Letters to the Editor

Am. J. Hum. Genet. 63:648–651, 1998

Repeat Expansion–Detection Analysis of Telomeric Uninterrupted $(TTAGGG)_n$ Arrays

To the Editor:

In humans, as in all vertebrates, telomeres are nucleoprotein complexes at the ends of chromosomes and primarily consist of tandemly repeated double-stranded $(TTAGGG)_n$ hexamers, to which are bound various specific proteins. The telomeric TTAGGG repeats are replicated by conventional DNA polymerases and by the reverse transcriptase telomerase, which catalyzes the addition of new TTAGGG repeats to the 3′ ends of chromosomes (de Lange 1995; Meyerson et al. 1997). The replenishing of telomeric terminal repeats by telomerase is required in order to compensate for the loss of these sequences in incomplete replication of linear chromosomes during the S phase. As part of the aging process, normal somatic cells undergo a progressive loss of TTAGGG repeats, both in vitro and in vivo, which correlates with lack of expression of telomerase in most somatic cells (de Lange 1995). In transformed cell lines and tumors, this loss is overcome by the up-regulation of the putative human telomerase catalytic-subunit gene (hEST2) and reactivation of telomerase, a process that is stabilized further by the involvement of the human telomeric repeat–binding factor, TRF1 (Meyerson et al. 1997; Nakamura et al. 1997; Van Steensel and de Lange 1997). However, recent data from telomerase knockout mice suggest that the proposed involvement of telomerase with tumorigenesis may be coincidental and of no functional significance (Blasco et al. 1997). In telomerase-negative immortal cell lines and tumors, other, still-unidentified mechanisms must operate to provide for telomere maintenance and elongation (Bryan et al. 1995, 1997). In human cells, functional telomeres have been found to form at previously interstitial sites almost exclusively after transfection with TTAGGG repeats (and, more rarely, after transfection with TTAGGG-related heterologous sequences), thus demonstrating that there are stringent sequence requirements for the formation of human telomeres (Hanish et al. 1994).

Uninterrupted telomeric TTAGGG-repeat arrays in somatic cells currently are believed to be in the range of ~10-kb or more (de Lange et al. 1990; de Lange 1995). The $(TTAGGG)_n$-repeat maximum length has been estimated by physical mapping strategies, on the basis of measurement of genomic DNA restriction-fragment lengths after Southern blotting and hybridization either with a $(TTAGGG)_n$ probe or by chromosome-specific subtelomeric probes (Brown et al. 1990; de Lange 1995; Notaro et al. 1997) and, more recently, by using quantitative FISH (Martens et al. 1998). Accurate measurements by cloning and sequencing have been hampered by the high instability of the telomeric $(TTAGGG)_n$ and have failed to show any clone containing an unadulterated $(TTAGGG)_n$ array >540 bp (de Lange et al. 1990). We have used the repeat expansion–detection (RED) assay (Schalling et al. 1993), originally described for detection of long trinucleotide repeats in the human genome, to measure maximum TTAGGG-repeat lengths and to monitor the relative stability of the repeats in DNA templates from both blood and transformed cell cultures. The RED assay uses genomic DNA as a template for the annealing and ligation of repeat-specific oligonucleotides, does not require flanking sequence determination or single-copy probes, and detects the longest repeat of a given type present in the genome. The RED method is highly reproducible in our hands (Sirugo and Kidd 1995; Sirugo et al. 1997; also see Epicentre Forum Website). We have analyzed DNA templates extracted from whole blood of 21 northern Europeans, for the maximum length of uninterrupted TTAGGG repeats, and have found that the longest uninterrupted $(TTAGGG)_n$ arrays in such templates ranged from ~220 bp to a maximum length of ~480 bp, well below the several-kilobase range described in the literature (de Lange 1995) (fig. 1 and table 1). This range was conserved (and the results were repeatable) in all samples analyzed, suggesting that this is the $(TTAGGG)_n$ maximum-length range that normally is present, in vivo, in human leukocytes.

Unlike the indirect physical-mapping strategies used for measuring the maximum length of the telomeric hexamer, the RED method gives a direct measure of the actual size of the longest uninterrupted TTAGGG repeat in the genome. The observed $(TTAGGG)_n$ maximum length of ~480 bp cannot be explained on the basis of aging. According to previously published reports (Vaziri et al. 1993), telomere shortening in leukocytes that is due to senescence would correlate with a size reduction of telomeric restriction fragments, from ~10 to ~5 kb,

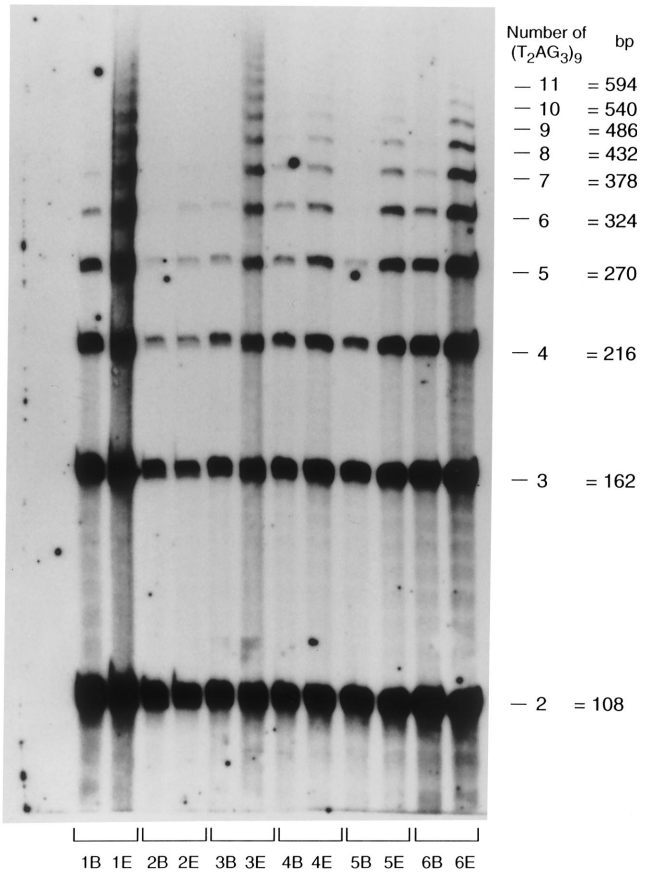

Figure 1 RED of $(TTAGGG)_n$ arrays, performed with a $(TTAGGG)_9$ oligonucleotide and 5 μg of DNA template in the conditions described by Sirugo and Kidd (1995; also see Epicentre Forum Website). Reaction products were fractionated by denaturing PAGE on 6% gels (20:1 acrylamide:bisacrylamide), were transferred onto a nylon membrane, and were detected by hybridization with a radiolabeled $(CCCTAA)_5$ probe. Lanes 1B–6E correspond to RED in DNA from blood/EBV-transformed lymphoblast pairs from six northern Europeans (Sirugo et al. 1997). An additional, seventh lymphoblast pair was analyzed (not shown). "B" and "E" denote, respectively, blood and EBV-cell-line DNA templates for each pair. The number of ligated $(TTAGGG)_9$ for each band is indicated in the figure, along with the corresponding size (in bp). All EBV-cell-line templates have been used for a previously published RED-based study of CTG/CAG-repeat maximum length in populations (Sirugo et al. 1997). Short CTG/CAG arrays were detected by RED in five of the seven lymphoblastoid cell samples (lanes 2E–6E) (Sirugo et al. 1997). We therefore can conclude that the repeat expansions detected in this study are specific for $(TTAGGG)_n$ arrays, after EBV transformation. One of these individuals (lanes 1B and 1E) was shown to carry a large CTG/CAG-trinucleotide–repeat expansion in templates from both blood and the EBV cell line. This rules out the possibility that the short $(TTAGGG)_9$ ligated ladders in blood are determined by poor quality of the DNA template with respect to the EBV-cell-line DNA. Fifteen additional northern European individuals also were analyzed; all carry $(TTAGGG)_n$ maximum lengths of 200–450 bp in their blood-extracted genomic DNAs (not shown) and provide confirmation that this is the normal $(TTAGGG)_n$ size range in this population sample.

implying a $(TTAGGG)_n$ maximum length much larger than the ~480 bp observed in our study. It is therefore likely that the ~10-kb, large regions so far believed to contain unadulterated $(TTAGGG)_n$ are, in fact, composed of shorter blocks of uninterrupted arrays intercalated with other motifs every 200–500 bp (Allshire et al. 1989).

From seven of the above-mentioned blood samples, Epstein-Barr virus (EBV)–transformed cell lines were established. The RED analysis of DNA templates extracted from these cell lines clearly demonstrated that the $(TTAGGG)_n$ arrays were systematically expanded when compared with the $(TTAGGG)_n$ maximum length de-

Table 1

$(TTAGGG)_n$ Maximum Lengths in DNA from Blood and EBV Cell Lines

	No. of Ligated $(TTAGGG)_9$ Repeats In DNA from	
Sample	Blood	EBV-Transformed Cell Lines
Paired samples:[a]		
1	7	13[b]
2	6	8
3	7	13 [b]
4	9	11
5	5	10
6	8	12
7	7	13[b]
Unpaired samples:		
8	9	
9	6	
10	8	
11	6	
12	6	
13[b]	7	
14	4	
15	6	
16	5	
17	6	
18	6	
19	6	
20	5	
21	6	
24		13[b]
25		9
26		9
27		13[b]
28		10

[a] $(TTAGGG)_n$ maximum length in templates extracted from blood is always elongated after EBV transformation, with expansions ranging from "moderate" (pairs 2, 4, and 5) to "large" or "very large" (pairs 1, 3, 6, and 7). By a simple one-direction sign test, the probability of all seven transformed cell lines being elongated by chance is significant ($P < .01$).

[b] Arbitrary upper size value for very long but nonmeasurable ligated ladders.

tected in DNA extracted from whole blood (fig. 1 and table 1), with maximum lengths in the range of 450–700 bp (or more). Five additional DNAs from EBV-transformed cell lines of other northern European subjects also showed maximum lengths in the same range, of 450–700 bp (or more). RED analysis of control samples from three different mice strains (DNAs extracted from spleen) revealed the presence of $(TTAGGG)_n$ arrays >700 bp (not shown), in agreement with the observation of ultralong telomeres in mouse (Zijlmans et al. 1997). A very long (up to the nonresolving area of the gel) $(TTAGGG)_n$ ladder also was detected when the RED method was tested on a 5–8-kb synthetic $(TTAGGG)_n$ / $(CCCTAA)_n$ polymer generated by PCR (fig. 2). Taken together, these data rule out the possibility that the "short" $(TTAGGG)_n$ maximum lengths detected by RED in templates from whole blood are artifacts due to the inability of the ligation process to proceed beyond secondary structure (e.g., intrastrand hairpins) of $(TTAGGG)_n$ arrays.

Overall, the maximum $(TTAGGG)_n$ length in the 21 samples from blood was found to be significantly shorter (200–500 bp) than the maximum length detected in the 12 EBV cell lines (Mann-Whitney test; $P < .001$). The $(TTAGGG)_n$-array expansions are in accord with telomere lengthening in transformed cell lines (de Lange 1995; Meyerson et al. 1997; Van Steensel and de Lange 1997) and largely consistent with the reported hEST2 up-regulation/telomerase reactivation in transformed cell cultures, although we cannot exclude the possibility that they originate from DNA polymerase slippage or by chromosomal recombination following EBV transformation.

