, i_{c+1}, \ldots, i_m, M) \\ &\quad \times P(i_z, i_{c+1}, \ldots, i_m|M) \ . \end{aligned}$$

With the important assumption of no crossover interference, the allele sharing at any locus depends only on the marker data and the neighboring locus:

$$P(i_k|i_{k+1}, \ldots, i_m, M) = P(i_k|i_{k+1}, M_k) = P(i_k, i_{k+1}, M_k)/P(i_{k+1}, M_k) = P(i_k|i_{k+1})P(i_k|M_k)/P(i_k) \ .$$

The unconditional probabilities $P(i)$ are $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ for $i = 0$, 1, and 2, respectively. The conditional probability $\Psi'_{kl} \equiv P(k|l)$ that the sib pair has k alleles IBD at one locus if this pair has l alleles IBD at another locus is based on the corresponding joint probability $\Psi''_{kl} \equiv P(k, l)$, derived by Haseman and Elston (1972, table 4). Therefore,

$$P(i_k|i_{k+1}, \ldots, i_m, M) = \Psi'_{i_k i_{k+1}} f_{i_i}/P(i_k) \equiv \Psi_{i_k i_{k+1}} f_{i_k} \ , \quad 1$$

where $\Psi_{i_k i_{k+1}} = \Psi'_{i_k i_{k+1}}/P(i_k) = \Psi''_{i_k i_{k+1}}/[P(i_k)P(i_{k+1})]$ or, in the matrix notation,

$$\Psi = \begin{bmatrix} 4\psi^2 & 4\psi(1-\psi) & 4(1-\psi)^2 \\ 4\psi(1-\psi) & 4(\frac{1}{2}-\psi+\psi^2) & 4\psi(1-\psi) \\ 4(1-\psi)^2 & 4\psi(1-\psi) & 4\psi^2 \end{bmatrix}, \quad 1$$

where $\psi = r^2 + (1-r)^2$, and r is the recombination fraction, calculated from the known distance between the marker loci studied, by use of the Haldane mapping function; the indices i_k and i_{k+1} are omitted for ψ, Ψ, and r.

The sum for the first marker is

$$\sum_{i_1} P(i_1|i_2, M_1) = \sum_{i_1} \Psi_{i_1 i_2} f_{i_1} \equiv f^{(1)}_{i_2} \ . \quad 1$$

The right notation emphasizes that the probabilities indexed for the second marker "picked up" information from the first one. For the second marker,

$$\sum_{i_1, i_2} P(i_1|i_2, M_1)P(i_2|i_3, M_2) = \sum_{i_2} \Psi_{i_2 i_3} f_{i_2} f^{(1)}_{i_2} \equiv f^{(1,2)}_{i_3} \quad 1$$

and so on, up to the closest marker to the left of the trait locus (included):

$$f^{(1,\ldots,c)}_{i_z} = \sum_{i_c} \Psi_{i_c i_z} f_{i_c} f^{(1,\ldots,c-1)}_{i_c} \ . \quad 1(1)$$

Remember that $\Psi_{i_c i_z}$ depends on z and that the Haldane mapping function is used. In the part of the chromosome to the right of point z, we proceed in the opposite direction:

$$P(i_z, i_{c+1}, \ldots, i_m|M) = P(i_m|i_{m-1}, M)P(i_{m-1}|i_{m-2}, M) \times \cdots \times P(i_{c+1}|i_z, M)P(i_z) \ . \quad 1$$

Again, by virtue of the assumption of no crossover interference,

$$P(i_k|i_{k-1}, \ldots, i_1, M) = P(i_k|i_{k-1}, M_k) = \Psi'_{i_{i-1} i_i} f_{i_i}/P(i_{k-1}) \equiv \Psi_{i_{i-1} i_i} f_{i_i} \ .$$

The sum for the last marker is

$$\sum_{i_m} P(i_m|i_{m-1}, M_m) = \sum_{i_m} \Psi_{i_m i_{m-1}} f_{i_m} \equiv f^{(m)}_{i_{m-1}} \ ;$$

then,

$$\sum_{i_{m-1}, i_m} P(i_{m-1}|i_{m-2}, M_{m-1}) = \sum_{i_{m-1}} \Psi_{i_{m-1} i_{m-2}} f_{i_{m-1}} f^{(m)}_{i_{m-1}} \equiv f^{(m-1,m)}_{i_{m-2}}$$

and so on, up to the closest marker to the right of the trait locus (included):

$$f^{(c+1,\ldots,m)}_{i_z} = \sum_{i_{c+1}} \Psi_{i_{c+1} i_z} f_{i_{c+1}} f^{(c+2,\ldots,m)}_{i_{c=1}} \ . \quad (2)$$

Finally, the probability at point z is the joint probability from the left (formula [1]) and right (formula [2]) parts of the chromosome:

$$P(i_z|M) = f^{(1,\ldots,c)}_{i_z} P(i_z) f^{(c+1,\ldots,m)}_{i_z} \equiv f^{(1,\ldots,m)}_{n_z} \ .$$

So, the prior probability $P(i_z)$ at point z is "corrected" by the marker data from both sides. When z is to the left of the first marker, $c = 0$ and the left factor disappears; when z is to the right of the last marker, $c = m$ and the right factor disappears; when z is at the position of the kth marker, $P(i_z|M_k)$ replaces $P(i_z)$, meaning that a noninformative marker receives information from its neighbors. If a marker is fully informative, only one of f_0, f_1, or f_2 is equal to 1; others are equal to 0, thus cutting the "probability chain."

The number of calculations is proportional to the number of marker loci in this multipoint method. If intermediate results are stored, this method leads to a fast algorithm for the calculation of allele IBD sharing at any point of a chromosome, for every sib pair. With this distribution, linkage tests for quantitative and qualitative traits may be derived, by use of likelihood, regression, scores, or other methods, which will be the subject of a separate communication.

Acknowledgments

This work was supported by research grant F05 TW05285 from the Fogarty International Center and resource grant P41 RR03655 from the National Center for Research Resources, National Institutes of Health.

EUGENE DRIGALENKO
Department of Epidemiology and Biostatistics
Rammelkamp Center for Education and Research
MetroHealth Campus
Case Western Reserve University
Cleveland

References

Goldgar DE (1990) Multipoint analysis of human quantitative genetic variation. Am J Hum Genet 47:957–967

Guo S-W (1994) Computation of identity-by-descent proportions shared by two siblings. Am J Hum Genet 54: 1104–1109

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439–454

Lander ES, Green P (1987) Construction of multipoint genetic linkage maps in humans. Proc Natl Acad Sci USA 84: 2363–2367

Whittemore AS (1996) Genome scanning for linkage: an overview. Am J Hum Genet 59:704–716

Address for correspondence and reprints: Dr. Eugene Drigalenko, Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, 2500 MetroHealth Drive, Room R258, Cleveland, OH 44109-1998. E-mail: dei@darwin.mhmc.cwru.edu

0002-9297/98/6304-0046$02.00