Sutherland GR, Richards RI (1992) Anticipation legitimized: unstable DNA to the rescue. Am J Hum Genet 51:7–9

0002-9297/93/5202-0023$02.00

Am. J. Hum. Genet. 52:434–436, 1993

Resolving an Apparent Paradox Concerning the Role of TGFA in CL/P

To the Editor:

Advances in molecular genetic technology have led to the discovery of the wealth of human polymorphism, which in turn has encouraged the genetic dissection of complex human traits. One recent success with this strategy involves the analysis of a common, severe birth defect, nonsyndromic cleft lip with or without cleft palate (CL/P). A variety of epidemiological studies have shown an important role for genetic susceptibility loci for CL/P. Although the genetics of CL/P are undoubtedly complex, several studies propose a major genetic locus for CL/P.

In an attempt to identify susceptibility loci for CL/P, Ardinger et al. (1989) found an association between RFLPs at transforming growth factor alpha (TGFA) and CL/P, in a patient-control study. This genotype-phenotype association with a TGFA *Taq*I RFLP has subsequently been confirmed in two independent studies (Chenevix-Trench et al. 1991; Holder et al. 1991). A further study detected significant association with a *Bam*HI RFLP (but not with the previously described *Taq*I RFLP) (Stoll et al. 1992). We conclude that there is very strong support, from these population-based genetic studies, for TGFA as a susceptibility locus for CL/P.

It is important to confirm reports of association of disease and polymorphisms at a putative susceptibility locus by means of a linkage study. This will help to resolve epistasis from linkage disequilibrium; furthermore, the proportion of alleles shared by affected relatives will define the contribution of the susceptibility locus to the familial clustering of the trait. Hecht et al. (1991*a*) have recently reported on such a linkage study investigating TGFA segregation in a series of multiplex CL/P families and have interpreted the outcome as excluding TGFA as a candidate locus. We would like to comment on several methodological issues that arise from this study and that may diminish the authors' challenge presented to TGFA's candidature as a susceptibility gene for CL/P.

Hecht et al. (1991*a*) presented lod scores computed for eight informative CL/P multiplex families with between two and five affected individuals in each family. For these calculations, they assumed a dominant mode of inheritance with a disease allele frequency of .001, a phenocopy rate of .001, and penetrances of .32 and .24 in male and female carriers, respectively. These parameters were chosen with reference to an accompanying segregation analysis (Hecht et al. 1991*b*).

Recurrence risks for brothers and sisters of CL/P probands can be calculated, from the parameters estimated in a POINTER mixed-model analysis (see table 6 of Hecht et al. 1991*b*), as 3.5% and 4.1%, respectively, by using the formula of James (1971). These risks are congruous with empirical published risks for siblings, which typically have a range of 2.2% (Carter et al. 1982) to 4.9% (Fogh-Andersen 1942) and which were based on very extensive surveys of CL/P families. The genetic parameters used by Hecht et al. (1991*a*) for linkage analysis predict slightly higher recurrence risks of 6.4% and 4.0% for brothers and sisters, respectively. The latter model therefore anticipates that all the familial clustering of CL/P can be explained by the effects of a single gene.

The validity of a linkage analysis under a major-locus model depends on the premise that the majority of multiplex CL/P families are due to the inheritance of a single locus. By testing for linkage in this way, Hecht et al. (1991*a*) are effectively testing whether all the genetic variance in their multiplex families can be explained by the effects of a single susceptibility locus, namely, TGFA. Failure to detect linkage therefore has several alternative explanations that include the following: (1) the candidate gene and disease are not tightly linked; (2) a single gene does not explain all the genetic variance in these multiplex families; and (3) the sample of pedigrees being tested has insufficient statistical power.