It has been stressed that "in considering telomere dynamics, it is clearly important to establish the length of the $(TTAGGG)_n$ repeat array" (de Lange 1995, p. 266) and that the instability of $(TTAGGG)_n$ might have "some mechanistic similarity to the instability of short tandem repeats that produce variable microsatellite or minisatellite loci, and the instability of trinucleotide repeats that underlie some human genetic diseases" (Kipling 1995, p. 196). It is worthwhile to hypothesize that the interruption of the unadulterated $(TTAGGG)_n$ arrays every 200–500 bp in DNA templates from whole blood may be important for the maintenance of repeat stability, whereas the loss of interruption could be associated with instability and expansion, as part of a dynamic process perhaps not dissimilar to that resulting in expansion mutation of some trinucleotide repeats (Ashley and Warren 1995; Gordenin et al. 1997). The RED method could be used to monitor telomere stability in transformed cell lines and tumors or to test the effect of genes involved in telomere maintenance or DNA repair after transfection in cell lines with abnormally short telomeres. In conclusion, the results generated by the RED method

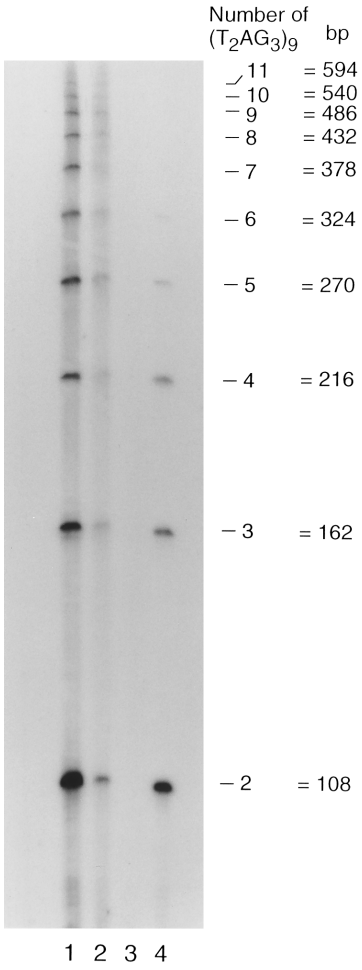

Figure 2 RED of $(TTAGGG)_n$ arrays, using 100 pg (lane 1) and 10 pg (lane 2) of a synthetic polymer generated by PCR with $(TTAGGG)_9/(CCCTAA)_5$ primers in the absence of DNA template. The PCR was performed in a 20-μl final volume, with 200 μM of each dNTP, 2 M Betaine, 5% dimethylsulfoxide, 2 μl of 10 × Klentaq buffer, and 0.2 μl of KlentaqLA-16 enzyme mix (15:1 Klentaq1:*Pfu* DNA polymerase) (Barnes 1994; Baskaran et al. 1996). Samples were taken through 40 cycles of 95°C for 1 min, 37°C for 1 min, and 72°C for 6 min, with a 3-s increase per cycle. The reaction product was run on a 1% agarose gel and was visualized, by ethidium bromide staining, as a smear with a size range of 1–20 kb. PCR products in the 5–8-kb range were gel purified and used as template in a RED reaction with $(TTAGGG)_9$ oligonucleotides. Decreasing amounts of synthetic template yield an equally long ladder but one that is of lower relative intensity (lanes 1 and 2). A control reaction in the absence of DNA template is shown in lane 3, demonstrating that the ladder does not result from self-annealing of $(TTAGGG)_9$ oligonucleotides. Lane 4 shows results for a RED replicate on a DNA template extracted from blood (sample 3B in fig. 1), demonstrating that the shorter length detected in DNA from whole blood is consistently reproduced across reactions.

offer a new perspective on $(TTAGGG)_n$ maximum length and on the relative stability of this telomeric hexamer, both in vitro and in vivo.

Acknowledgments

We thank Drs E. Fritz , S. Meyn, F. Calafell, S. Baserga, G. Isaya, and T. Ashley for helpful discussion and advice, and we thank B. King for her gift of DNA samples from mice. This work was supported in part by National Institutes of Health grants MH44876, MH39239, and MH50390.

GIORGIO SIRUGO AND KENNETH K. KIDD
Department of Genetics
Yale University School of Medicine
New Haven

Electronic-Database Information

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

Epicentre Forum, http://www.epicentre.com/f2_3/f2_3al.html

References

Allshire RC, Dempster M, Hastie ND (1989) Human telomeres contain at least three types of G-rich repeat distributed nonrandomly. Nucleic Acids Res 17:4611–4627

Ashley CJ, Warren ST (1995) Trinucleotide repeat expansion and human disease. Annu Rev Genet 29:703–704

Barnes WM (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. Proc Natl Acad Sci USA 91:2216–2220

Baskaran N, Kandpal RP, Bhargava AK, Glynn MW, Bale A, Weissman SM (1996) Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. Genome Res 6:633–638

Blasco MA, Lee HW, Hande MP, Samper E, Lansdorp PM, Depinho RA, Greider CW (1997) Telomere shortening and tumour formation by mouse cells lacking telomerase RNA. Cell 91:25–34

Brown WR, MacKinnon PJ, Villasante A, Spurr N , Buckle VJ, Dobson MJ (1990) Structure and polymorphism of human telomere-associated DNA. Cell 63:119–132

Bryan TM, Englezou A, dalla-Pozza L, Dunham MA, Reddel RR (1997) Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. Nat Med 3:1271–1274

Bryan TM, Englezou A, Gupta J, Bacchetti S, Reddel RR (1995) Telomere elongation in immortal human cells without detectable telomerase activity. EMBO J 14:4240–4248

de Lange T (1995) Telomere dynamics and genome instability in human cancer. In: Blackburn EH, Greider CW (eds) Telomeres. Cold Spring Harbor Laboratory Press, New York, NY, pp 265–293

de Lange T, Shiue L, Myers RM, Cox DR, Naylor SL, Killery AM, Varmus H (1990) Structure and variability of human chromosome ends. Mol Cell Biol 10:518–527

Gordenin DA, Kunkel TA, Resnik M (1997) Repeat expansion—all in a flap? Nat Genet 16:116–118

Hanish JP, Yanowitz JL, de Lange T (1994) Stringent sequence requirements for the formation of human telomeres. Proc Natl Acad Sci USA 91:8861–8865

Kipling D (1995) The telomere. Oxford University Press, New York

Martens UM, Zijlmans JMJM, Poon SSS, Dragowska W, Yui J, Chavez E, Ward RK, et al (1998) Short telomeres on human chromosome 17p. Nat Genet 18:76–80

Meyerson M, Counter CM, Eaton EN, Ellisen LW, Steiner P, Caddle SD, Ziaugra L, et al (1997) hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. Cell 90:785–795

Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB, et al (1997) Telomerase catalytic subunit homologs from fission yeast and human. Science 277:955–959

Notaro R, Cimmino A, Tabarini D, Rotoli B, Luzzatto L (1997) *In vivo* telomere dynamics of human hematopoietic stem cells. Proc Natl Acad Sci USA 94:13782–13785

Schalling M, Hudson TJ, Buetow KH, Housman DE (1993) Direct detection of novel expanded trinucleotide repeats in the human genome. Nat Genet 4:135–139

Sirugo G, Kidd KK (1995) Repeat expansion detection using ampligase thermostable DNA ligase. Epicentre Forum 2:1–3

Sirugo G, Deinard AS, Kidd JR, Kidd KK (1997) Survey of maximum CTG/CAG repeat lengths in humans and nonhuman primates: total genome scan in populations using the repeat expansion detection method. Hum Mol Genet 6: 403–408

Van Steensel B, de Lange T (1997) Control of telomere length by the human telomeric protein TRF1. Nature 385:740–743

Vaziri H, Schachter F, Uchida I, Wei L, Zhu X, Effros R, Cohen D, et al (1993) Loss of telomeric DNA during aging of normal and trisomy 21 human lymphocytes. Am J Hum Genet 52:661–667

Zijlmans JMJM, Martens U, Poon SSS, Raap AK, Tanke HJ, Ward RK, Lansdorp PM (1997) Telomeres in the mouse have large inter-chromosomal variations in the number of T_2AG_3 repeats. Proc Natl Acad Sci USA 94:7423–7428

Address for correspondence and reprints: Dr. G. Sirugo, Department of Genetics, Yale University School of Medicine, New Haven, CT 06520-8005. E-mail: Sirugo@biomed.med.Yale.edu

0002-9297/98/6302-0040$02.00

Am. J. Hum. Genet. 63:651–654, 1998

A retGC-1 Mutation in Autosomal Dominant Cone-Rod Dystrophy

To the Editor:

Choroidoretinal dystrophies represent a clinically and genetically heterogeneous group of disorders that in-

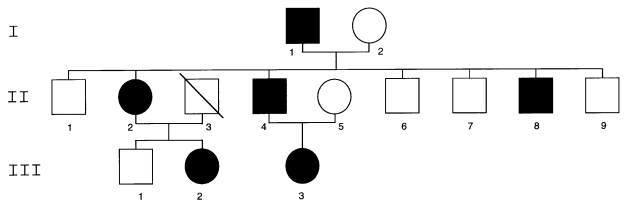

Figure 1 Pedigree of family segregating autosomal dominant cone-rod dystrophy (CORD6).

cludes retinitis pigmentosa (RP). On the other hand, cone-rod dystrophies (CRDs) long have been regarded as inverse RP and are characterized clinically by an initial cone dysfunction followed by a progressive peripheral disease (Rabb et al. 1986; Heckenlively 1987). The main symptoms at onset of the disease are a decrease of visual acuity, with loss of color discrimination and photophobia. As the disease progresses, nyctalopia, progressive peripheral visual field deficit, and decreasing scotopic electroretinogram (ERG) amplitudes are observed (Moore 1995). Autosomal dominant, autosomal recessive, and X-linked recessive patterns of inheritance have been observed (Bird 1995), and five CRD loci have been mapped: CORD1 to 18q21.1 (Warburg et al. 1991), CORD2 to 19q13 (Evans et al. 1994), CORD3 to Xp22.13-p22.11 (McGuire et al. 1995), the peripherin/retinal degeneration slow (RDS) gene to 6p21.2-cen (Travis et al. 1991), and CORD6 to 17p12-p13 (Kelsell et al. 1997). Yet, only two disease-causing genes have been identified for CRD—namely, the peripherin/RDS gene (Nakazawa et al. 1994, 1996; Kohl et al. 1997) and the photoreceptor-specific homeobox gene, CRX, corresponding to CORD2 (Freund et al. 1997).

Since CORD6 maps to the genetic interval encompassing the retinal-specific guanylate cyclase gene (retGC-1) and especially since retGC-1 mutations have been reported elsewhere for Leber congenital amaurosis (LCA1) (Perrault et al. 1996), we screened retGC-1 for mutations in a large CRD pedigree consistent with linkage to CORD6. In addition, very recently a large deletion of the GC1 gene, the avian orthologue of retGC-1, had been reported in the rd/rd chicken affected with a congenital retinal degeneration similar to LCA (Semple-Rowland et al. 1997).

All affected individuals displayed an early cone dysfunction characterized by decreased vision acuity, with severe color dyschromatopsia and photophobia, during the 1st decade of life. At this stage, ophthalmoscopical examinations were not specific. By contrast, electrophysiological testing revealed marked loss of photopic function, with scotopic function relatively well preserved, and the visual field showed a consistent central scotoma. During the 2d and 3d decades, visual acuity decreased dramatically, and the color-vision defect was confined to achromatopsia, hampering normal schooling and professional insertion. After 40 years, peripheral visual field loss and progressive night blindness were observed, and the ERG became unrecordable (individual I-1; fig. 1).

For PCR-based genotypic analyses, genomic DNA

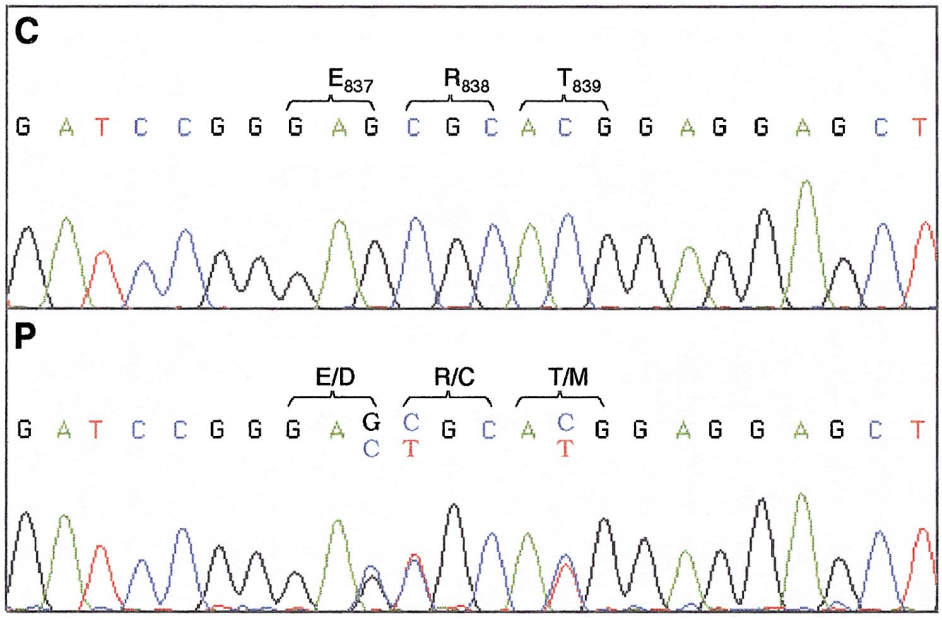

Figure 2 Identification of a heterozygote mutant genotype of the retGC-1 gene in CORD6. *C*, Control. *P*, Patient.

(200 ng) was tested as described elsewhere (Perrault et al. 1998), and linkage analyses were performed by use of the MLINK and LINKMAP options of the LINKAGE program, version 5.1 (Lathrop et al. 1985). For mutation screening of the 18 coding exons of the retGC-1 gene, genomic DNA (200 ng) was PCR amplified by use of 1 μM of the intronic primers, as described elsewhere (Perrault et al. 1996). Amplification products were loaded onto a 1% agarose low–melting-temperature gel, were purified by phenol-chloroform extraction, and were recovered by ethanol precipitation. Purified fragments were sequenced directly, by use of primers specific for the cDNA sequence and the Big Dye Terminator Cycle Sequencing kit (Perkin Elmer), on an automatic fluorometric DNA sequencer (Applied Biosystems).

Positive LOD-score values were obtained with polymorphic markers flanking retGC-1 at loci D17S1796 and D17S1881 (maximum LOD score of 2.71 at recombination fraction 0, for both markers). The coding sequence of the retGC-1 gene was screened for point mutations or minute changes, by direct-sequencing analysis of genomic DNA. The proband was heterozygous for a complex mutational event including three consecutive missense mutations in exon 13: (1) a G→C transversion at nucleotide 2584, changing a glutamate to an aspartate at codon 837 (E837D); (2) a C→T transition at nucleotide 2585, changing an arginine to a cysteine at codon 838 (R838C); and (3) a C→T transition at nucleotide 2589, changing a threonine to a methionine at codon 839 (T839M) (see fig. 2). This mutational event was found in all affected individuals and was absent in all healthy members of the family. No base change was found in the remaining exons.

retGC-1 mutations previously had been shown to account for LCA1. Interestingly, none of the 17 retGC-1 mutations identified in 20 unrelated LCA1 families involved the putative dimerization domain encoded by exons 11–13 (Laura et al. 1996). Conversely, no visual impairment was present in individuals heterozygous for the LCA1 mutations. We speculate that mutations at these codons led to the production of a mutant cyclase that interfered with normal protein dimerization, thereby limiting the production of cGMP in the retina, via a dominant negative effect of the mutant protein on the wild-type gene product.

In conclusion, it appears that the same gene—namely, retGC-1—can result in either an autosomal dominant cone-rod dystrophy or an autosomal recessive retinal degeneration (Leber disease), depending on the location of the mutation in the gene. The wide clinical spectrum of retGC-1 mutations gives additional support to the relevance of visual-transduction–cascade genes in a variety of retinal diseases.

Acknowledgments

This study was supported by the Association Retina France.

Isabelle Perrault,[1] Jean-Michel Rozet,[1] Sylvie Gerber,[1] Rosemary E. Kelsell,[2] Eric Souied,[1] Annick Cabot,[1] David M. Hunt,[2] Arnold Munnich,[1] and Josseline Kaplan[1]

[1]*Unité de Recherches sur les Handicaps Génétiques de l'Enfant, Institut National de la Santé et de la Recherche Médicale U393, Hôpital des Enfants Malades, Paris; and* [2]*Department of Molecular Genetics, Institute of Ophthalmology, University College London, London*

References

Bird AC (1995) Retinal photoreceptor dystrophies LI. Edward Jackson Memorial Lecture. Am J Ophthalmol 119:543–562

Evans K, Fryer A, Inglehearn C, Duvall-Young J, Wittaker JL, Gregory CY, Butler R, et al (1994) Genetic linkage of cone-rod retinal dystrophy to chromosome 19q and evidence for segregation distortion. Nat Genet 6:210–213

Freund CL, Gregory-Evans CY, Furukawa T, Papaioannou M, Looser J, Ploder L, Bellingham J, et al (1997) Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. Cell 91:543–553

Heckenlively JR (1987) RP cone-rod degeneration. Trans Am Ophthalmol Soc 85:438–470

Kelsell RE, Evans K, Gregory CY, Moore AT, Bird AC, Hunt DM (1997) Localisation of a gene for dominant cone-rod dystrophy (CORD6) to chromosome 17p. Hum Mol Genet 6:597–600

Kohl S, Christ-Adler M, Apfelstedt-Stylla E, Kellner U, Eckstein A, Zrenner E, Wissinger B (1997) RDS/peripherin gene mutations are frequent causes of central retinal dystrophies. J Med Genet 34:620–626

Lathrop GM, Lalouel JM, Julier C, Ott J (1985) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am J Hum Genet 37:482–498

Laura RP, Dizhoor A, Hurley JB (1996) The membrane guanylyl cyclase, retinal guanylyl cyclase-1, is activated through its intracellular domain. J Biol Chem 271:11646–11651

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

Moore AT (1992) Cone and cone-rod dystrophies. J Med Genet 29:289–290

Nakazawa M, Kikawa E, Chida Y, Shiono T, Tamai M (1996) Autosomal dominant cone-rod dystrophy associated with mutations in codon 244 (Asn244His) and codon 184 (Tyr184Ser) of the peripherin/RDS gene. Arch Ophthalmol 114:72–78

Nakazawa M, Kikawa E, Chida Y, Tamai M (1994) Asn244His mutation of peripherin/RDS gene causing autosomal dominant cone-rod degeneration. Hum Mol Genet 3:1195–1196

Perrault I, Châtelin S, Nancy V, Rozet JM, Gerber S, Ghazi I,

Souied E, et al (1998) Exclusion of five subunits of cGMP phosphodiesterase in Leber's congenital amaurosis. Hum Genet 102:322–326

Perrault I, Rozet JM, Calvas P, Gerber S, Camuzat A, Dollfus H, Châtelin S, et al (1996) Retinal-specific guanylate cyclase gene mutations in Leber's congenital amaurosis. Nat Genet 14:461–464

Rabb MF, Tso MO, Fishman GA (1986) Cone-rod dystrophy: a clinical and histopathologic report. Ophthalmology 93: 1443–1451

Semple-Rowland SL, Lee NR, Van Hooser JP, Palczewski K, Baehr W (1998) A null mutation in the photoreceptor guanylate cyclase gene causes the retinal degeneration chicken phenotype. Proc Natl Acad Sci USA 95:1271–1276

Travis GH, Christerson L, Danielson PE, Klisak I, Sparkes RS, Hahn LB, Dryja TP, et al (1991) The human retinal degeneration slow rds gene: chromosome assignment and structure of the mRNA. Genomics 10:733–739

Warburg M, Sjo O, Fledelius HC (1991) Deletion mapping of a retinal cone-rod dystrophy: assignment to 18q21.1. Am J Med Genet 39:288–293

Address for correspondence and reprints: Dr. Josseline Kaplan, Unité de Recherches sur les Handicaps Génétiques de l'Enfant, INSERM U393, Hôpital des Enfants Malades, 149 rue de Sèvres, 75743 Paris Cedex 15, France. E-mail: munnich3@citi2.fr

0002-9297/98/6302-0041$02.00

Am. J. Hum. Genet. 63:654–656, 1998

Mapping Genes by Drift-Generated Linkage Disequilibrium

To the Editor:

In human populations that have remained of small and constant size, high levels of linkage disequilibrium (LD) are generated by genetic drift (Slatkin 1994; Laan and Pääbo 1997). Theoretical considerations suggest that such LD can be used to identify chromosomal regions involved in diseases or other traits, by "drift mapping" (Terwilliger et al. 1998). This concept relies on the assumption that when "cases" and "controls" are compared within a population in which extensive LD exists, disequilibrium will be observed between the trait and marker loci close to the gene(s) that contributes to the trait. Furthermore, genetic differentiation between the cases and controls will be observed in genomic regions contributing to the trait, whereas no differentiation will be seen in other parts of the genome. Computer simulations indicate that, under reasonable assumptions with regard to population size, population age, and marker heterozygosity (Terwilliger et al. 1998), it might be possible to map genes by use of this approach.

To empirically evaluate this idea, we have studied polymorphic loci in and around the gene that encodes the renin-binding protein (RnBP), a component of the renin-angiotensin system involved in the regulation of blood pressure. The RnBP gene is located on Xq28 and contains a point mutation, T61C, that occurs with a frequency of .18 in Germans (Knöll et al. 1997). We scored this polymorphism in males from the Saami and the Finns, two populations that differ radically in their demographic history. Whereas the Saami have not expanded during historical times and show no indication of expansion in tests based on DNA sequence variability (von Haeseler et al. 1996), the Finns are thought to have expanded drastically during the past few thousand years, on the basis of both epidemiological (Peltonen et al. 1995) and genetic evidence (Sajantila et al. 1996). The frequencies of the C allele were found to be .21 and .19 in the Saami and the Finns, respectively. The fact that the C allele occurs at appreciable frequencies in three European populations indicates that it is older than these populations. It is therefore a useful model of alleles involved in complex traits, since such alleles are expected to be both frequent in the population and of old age.

Four microsatellites located ~1.0–7.8 cM from the RnBP gene (fig. 1), as well as the T61C polymorphism, were typed in 53 Saami and 80 Finns. In addition, 10 microsatellite loci on Xp22 and Xq13, which had numbers of alleles comparable to the numbers of those around the RnBP gene, were typed in the same individuals (Laan and Pääbo 1997; authors' unpublished data), to assess the extent to which loci situated far from the RnBP gene might yield spurious associations with the T61C polymorphism. When the RnBP polymorphism and the microsatellite loci were analyzed for allelic as-

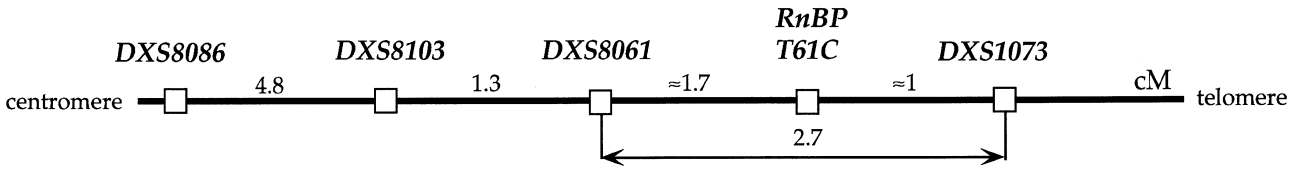

Figure 1 Genetic map (Nelson et al. 1995; Dib et al. 1996; Esposito et al. 1997; Nagaraja et al. 1997) of studied microsatellite loci around the RnBP gene.

sociation (table 1), DXS8061, located ~1.7 cM from the RnBP gene, showed LD in the Saami, at a significance level that would allow for gene mapping ($P = .00002$). All other P values, including that for DXS8061 in the Finns, were $\geqslant .003$. To evaluate the false-positive rate for the observed data, a randomization analysis was performed, in which the microsatellite haplotypes across the three regions on the X chromosome were kept together while the RnBP C/T alleles were randomly shuffled between the different haplotypes. The LD statistic was calculated for each microsatellite locus, and the most significant marker from each replicate was retained. After 10,000,000 randomizations, the significance of the association between DXS8061 remained significant at the .001 level in the Saami, whereas no association in the Finns was significant at the .05 level. It is noteworthy that, in contrast to DXS8061, DXS1073, located ~1.0 cM from the RnBP gene, showed no LD in the Saami. This result can be attributed to the fact that DXS1073 displays a single allele in the Saami carrying the C allele; it also underscores the fact that, because the extent of LD between closely linked loci is highly stochastic in constant populations (Weir 1996; Terwilliger et al. 1998), drift mapping is unlikely to allow fine-scale localization of genes.

When F_{ST} (Weir 1996) was used as a measure of genetic differentiation between the C and T chromosomes (table 1), differentiation ($P = .00000$) was seen for DXS8061 in the Saami. Furthermore, it was tested whether haplotypes consisting of four microsatellite loci would still show significant differentiation between the C and T alleles. This was the case, at a high level of significance ($P = .00000$), for the Saami but not for the Finns ($P = .02760$). Furthermore, no differentiation was detected in either the Saami or the Finns, for any of the four-locus haplotypes on Xp22 and Xq13 (data not shown).

It is noteworthy that the C allele is associated with different alleles (and haplotypes) of flanking microsatellites in the Saami and the Finns (data not shown). Thus, the ancestral haplotype on which the T61C mutation originally occurred has been lost in at least one population—and, possibly, in both. As predicted elsewhere (Terwilliger et al. 1998), this does not affect the efficiency of "drift mapping," whereas "shared segment" analyses of the combined data set would fail to show positive evidence of a gene in this region.

In conclusion, the Saami exhibit P values, both for LD between a marker locus and the T61C polymorphism and for population differentiation between the C and T chromosomes around the RnBP locus, of <.0001; this would allow ~10,000 tests to be performed without an unreasonably high background of false positives. By contrast, no comparable signals are seen in the Finns. Thus, these results show that, in principle, LD generated by drift in a small and constant population can be used to localize a gene, whereas it is difficult, if not impossible, in a population that has expanded. A further potential advantage of populations such as the Saami is that they may retain much of their statistical power in the presence of allelic heterogeneity (Terwilliger et al. 1998). Although several issues not addressed here complicate real mapping projects, we are hopeful that chromosomal regions involved in complex traits could be identified by an approach based on LD in populations such as the Saami.

Table 1

LD between T61C and Microsatellites and Genetic Differentiation between C and T Chromosomes

Region and Marker	LD (P)		F_{ST} (P)	
	Saami	Finns	Saami	Finns
Xq28:				
DXS1073	.09244	.00878	.169 (.04870)	.144 (.00540)
DXS8061	.00002	.03446	.443 (.00000)	.010 (.28567)
DXS8103	.13335	.32284	.089 (.06949)	.010 (.26387)
DXS8086	.40300	.72854	.053 (.19008)	−.022 (.78802)
Haplotype	N/A	N/A	.213 (.00000)	.034 (.02760)
Xq13:				
DXS983	.68754	.55828	−.010 (.42876)	−.032 (.94841)
DXS1225	.28491	.25807	.019 (.27327)	−.014 (.65133)
DXS8082	.13440	.43646	.039 (.17868)	.005 (.32887)
DXS8037	.26383	.18007	−.013 (.43976)	.019 (.17988)
DXS995	.05786	.21628	.205 (.03950)	.033 (.12239)
Xp22:				
DXS7105	.54625	.79768	−.009 (.42616)	−.016 (.62774)
DXS7163	.00316	.49793	.139 (.01850)	−.014 (.56234)
DXS1052	.69146	.39611	−.038 (.60054)	.015 (.22038)
DXS1229	1.00000	.02095	−.062 (.99990)	.113 (.01220)
DXS999	.82214	.00660	−.022 (.60774)	.050 (.05139)

Note.—P values were computed by use of ARLEQUIN software (Schneider et al. 1997).

Acknowledgments

We are indebted to Drs. Joseph Terwilliger and Sebastian Zöllner, for continuous discussion and statistical help; to Drs. Leena Peltonen and Lars Beckman, for samples; to the Alexander-von-Humboldt Stiftung, for a fellowship to M.L.; and to the Deutsche Forschungsgemeinschaft grant Pa 452/3-1, for financial support.

Maris Laan and Svante Pääbo
Max-Planck-Institute for Evolutionary Anthropology and Institute of Zoology
University of Munich
Munich

References

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
Esposito T, Ciccodicola A, Flagiello L, Matarazzo MR, Migliaccio C, Cifarelli RA, Visone R, et al (1997) Expressed STSs and transcription of human Xq28. Gene 187:185–191
Knöll A, Schunkert H, Reichwald K, Danser AHJ, Bauer D, Platzer M, Stein G, et al (1997) Human renin binding protein: complete genomic sequence and association of an intronic T/C polymorphism with the prorenin levels in males. Hum Mol Genet 6:1527–1534
Laan M, Pääbo S (1997) Demographic history and linkage disequilibrium in human populations. Nat Genet 17: 435–438
Nagaraja R, MacMillan S, Kere J, Jones C, Griffin S, Schmatz M, Terrell J, et al (1997) X chromosome map at 75-kb STS resolution, revealing extremes of recombination and CG content. Genome Res 7:210–222
Nelson DL, Ballabio A, Cremers F, Monaco AP, Schlessinger D (1995) Report on the 6th international workshop on X-chromosome mapping. Cytogenet Cell Genet 71:308–342
Peltonen L, Pekkarinen P, Aaltonen J (1995) Messages from an isolate: lessons from the Finnish gene pool. Biol Chem 376:697–704
Sajantila A, Salem A-H, Savolainen P, Bauer K, Gierig C, Pääbo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci USA 93:12035–12039
Schneider S, Kueffer J-M, Roessli D, Excoffier L (1997) ARLEQUIN, version 1.1. Genetics and Biometry Laboratory, University of Geneva, Geneva
Slatkin M (1994) Linkage disequilibrium in growing and stable populations. Genetics 137:331–336
Terwilliger JD, Zöllner S, Laan M, Pääbo S (1998) Mapping genes through the use of linkage disequilibrium generated by genetic drift: "drift mapping" in small populations with no demographic expansion. Hum Hered 48:138–154
von Haeseler A, Sajantila A, Pääbo S (1996) The genetical archaeology of the human genome. Nat Genet 14:135–140
Weir BS (1996) Genetic data analysis II. Sinauer Associates, Sunderland, MA

Address for correspondence and reprints: Dr. Svante Pääbo, Max-Planck-Institute for Evolutionary Anthropology and Institute of Zoology, University of Munich, Luisenstrasse 14, D-80333 Munich, Germany. E-mail: paabo@eva.mpg.de

0002-9297/98/6302-0042$02.00

Am. J. Hum. Genet. 63:656–662, 1998

Evidence for a Common Ethnic Origin of Cystic Fibrosis Mutation 3120+1G→A in Diverse Populations

To the Editor:

Cystic fibrosis (CF) is a common recessive disorder in Caucasians, but little is known about its incidence in other populations (Welsh et al. 1995). In a recent study, however, Macek et al. (1997*b*) described a subset of specific CF transmembrane-conductance regulator (CFTR) gene mutations in African American CF patients. One splicing mutation, 3120+1G→A in intron 16, was particularly frequent and accounted for approximately half the "African" CF chromosomes in the group that Macek et al. studied (Macek et al. 1997*b*). This mutation also has been identified in four native African CF patients, on 5/8 chromosomes (Carles et al. 1996). Furthermore, it has been demonstrated that 3120+1G→A is a predominant CF mutation in the Eastern Oasis population of Saudi Arabia (El-Harith et al. 1997). Finally, three Greek CF families have been reported to harbor this mutation (Tzetis et al. 1997). These observations indicate that CF mutation 3120+1G→A is present in diverse populations from different continents.

To examine whether the 3120+1G→A mutation has a common origin in all these populations or whether its widespread distribution is the result of recurrent mutational events, we analyzed DNA samples obtained from 17 unrelated CF patients in four different populations and from 8 unrelated African CF carriers (fig. 1). In the first cohort, six CF patients were of African American descent, three CF patients originated from Saudi-Arabia, three CF patients were of Greek origin, and five CF patients were native Africans (four families came from South Africa, and one family came from Cameroon). In the second cohort, eight native African individuals who had been identified as mutation carriers in a population-based screening in South Africa (C. Padoa and M. Ramsay, unpublished data) were included here as a confirmatory group. The presence of the 3120+1G→A mutation in these different ethnic groups was confirmed by direct sequencing. We have typed six intra- and six extragenic RF40LP markers that had been useful in previous studies to characterize the origins of numerous other CFTR mutations (Estivill et al. 1987; Dörk et al. 1992, 1994; Ramsay et al. 1993; Sereth et al. 1993; Cuppens et al. 1994; Morral et al. 1996). In addition, we investigated the three highly informative intragenic CFTR microsatellites that are located in intron 8 (IVS8CA) and intron 17b (IVS17bTA and IVS17bCA) of the CFTR gene (Zielenski et al. 1991; Morral and Estivill 1992; Morral et al. 1993).

A common extended 3120+1G→A–associated haplotype could be derived in each of the four study populations (table 1). The phasing of haplotypes was based either on homozygosity or on the analysis of parental samples in all African and Arab CF families, as well as in two African American and two Greek CF families. In the remaining single CF patients, other haplotypes for the 3120+1G→A allele than those deduced in table 1 would be formally possible. Three of the four single African American patients, however, were compound

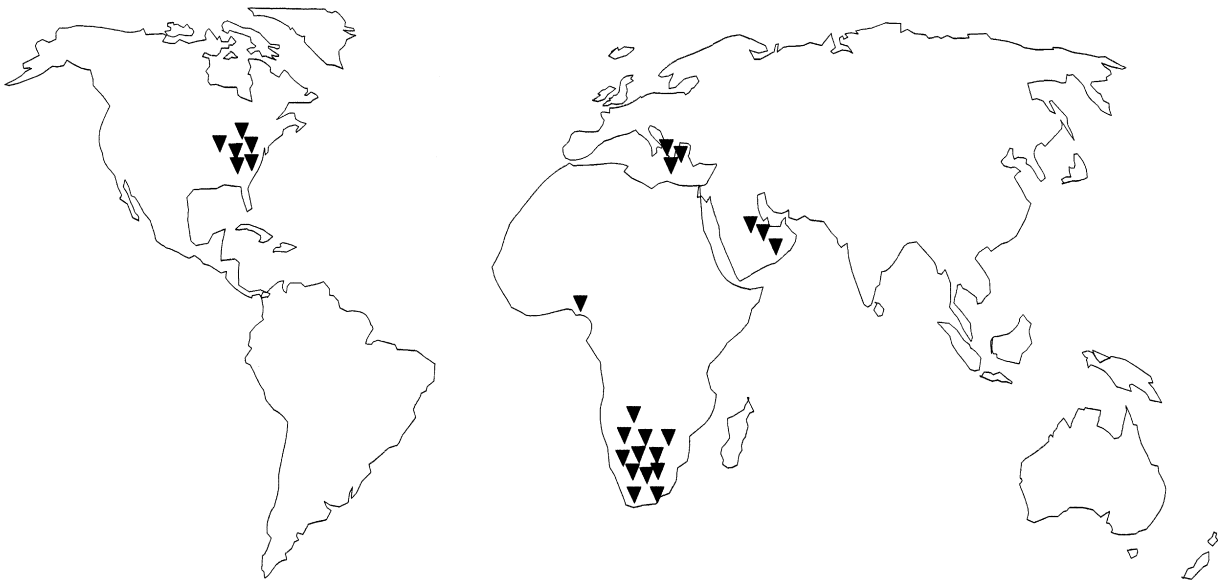

Figure 1 Geographic origins of 3120+1G→A–carrying individuals whose DNA samples were contributed to this study

heterozygotes for 3120+1G→A and for the major CF mutation, ΔF508. The extensively studied ΔF508 mutation has been shown to have a single origin in several investigated populations, with a common dimorphic marker haplotype (Kerem et al. 1989; Dörk et al. 1992; Claustres et al. 1996; Morral et al. 1996) and three major intragenic microsatellite haplotypes—23-31-13, 17-32-13, and 17-31-13—accounting for >85% of ΔF508 chromosomes (Zielenski et al. 1991; Morral et al. 1993, 1994; Claustres et al. 1996; Hughes et al. 1996). Under the assumption that the ΔF508 mutation has occurred only once and has been introduced into the African American population by ethnic admixture, we were able to use the known major ΔF508 haplotypes to deduce the most likely haplotypes for the 3120+1G→A allele in the three additional single African American patients. Within these limitations, all results obtained with the intragenic CFTR markers were consistent with an identical intragenic haplotype for all investigated 3120+1G→A alleles—with the exception of Greek haplotypes, which differed at a single microsatellite locus (IVS8CA) by one repeat unit (16 vs. 17 CA repeats). These two related intragenic haplotypes together account for <15% of non-CF alleles in the general Caucasian population (Morral et al. 1993, 1994, 1996; Russo et al. 1995; Claustres et al. 1996; Hughes et al. 1996), and we have observed these intragenic haplotypes only once in a preliminary study of 12 Arab and 10 African non-CF alleles. Thus, our analysis of intragenic markers indicates that the 3120+1G→A mutation in the four study populations most likely derives from a common ancestor. On the 5′ side, the shared haplotype extends beyond the CFTR gene in all populations, up to a point where a difference is again observed between the Greek and the Arab/African CF alleles—namely, distal to the marker CS.7, which is located >220 kb upstream of the CFTR gene. In previous studies, this region between KM.19 and XV-2c had been found to be prone to recombinations (Estivill et al. 1987; Dörk et al. 1992). On the 3′ side, at the extragenic locus pJ3.ll, which is located some 660 kb downstream of the CFTR gene, haplotypic variability is even seen within the Arab and native African patient groups. The 3120+1G→A chromosomes of the African American patients, however, carried the pJ3.11 *Msp*I allele 1 in every informative case.

The observed identity of extended CFTR haplotypes for the 3120+1G→A alleles in the Arab, African, and African American patients strongly suggests that this mutation has a common origin in these groups. This finding is not surprising in the case of Africans and African Americans, since the latter group has originated mostly from the western African coast and came to North America between the 16th and 19th centuries, which is too recent to allow origination of significant CFTR-mutation haplotype changes restricted to African Americans. It is not quite so simple to explain the presence of the 3120+1G→A mutation in African and Saudi Arab patients. Although recent ethnic admixture accounts for a few percent of Africans in Saudi Arabia, this is very unlikely to explain our findings, since none of the Saudi families had any anthropomorphological signs of an African descent. However, a continuous gene flow between Arab and African populations probably

Table 1

Intra- and Extragenic CFTR Marker Haplotypes of the 3120+1G→A Mutation in Diverse Populations

Group and Mutations	Haplotype[a]														
	MetH (*Msp*I)	XV2c (*Taq*I)	CS.7 (*Hha*I)	KM.19 (*Pst*I)	J44 (*Xba*I)	IVS8CA	TUB9 (*Mnl*I)	M470 (*Hph*I)	T854 (*Ava*II)	TUB15 (*Nsi*I)[b]	IVS17bTA	IVS17bCA	TUB18 (*Hin*fI)	Q1463 (*Tsp*509I)	J3.11 (*Msp*I)
CF families:															
African American:															
Bal236:															
3120+1G→A	1	1	2	2	1	17	(2)	1	2	(2)	7	17	(2)	2	1
ΔF508	1	1	2	2	1	17	(1)	1	1	(1)	31	13	(1)	1	1
Bal719:															
3120+1G→A	1	1	2	2	1	17	2	1	(2)	2	7	17	2	2	1
ΔF508	1	1	2	2	1	17	1	1	(1)	1	32	13	1	1	1
Bal962:															
3120+1G→A	1	(1)	(2)	(2)	(1)	17	2	1	2	(2)	(7)	(17)	(2)	(2)	…
405+3A→C	1	(2)	(1)	(1)	(2)	16	2	1	2	(1)	(31)	(13)	(1)	(1)	…
Bal963:															
3120+1G→A	(1)	1	2	2	1	(17)	(2)	1	(2)	(2)	(7)	(17)	(2)	(2)	…
ΔF508	(2)	1	2	2	1	(23)	(1)	1	(1)	(1)	(31)	(13)	(1)	(1)	…
Bal964:															
3120+1G→A	1	1	2	2	1	17	(2)	1	(2)	(2)	(7)	(17)	(2)	(2)	1
ΔF508	1	1	2	2	1	17	(1)	1	(1)	(1)	(31)	(13)	(1)	(1)	1
Bal965:															
3120+1G→A	1	1	2	2	1	(17)	(2)	1	(2)	(2)	(7)	(17)	(2)	(2)	…
ΔF508	1	1	2	2	1	(23)	(1)	1	(1)	(1)	(31)	(13)	(1)	(1)	…
Saudi Arabian:															
CF10:															
3120+1G→A	1	1	2	2	1	17	2	1	2	2	7	17	2	2	2
3120+1G→A	1	1	2	2	1	17	2	1	2	2	7	17	2	2	2
CF16:															
3120+1G→A	1	1	2	2	1	17	2	1	2	2	7	17	2	2	1
3120+1G→A	1	1	2	2	1	17	2	1	2	2	7	17	2	2	1
CF46:															
3120+1G→A	1	1	2	2	1	17	2	1	2	2	7	17	2	2	1
3120+1G→A	1	1	2	2	1	17	2	1	2	2	7	17	2	2	1

Greek:															
CF17:															
3120+1G→A	2	2	2	2	1	16	2	1	2	2	7	17	2	2	1
1497delGG	1	1	2	2	1	17	2	1	2	2	7	17	2	2	1
CF541:															
3120+1G→A	2	2	(2)	(2)	(1)	16	2	(1)	(2)	(2)	7	(17)	(2)	(2)	1
711+3A→G	1	2	(1)	(1)	(2)	16	2	(2)	(1)	(1)	33	(13)	(1)	(1)	2
CF294:															
3120+1G→A	…	…	(2)	(2)	(1)	16	2	(1)	(2)	(2)	7	(17)	(2)	(2)	…
296+1G→C	…	…	(1)	(1)	(2)	16	2	(2)	(1)	(1)	31	(13)	(1)	(1)	…
Native African:															
IM:															
3120+1G→A	1	1	2	2	1	17	2	1	2	2	7	17	2	2	1
3120+1G→A	1	1	2	2	1	17	2	1	2	2	7	17	2	2	1
CL:															
3120+1G→A	1	1	2	2	1	17	2	1	2	2	7	17	2	2	1
G1249E	1	1	2	2	1	16	2	1	1	1	22	16	1	1	2
MC1:															
3120+1G→A	…	1	…	…	1	17	…	1	(2)	…	7	17	2	…	…
3196del54	…	1	…	…	1	16	…	1	(1)	…	7	17	2	…	…
MC2:															
3120+1G→A	…	1	…	2	1	17	…	1	(2)	…	7	17	(2)	…	…
2183delAA	…	2	…	2	2	16	…	1	(1)	…	21	16	(1)	…	…
M1115:															
3120+1G→A	1	1	…	2	1	17	…	1	2	2	7	17	2	…	…
ΔF508	1	1	…	2	1	23	…	1	1	1	34	14	1	…	…
African carriers:															
SABL1:															
3120+1G→A	1	…	(2)	(2)	(1)	(17)	2	1	2	(2)	(7)	(17)	(2)	(2)	…
Non-CF	1	…	(1)	(1)	(2)	(16)	2	1	2	(1)	(33)	(13)	(1)	(1)	…
SABL2:															
3120+1G→A	1	…	2	2	1	(17)	2	1	(2)	(2)	(7)	(17)	(2)	(2)	…
Non-CF	1	…	2	2	1	(16)	2	1	(1)	(1)	(34)	(13)	(1)	(1)	…

(*continued*)

Table 1 (continued)

Group and Mutations	Haplotype[a]														
	MetH (*Msp*I)	XV2c (*Taq*I)	CS.7 (*Hha*I)	KM.19 (*Pst*I)	J44 (*Xba*I)	IVS8CA	TUB9 (*Mnl*I)	M470 (*Hph*I)	T854 (*Ava*II)	TUB15 (*Nsi*I)[b]	IVS17bTA	IVS17bCA	TUB18 (*Hin*fI)	Q1463 (*Tsp*509I)	J3.11 (*Msp*I)
SABL3:															
3120+1G→A	1	…	2	2	1	(17)	(2)	1	(2)	(2)	(7)	(17)	(2)	2	…
Non-CF	1	…	2	2	1	(23)	(1)	1	(1)	(1)	(19)	(22)	(1)	2	…
SABL4:															
3120+1G→A	1	1	2	2	1	(17)	(2)	1	2	(2)	(7)	(17)	(2)	2	2
Non-CF	1	1	2	2	1	(19)	(1)	1	2	(1)	(20)	(16)	(1)	2	2
SABL5:															
3120+1G→A	1	…	2	2	1	(17)	(2)	1	(2)	(2)	(7)	(17)	(2)	(2)	1
Non-CF	1	…	2	2	1	(19)	(1)	1	(1)	(1)	(35)	(13)	(1)	(1)	1
SABL6:															
3120+1G→A	1	…	2	2	1	(17)	2	1	(2)	(2)	(7)	(17)	(2)	(2)	…
Non-CF	1	…	2	2	1	(16)	2	1	(1)	(1)	(30)	(13)	(1)	(1)	…
SABL7:															
3120+1G→A	1	…	2	2	(1)	17	(2)	1	(2)	(2)	(7)	(17)	(2)	2	…
Non-CF	1	…	2	2	(2)	17	(1)	1	(1)	(1)	(19)	(19)	(1)	2	…
SABL8:															
3120+1G→A	1	1	2	2	(1)	17	2	1	2	(2)	7	(17)	(2)	(2)	…
Non-CF	1	1	2	2	(2)	17	2	1	2	(1)	7	(18)	(1)	(1)	.....

Note.—Microsatellite analysis was perf40ormed as described (Morral et al. 1993), except that fluorescein-labeled forward primers were used and the products were analyzed on an ALF sequencer (Pharmacia).

[a] Dimorphic markers were typed as described elsewhere (Estivill et al. 1987; Williams et al. 1988; Dörk et al. 1992; Morral et al. 1996). Markers located within the CFTR gene are underlined. Combined haplotypes were constructed from the analysis of homozygous patients (3 Arabs and 1 African) and from the analysis of parental samples (2 African Americans, 2 Greeks, and 4 Africans). The common haplotypes of the 3120+1G→A alleles are within boxes. In those cases in which the phase could not be established on the basis of family analysis, numbers shown in parentheses indicate the most likely haplotypes that have been inferred on the basis of the known haplotypes of the ΔF508 mutation (Dörk et al. 1992; Morral et al. 1994) and with consideration of the strong linkage disequilibrium between single markers (Dörk et al. 1992; Cuppens et al. 1994; Morral et al. 1996).

[b] Nucleotide substitution 3041-92G/A in intron 15, located 173 bp upstream of mutation 3120+1G→A and amplifiable together with the mutation in the same PCR product.

has persisted for many centuries, in association with trading and with the spread of the Islamic religion. Thus far, the Greeks are the only Caucasian population in which the 3120+1G→A mutation has been identified. A recurrent mutational event seems to be unlikely, because the Greek haplotype differs from the others in only two minor respects: there is a difference of one dinucleotide unit at the intragenic IVS8CA repeat, a difference that could result from a single slippage mutation; and the Greek alleles carry a different extragenic Met-XV2c haplotype that probably is due to a single recombination event. Similar events at these two marker loci also account for much of the haplotypic variability associated with the ΔF508 mutation, which has been shown to have a single origin (Morral et al. 1994). Greek and Arab/African haplotypes of the 3120+1G→A mutation thus may have diverged from a common ancestor and then evolved separately in the respective populations. In this context, it is interesting to note that there are other rare mutations shared by Saudi Arabs and Greeks, such as a polyadenylation-signal mutation in the α-globin gene of Saudi and Greek thalassemia patients (Traeger-Synodinos et al. 1993). Historical contacts—for example, under Alexander the Great or during the ancient Minoan civilization—may provide an explanation for the common ancestry of disease mutations in these ethnically diverse populations.

Current theories of a heterozygote advantage for CF carriers of frequent CFTR mutations include increased survival from diarrheal diseases, genetic drift, and hitchhiking (Romeo et al. 1989; Sereth et al. 1993; Gabriel et al. 1994; Macek et al. 1997*a*). The presence of a common ancient CF mutation in African, Saudi Arab, and Greek populations suggests that this mutation too may have been selected. This study demonstrates that the 3120+1G→A mutation shares the same extragenic CS.7-KM.19 "risk" haplotype with the other frequent and ancient CF mutations—ΔF508, N1303K, and G542X (Dörk et al. 1992; Morral et al. 1993)—but that it differs from these latter mutations with respect to intragenic CFTR markers. The extragenic CS.7-KM.19 "risk" haplotype recently has been associated with a selective advantage to the postnatal survival of female carriers without a family history of CF (Macek et al. 1997*a*). In summary, our present analysis provides the first evidence for a common origin of CF among African, Arab, Greek, and African American populations. The shared extra- and intragenic 3120+1G→A–associated haplotype is most easily explained by the assumption of a single origin for this mutation. 3120+1G→A appears to be an ancient mutation that may be more common than previously thought, in populations of the tropical and subtropical belt, where CF probably is an underdiagnosed disorder.

Acknowledgments

Part of this work was supported by an Alexander von Humboldt Foundation grant (to E.-H.A.E.-H.); by IGA MZ CR grants 2899-5, 3526-3, and 4124-3, GA CR grant 301/66/1606; and Barrande grant 970157 (all to M.M.).

Thilo Dörk,[1] El-Harith A. El-Harith,[1,2] Manfred Stuhrmann,[2] Milan Macek, Jr.,[3] Michelle Egan,[4] Garry R. Cutting,[4] Maria Tzetis,[5] Emmanuel Kanavakis,[5] Soukeyna Carles,[6] Mireille Claustres,[6] Carolyn Padoa,[7] Michèle Ramsay,[7] and Jörg Schmidtke[1]

[1]*Institute of Human Genetics, Medical School Hannover, Hannover;* [2]*College of Health Sciences, Dammam, Saudi Arabia;* [3]*Department of Medical Genetics II, University Hospital Prague-Motol, Prague;* [4]*Center for Medical Genetics, Johns Hopkins Hospital, Baltimore;* [5]*First Department of Pediatrics, Athens University, Athens; Laboratoire de Biochimie Genetique, Institut de Biologie, Centre Hospitalier Universitaire, Montpellier, France; and* [7]*Department of Human Genetics, South African Institute for Medical Research and School of Pathology, University of the Witwatersrand, Johannesburg*

References

Carles S, Desgeorges M, Goldman A, Thiart R, Guittard C, Kitazos CA, de Ravel TJL, et al (1996) First report of CFTR mutations in black cystic fibrosis patients of southern African origin. J Med Genet 33:802–804

Claustres M, Desgeorges M, Moine P, Morral N, Estivill X (1996) *CFTR* haplotypic variability for normal and mutant genes in cystic fibrosis families from southern France. Hum Genet 98:336–344

Cuppens H, Teng H, Raeymaekers P, DeBoeck C, Cassiman JJ (1994) CFTR haplotype backgrounds on normal and mutant CFTR genes. Hum Mol Genet 3:607–614

Dörk T, Mekus F, Schmidt K, Boßhammer J, Fislage R, Heuer T, Dziadek V, et al (1994) Detection of more than 50 different CFTR mutations in a large group of cystic fibrosis patients. Hum Genet 94:533–542

Dörk T, Neumann T, Wulbrand U, Wulf B, Kalin N, Maass G, Krawczak M, et al (1992) Intra- and extragenic marker haplotypes of CFTR mutations in cystic fibrosis. Hum Genet 88:417–425

El-Harith EA, Dörk T, Stuhrmann M, Abu-Srair H, Al-Shahri A, Keller K-M, Lentze MJ, et al (1997) Novel and characteristic CFTR mutations in Saudi Arab children with severe presentation of cystic fibrosis. J Med Genet 34:996–999

Estivill X, Scambler PJ, Wainwright BJ, Hawley K, Frederick P, Schwartz M, Baiget M, et al (1987) Patterns of polymorphism and linkage disequilibrium for cystic fibrosis. Genomics 1:257–263

Gabriel SE, Brigman KN, Koller BH, Boucher RC, Stutts MJ (1994) Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. Science 266: 107–109

Hughes D, Wallace A, Taylor J, Tassabehji M, McMahon R, Hill A, Nevin N, et al (1996) Fluorescent multiplex micro-

satellites used to define haplotypes associated with 75 CFTR mutations from the UK on 437 CF chromosomes. Hum Mutat 8:229–235

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, et al (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245: 1073–1080

Macek M Jr, Macek M Sr, Krebsová A, Nash E, Hamosh A, Reis A, Varon-Mateeva R, et al (1997*a*) Possible association of the allele status of the CS.7/*Hha*I polymorphism 5′ of the *CFTR* gene with postnatal female survival. Hum Genet 99: 565–572

Macek M Jr, Mackova A, Hamosh A, Hilman BC, Selden RF40, Lucotte G, Friedman KJ, et al (1997*b*) Identification of common cystic fibrosis mutations in African-Americans with cystic fibrosis increases the detection rate to 75%. Am J Hum Genet 60:1122–1127

Morral N, Bertranpetit J, Estivill X, Nunes V, Casals T, Giménez J, Reis A, et al (1994) The origin of the major cystic fibrosis mutation (delta F508) in European populations. Nat Genet 7:169–175

Morral N, Dörk T, Llevadot R, Dziadek V, Mercier B, Ferec C, Costes B, et al (1996) Haplotype analysis of 94 cystic fibrosis mutations with seven polymorphic CFTR DNA markers. Hum Mutat 8:149–159

Morral N, Estivill X (1992) Multiplex PCR amplification of three microsatellites within the CFTR gene. Genomics 13: 1362–1364

Morral N, Nunes V, Casals T, Chillón M, Giménez J, Bertranpetit J, Estivill X (1993) Microsatellite haplotypes for cystic fibrosis: mutation frameworks and evolutionary tracers. Hum Mol Genet 2:1015–1022

Ramsay M, Williamson R, Estivill X, Wainwright BJ, Ho M, Halford S, Kere J, et al (1993) Haplotype analysis to determine the position of a mutation among closely linked DNA markers. Hum Mol Genet 2:1007–1014

Romeo G, Devoto M, Galietta LJV (1989) Why is the cystic fibrosis gene so frequent? Hum Genet 84:1–5

Russo MP, Romeo G, Devoto M, Barbujani G, Cabrini G, Giunta A, D'Alcamo E, et al (1995) Analysis of linkage disequilibrium between different cystic fibrosis mutations and three intragenic microsatellites in the Italian population. Hum Mutat 5:23–27

Sereth H, Shoshani T, Bashan N, Kerem B (1993) Extended haplotype analysis of cystic fibrosis mutations and its implications for the selective advantage hypothesis. Hum Genet 92:289–295

Traeger-Synodinos J, Kanavakis E, Tzetis M, Kattamis A, Kattamis C (1993) Characterization of nondeletion α-thalassemia mutations in the Greek population. Am J Hematol 44:162–167

Tzetis M, Kanavakis E, Antoniadi T, Adam G, Kattamis C (1997) Characterization of more than 85% of cystic fibrosis alleles in the Greek population, including five novel mutations. Hum Genet 99:121–125

Welsh MJ, Tsui LC, Boat TF, Beaudet A (1995) Cystic fibrosis. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic and molecular bases of inherited disease, 7th ed. McGraw-Hill, New York, pp 3799–3876

Williams C, Williamson R, Coutelle C, Loeffler F, Smith J, Ivinson A (1988) Same-day, first-trimester antenatal diagnosis for cystic fibrosis by gene amplification. Lancet 2: 102–103

Zielenski J, Markiewicz D, Rininsland F, Rommens J, Tsui L-C (1991) A cluster of highly polymorphic dinucleotide repeats in intron 17b of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Am J Hum Genet 49: 1256–1262

Address for correspondence and reprints: Dr. Thilo Dörk, Institute of Human Genetics, Medical School Hannover, D-30625 Hannover, Germany. E-mail: doerk.thilo@mh-hannover.de

0002-9297/98/6302-0043$02.00

Am. J. Hum. Genet. 63:662–663, 1998

Media Portrayals of Genetics

To the Editor:

The article by Condit et al. (1998) demonstrates some of the limitations of quantitative analysis. The authors select from *Reader's Guide* articles listed under "heredity" in various time periods. Not surprisingly, such articles consistently attribute characteristics to genes. When the 50 articles selected from the eugenic period attribute human characteristics to heredity at almost the same rate as those selected from the 1990s, the authors conclude that nothing has changed. Predictably, they find that the "degree of determinism" (which they calculate to the fifth decimal) is consistent over 90 years of profound scientific and social change.

The paper is an example of the problem of trying to quantitate what is most compellingly understood in qualitative terms. Our study of the gene in popular culture (Nelkin and Lindee 1995), a target of Condit et al.'s paper, was not a quantitative study for the precise reason that the counting of such ambiguous and heterogeneous materials provides little insight into the public meaning of science. We focused on qualitative changes in a much broader literature, to suggest that the gene has acquired new powers as a guide to social policy. In the 1990s, the cultural meanings attached to the gene are shaping employment practices, educational policies, and decisions in the courts. The serious issues raised by the high-profile gene deserve more serious analysis.

DOROTHY NELKIN AND M. SUSAN LINDEE
Department of Sociology
New York University
New York

References

Condit CM, Ofulue N, Sheedy KM (1998) Determinism and mass-media portrayals of genetics. Am J Hum Genet 62: 979–984

Nelkin D, Lindee SM (1995) The DNA mystique: the gene as a cultural icon. WH Freeman, New York

Address for correspondence and reprints: Dr. Dorothy Nelkin, Department of Sociology, New York University, 269 Mercer Street, Room 404, New York, NY 10003. E-mail: Dorothy.Nelkin@nyu.edu

0002-9297/98/6301-0044$02.00

Am. J. Hum. Genet. 63:663, 1998

Reply to Nelkin and Lindee

To the Editor:

The letter by Nelkin and Lindee, like their book (Nelkin and Lindee 1995), aptly demonstrates that qualitative methods can be at least as reductionistic as quantitative methods. Their reduction of our multiple historically sensitive index headings to the single heading of "heredity" is a misleading oversimplification. Furthermore, their claim that our article concluded that "nothing has changed" is false. Our study did show that contemporary public discourse about heredity, based as that discourse is in the accounts provided by molecular genetics and medical genetics, is not significantly more deterministic than were earlier accounts of human heredity. That, however, is not equivalent to a statement that there has been no change. In fact, our study demonstrates that contemporary presentations of genetics are more likely to assign different levels of genetic influences to different conditions. Contemporary accounts are also less likely to attribute genetic causation to simplistic behavioral characteristics. Moreover, our study demonstrates that in all periods the majority of popular representations do not attribute human characteristics solely to genetics but, rather, explicitly recognize that genes are only one factor in human outcomes.

Both quantitative and qualitative methods have useful contributions to make toward an understanding of the social implications of genetic science. To draw conclusions about the relative proportions of various types of discursive elements appearing in various venues requires that one make a quantitative assessment, no matter how informally. Formalizing one's quantitative method by employing multiple coders and randomized article selection is useful for checking the researcher's preconceptions by providing counterforces to the well-known tendencies toward selective perception of discourse. Certainly the quantitative findings of our study helped to modify our own preconceptions and to produce a more detailed, complex, and accurate qualitative account of the public discourse about biological heredity.

The qualitative portion of our study also indicates that reductionistic claims about increased determinism, of the sort made by Nelkin and Lindee (1995), fail to capture the complexities of the changes in public discussions about human heredity. Public accounts of the biological mechanisms of inheritance have shifted across the four eras in this century, from explanations centered on "germplasm" to "genes" to "DNA" to the "genome." Accompanying these shifts have been changes in models of the relationship between genetic material and various environmental inputs. These models have posited increasingly fluid relationships between genetics and other forces across time, beginning with a model of the gene as boundary setter, moving to a model of DNA as a starting point, and, most recently, featuring models of genome and environment as coactive contributors to a normatively judged outcome. Space (not methodological choice) does not allow a full elaboration of these models and their complex relationships to other parts of the public discourse. Because of the enormous delay times in academic book publishing, we will be happy to make available, to anyone who requests it and pays postage and photocopying costs, the manuscript describing these features.

Nelkin and Lindee are correct that the new scientific information about genetics and the accompanying technological capabilities raise serious social questions, and their role in raising those questions has been valuable. However, these questions are best answered by approaches employing multiple methodologies and multiple perspectives.

Celeste M. Condit
Department of Speech Communication
University of Georgia
Athens

Reference

Nelkin D, Lindee S (1995) The DNA mystique: the gene as a cultural icon. WH Freeman, New York

Address for correspondence and reprints: Dr. Dorothy Nelkin, Department of Sociology, New York University, 269 Mercer Street, Room 404, New York, NY 10003. E-mail: Dorothy.Nelkin@nyu.edu

0002-9297/98/6302-0045$02.00

Am. J. Hum. Genet. 63:664–667, 1998

Maximum-Likelihood Expression of the Transmission/Disequilibrium Test and Power Considerations

To the Editor:

The classic transmission/disequilibrium test (TDT) proposed by Spielman et al. (1993) for analysis of family-based case-control studies is a matched χ^2 test referred to as "McNemar's test." However, the same data also could be analyzed by a likelihood model, as suggested by Terwilliger (1995). In the present letter, we study the maximum likelihood (ML) statistic derived from this model and show that, when a common asymptotic threshold is used, the ML statistic is expected to be slightly more powerful than the classic McNemar χ^2. We also investigate the influence of linkage disequilibrium and allelic frequencies on the power of the ML-TDT, with comparison with the results obtained by means of the classic TDT in a recent study (Risch and Merikangas 1996) .

Using the notation in Spielman et al.'s (1993) table 2, we consider a marker locus with codominant alleles M_1 and M_2, where b (c) is the number of alleles M_1 (M_2) transmitted from a heterozygous M_1M_2 parent to an affected child. The classic TDT test is $\chi^2_{TD} = (b - c)^2/n$, where $n = b + c$ is the number of informative heterozygous parents. Let π be the probability that allele M_1 is transmitted from an M_1M_2 parent to an affected child; then $L(\pi)$—the likelihood for the same data—can be written as $L(\pi) = \pi^b(1 - \pi)^c$. The ML estimator of π, denoted as "p," is equal to b/n, and, in the ML-TDT, the test of the null hypothesis $\pi = .5$ is performed by use of a standard likelihood-ratio test denoted as "Λ":

$$\Lambda = 2Ln[L(p)/L(.5)]$$
$$= 2n[pLn(p) + (1 - p)Ln(1 - p) - Ln(.5)] .$$

When the alternative hypothesis, H_1, is $\pi \neq .5$, $\Lambda_{\pi \neq .5}$ is asymptotically distributed as a χ^2 with 1 df, and this procedure allows us to assess the effect of both alleles in a single two-sided test; when H_1 is $\pi > .5$, p is bounded at .5 when $b < n/2$, and $\Lambda_{\pi \neq .5}$ is asymptotically distributed as a 50:50 mixture of χ^2(0 df) and χ^2(1 df). In this latter procedure, the effect of each allele should be assessed separately, leading to two one-sided tests.

The two statistics, $\Lambda_{\pi \neq .5}$ and χ^2_{TD}, are strictly monotonic, increasing with an increasing departure of p from .5, and are perfectly correlated in rank, and they are therefore equivalent in the sense discussed by Knapp et al. (1994). In particular, it is possible to find a critical threshold, denoted as c_{ML}, for $\Lambda_{\pi \neq .5}$, and another one, c_{TD}, for χ^2_{TD}, such that the tests derived from the two statistics have identical sizes (and consequently equal power). However, this equivalence does not necessarily imply the equality of the power of the tests when the *same* critical threshold is used for the two statistics (i.e., when $c_{ML} = c_{TD}$). This corresponds to the common situation in which an asymptotic threshold is considered; for example, for a .05 type I error, the critical value of 3.84 is used for both $\Lambda_{\pi \neq .5}$ and χ^2_{TD}. In this case, to determine the statistic producing the highest value for a given p, we are interested in the difference between $\Lambda_{\pi \neq .5}$ and χ^2_{TD}, denoted as "$d(p)$," which is equal to

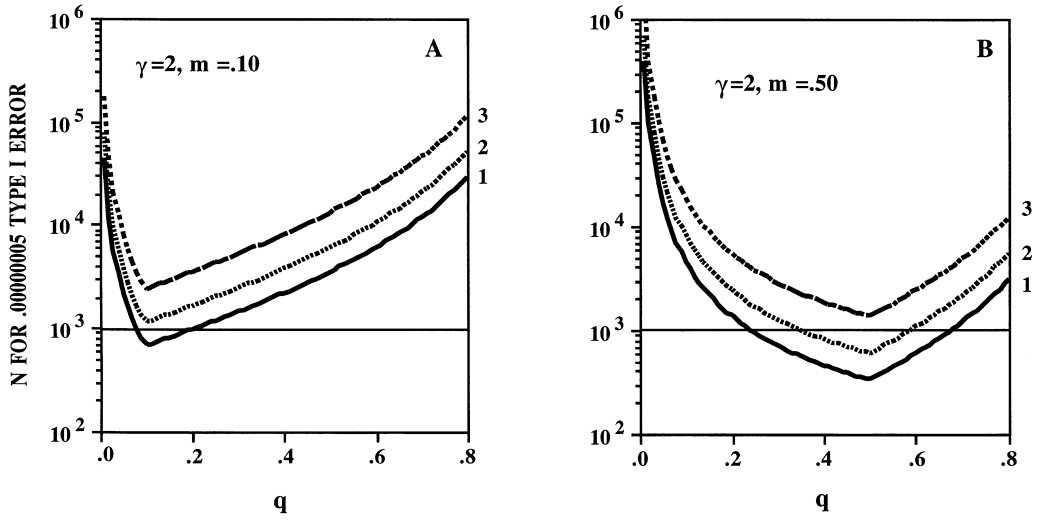

Figure 1 *N* (log scale) required for detection of association, for a type I error of 5×10^{-8} (genomewide-screening strategy) and a power of 80%, according to q, for $\gamma = 2$. The curves are drawn for $m = .10$ (*A*) and $m = .50$ (*B*), under the assumption that linkage disequilibrium is equal to its maximum, δ_{max} (curve 1), $.75\delta_{max}$ (curve 2), and $.5\delta_{max}$ (curve 3).

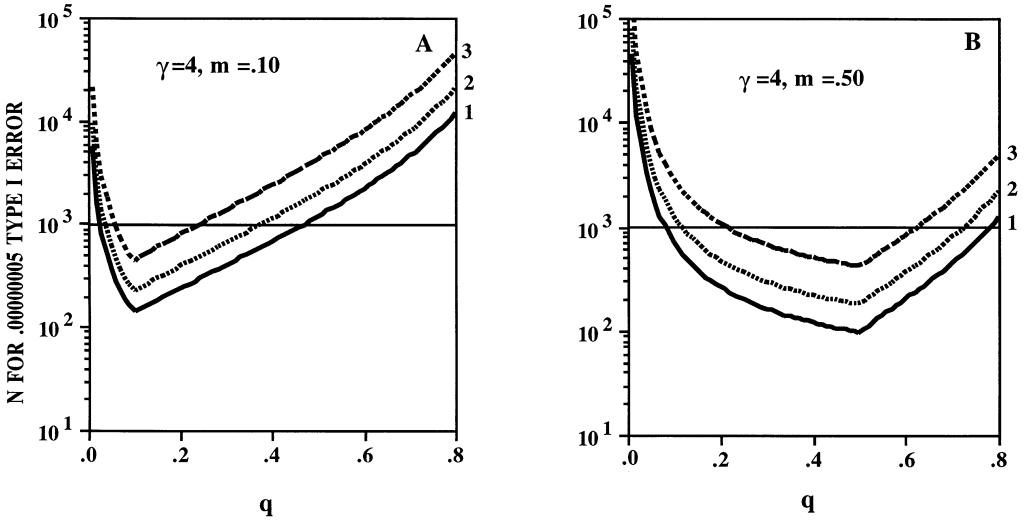

Figure 2 *N* (log scale) required for detection of association, for a type I error of 5×10^{-8} (genomewide-screening strategy) and a power of 80%, according to q, for $\gamma = 4$. The curves are as described in figure 1.

$$\mathrm{d}(p) = 2n[p\mathrm{L}n(p) + (1-p) \times \mathrm{L}n(1-p) - \mathrm{L}n(.5) - 2(p-.5)^2] \ .$$

The second derivative of $d(p)$, which is equal to $2n\{[p(1-p)]^{-1} - 4\}$, is a positive function for $p \neq .5$, since, for $p \neq .5$, $p(1-p)$ is always $<.25$. Consequently, its primitive, which is the first derivative of $d(p)$ and is equal to $2n[\mathrm{L}n(p) - \mathrm{L}n(1-p) - 4p + 2]$, is an increasing function as p departs from .5 and is also positive, since this first derivative is equal to 0 for $p = .5$. By the same reasoning, $d(p)$ is found to be a positive, increasing function as p departs from .5. This result demonstrates that, for $p \neq .5$, $\Lambda_{\pi \neq .5}$ is always $> \chi^2_{\mathrm{TD}}$. As an example, we consider the data for insulin-dependent diabetes mellitus and the insulin-gene region, presented in Spielman et al.'s (1993) table 5, where $b = 78$ and $c = 46$. The analysis by classic TDT provided a χ^2_{TD} of 8.26, whereas $\Lambda_{\pi \neq .5}$ is equal to 8.35. This is not, of course, a large difference, but it makes the point that, when asymptotic thresholds are used, the ML-TDT is expected to be slightly more powerful that the McNemar test.

We also compare the numbers of families that the ML-TDT requires for detection of an association in different situations versus those given in a recent paper by Risch and Merikangas (1996), who used the classic McNemar test. We use the same genetic model as was used by Risch and Merikangas (1996), which includes (1) a disease locus with two alleles, A and a, with population frequencies q and $1 - q$, and a multiplicative model with genotypic relative risk γ and γ^2 for Aa and AA subjects, respectively; and (2) a closely linked diallelic marker (recombination fraction 0) with alleles M_1 and M_2 with respective frequencies m and $1 - m$. For reasons of comparability, we consider the one-sided $\Lambda_{\pi \neq .5}$ test and denote as "Z_α" and "$Z_{1-\beta}$" the standard normal deviates corresponding to a type I error of α and a power of $1 - \beta$, respectively (e.g., $Z_{1-\beta} = .842$ to achieve a power of 80%). If it is assumed that, under H_1, p follows a normal distribution with expectation p_1 and variance $\sigma^2 = p_1(1-p_1)/n$, then the number of heterozygous M_1M_2 parents, n, is obtained by solving the following equation:

$$2n[p_\beta \mathrm{L}n(p_\beta) + (1-p_\beta) \times \mathrm{L}n(1-p_\beta) - \mathrm{L}n(.5)] = (Z_\alpha)^2 \ ,$$

where $p_\beta = p_1 - \sigma Z_{1-\beta}$. This equation, which has no simple analytical solution, can be solved by a straightforward iterative procedure. From n, the number of necessary families, denoted as "N," is obtained as $N = n/2h$, where h, the probability that a parent with an affected child is M_1M_2, is computed by use of formulas developed by Risch and Merikangas (1996). For example, when $\gamma = 4$ and α is fixed at 5×10^{-8} ($Z_\alpha = -5.33$), we obtain, with the ML-TDT, $N = 139$ for $q = m = .10$ and $N = 96$ for $q = m = .50$, compared with 150 and 103, respectively, for the classic TDT (Risch and Merikangas 1996); the corresponding numbers for $\gamma = 2$ are 680 and 334 for the ML-TDT, compared with 695 and 340.

However, as we have pointed out in a comment elsewhere (Müller-Myhsok and Abel 1997), all computations performed by Risch and Merikangas (1996) were

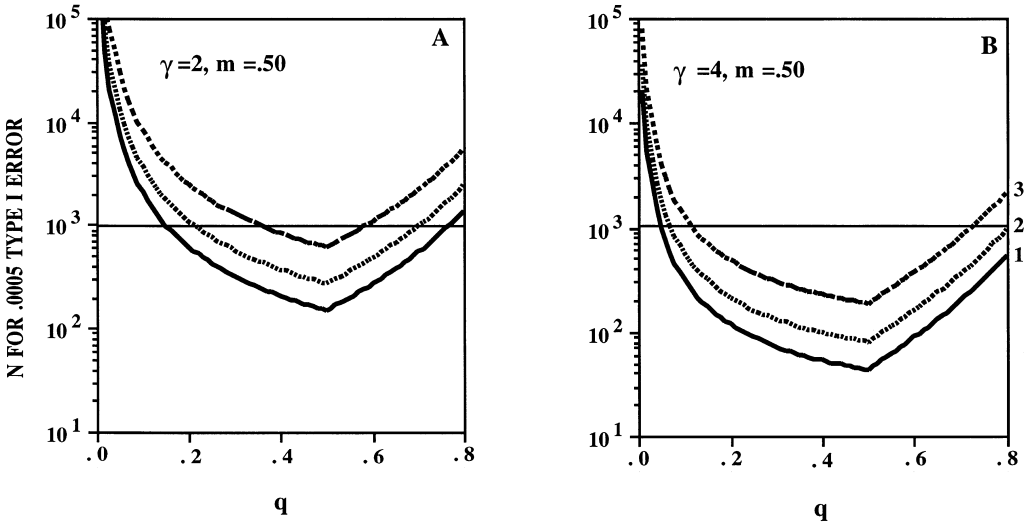

Figure 3 *N* (log scale) required for detection of an association, for a type I error of 5×10^{-4} (candidate-gene strategy) and a power of 80%, according to *q*, for $\gamma = 2$ (*A*) and $\gamma = 4$ (*B*). The curves are drawn for $m = .50$, under the assumption that linkage disequilibrium is equal to δ_{max}, (curve 1), $.75\delta_{max}$ (curve 2), and $.5\delta_{max}$ (curve 3).

based on the optimal assumption that the analyzed allele is the disease allele itself, whereas a more common situation is the analysis of polymorphisms that have a low prior probability of being the disease allele. In this case, we showed by theoretical means that the power of TDT is highly dependent not only on the linkage disequilibrium between the disease allele and the analyzed allele but also on the relative frequencies of both these alleles. In the present paper we illustrate these findings when using the ML-TDT. The coefficient of linkage disequilibrium, δ, is defined as $\text{freq}(AM_1) - qm$, and the maximum value of δ, denoted as "δ_{max}," is reached when $\text{freq}(AM_1)$ is the lowest of the two frequencies *m* and *q*. When the formulas of Müller-Myhsok and Abel (1997) are used, p_1, the expectation of *p*, and *h*, the proportion of heterozygous parents, have the following expressions:

$$p_1 = [1 + (\gamma - 1)\alpha_1]/[2 + (\gamma - 1)(\alpha_1 + \alpha_2)] \ ,$$

where $\alpha_1 = q + (\delta/m)$ and $\alpha_2 = q - [\delta/(1 - m)]$, and

$$h = u/\{u + m^2[1 + (\gamma - 1)\alpha_1] + (1 - m)^2[1 + (\gamma - 1)\alpha_2]\} \ ,$$

where $u = m(1 - m)[2 + (\gamma - 1)(\alpha_1 + \alpha_2)]$.

In the context of a genomewide search, as proposed by Risch and Merikangas (1996)—that is, $\alpha = 5 \times 10^{-8}$—figures 1 and 2 show the variations of *N* as a function of *q* for two values of *m* (.10 and .50), with various strengths of linkage disequilibrium. When $\gamma = 2$ (fig. 1), the required sample size is generally significantly >1,000 families, except for situations close to the optimal case ($m = q$ and $\delta = \delta_{max}$), representing not only a technological challenge but also a major fieldworking effort. A more pronounced gene effect ($\gamma = 4$; fig. 2) allows detection of an association for a larger range of disease-allele frequencies in samples of realistic size (e.g., for $m = .5$, a sample of 1,000 families will lead to the detection of deleterious alleles having frequencies .12–.72 when δ is $\geqslant 75\%$ of δ_{max}). We also examined the strategy of a candidate-gene approach investigating 10 genes with five diallelic markers per gene, which leads to 100 one-sided tests and, consequently, to a required nominal type I error of .0005 for each test, for an overall type I error of .05. Results for $m = .5$ are presented in figure 3 and show that a sample of 1,000 families will allow detection of most alleles when $\gamma = 4$ and $\delta \geqslant .75\delta_{max}$ and also will allow detection of a large number of alleles when $\gamma = 2$, either having a frequency close to *m* or presenting a δ close to δ_{max}. Therefore, at present, unless there is a high chance that the disease allele is among the alleles analyzed in a genomewide search, the candidate-gene approach is a more promising strategy for TDT association studies. Alternatively, the initial use of a lower critical threshold in a genomewide setting (e.g., use of a two-stage strategy, as is commonly done in linkage analysis) can indicate follow-up regions of interest to be tested and can reduce both the genotyping efforts and the necessary sample sizes. It is also important to note that the power of such studies can be greatly influenced by both the dominance model at the disease locus and the resulting genotypic relative risks (Camp 1997). In any case, the ML-TDT appears to be an in-

teresting alternative that can take into account multiallelic markers (Terwilliger 1995) and that easily can be extended to introduce different π parameters, according to some measured factors such as parent gender.

LAURENT ABEL[1] AND BERTRAM MÜLLER-MYHSOK[2]
[1]*INSERM U.436, Mathematical and Statistical Modeling in Biology and Medicine, Hôpital Pitié-Salpêtrière, Paris; and* [2]*Department of Molecular Genetics, Bernhard Nocht Institute for Tropical Medicine, Hamburg*

References

Camp NJ (1997) Genomewide transmission/disequilibrium testing—consideration of the genotypic relative risks at disease loci. Am J Hum Genet 61:1424–1430

Knapp M, Seuchter SA, Baur MP (1994) Linkage analysis in nuclear families. II. Relationship between affected sib-pair tests and lod-score analysis. Hum Hered 44:44–51

Müller-Myhsok B, Abel L (1997) Genetic analysis of complex diseases. Science 275:1328–1329

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

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

Terwilliger JD (1995) A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am J Hum Genet 56: 777–787

Address for correspondence and reprints: Dr. Laurent Abel, INSERM U.436, Hôpital Pitié-Salpêtrière, 91 Boulevard de l'Hôpital, 75013 Paris, France. E-mail: abel@biomath.jussieu.fr

0002-9297/98/6302-0046$02.00