Published data show that a *Taq*I RFLP at the TGFA locus (with alleles C1 [3.0 kb] and C2 [2.7 kb]) shows a highly significant association with CL/P (Ardinger et al. 1989; Chenevix-Trench et al. 1991; Holder et al. 1991). If the genotype data from these three studies are pooled, and if the population prevalence of CL/P is assumed to be .001, then the penetrances of C1C1, C1C2, and C2C2 genotypes are 7.8×10^{-4}, 3.0×10^{-3}, and 5.1×10^{-3}, respectively (TGFA model). If the population allele frequency of the "high-risk" C2 allele is .05, then the recurrence-risk ratio for siblings (λ_{Sib},

which is defined as the risk to a sibling divided by the population prevalence) that is attributable to TGFA is 1.233. In other words, the TGFA susceptibility locus determines a minor fraction of the familial recurrence of CL/P (between 2.5% and 5.6%, depending on which population estimate of λ_{Sib} is taken as the denominator). This point has also been noted by Mitchell and Risch (1992). It is important to remember that these latter estimates are lower limits, as new polymorphisms may be found that show a stronger association with CL/P (e.g., a three-allele single-strand conformation polymorphism, reported by Shiang et al. (1991), indeed promises to show a stronger association).

The effect that substituting the major-locus model and the TGFA model has on linkage calculations is dramatic. For example, for nuclear "dominant" multiplex families, roughly six times as many families would be needed to establish linkage under the TGFA model than under the major-locus model. The implications for excluding a susceptibility locus are equally serious, as more than seven times as many families would be needed to exclude linkage under the TGFA model than under the major-locus model. This is because the observation of an "obligate" crossover carries relatively little weight when analyzed under the TGFA model but would generate an appreciably negative score when analyzed under a major-locus model.

The estimate of the recurrence-risk ratio attributable to TGFA, which was calculated from the TGFA genotype association data, is the lower limit of the true ratio. This is because linkage disequilibrium and/or pleiotropy are genetic mechanisms that explain why the association of RFLP and disease is not absolute (Risch 1987). These alternatives can be resolved by examining the proportion of alleles identical by descent in affected relatives. For example, HLA shows both linkage and association with multiple sclerosis (MS) when λ_{Sib} computed from haplotype-sharing data is 2.4 and λ_{Sib} computed from patient-control DR2 genotype association data is 1.46 (Risch 1987). These results are consistent with linkage disequilibrium between the DR2 allele and MS. Risch (1987) has also commented on a similar phenomenon for linkage and association data for HLA and insulin-dependent diabetes mellitus. If an analogous argument is followed for TGFA-determined susceptibility to CL/P, then the true λ_{Sib} attributable to TGFA in CL/P is likely to be greater than 1.233 but probably not much greater than 2.5.

Risch (1990) has shown that the power of an affected-relative-pair linkage study depends on the recurrence-risk ratio attributable to the susceptibility locus. For example, if λ_{Sib} is 1.233, then 1,650 affected sib pairs would be needed for an 80% probability of generating a lod score greater than 3.0 (with a completely informative marker locus and no recombination). The stringent lod 3.0 threshold is appropriate when the prior chance of linkage is low (1 in 50) (Morton 1955); when there is evidence of an association, then a lower threshold (e.g., lod = 0.83) may be acceptable (Chotai 1984). With this relaxed threshold, about 600 affected sib pairs would be needed to confirm linkage. If the true λ_{Sib} attributable to TGFA is, in fact, greater than 1.233 (e.g., 2.5), then only 130 sib pairs (under the stringent lod 3.0 threshold) or 40 sib pairs (under the relaxed threshold) would be needed.

The strategy of collecting multiplex families, which apparently segregate in a manner consistent with a fairly simple Mendelian model, for a complex genetic trait, with the intent to search for linkage with candidate genes or gene regions, has become increasingly popular in recent years. For example, linkage was found in families with young-onset Alzheimer disease with markers on chromosome 21 or in dominant atopy families with markers on chromosome 11q; in both cases there have also been reports of failure to detect linkage in some follow-up studies. Readers will no doubt also recall that two early claims of linkage to certain psychiatric disorders have been subsequently disproved (e.g., affective disorder and chromosome 11 and schizophrenia and chromosome 5). This has stimulated considerable theoretical interest in the design and execution of such studies and, in particular, in the estimation of type I errors when linkage is evaluated under an incorrect model. We note with regret that there has been relatively little interest in the related and, in our opinion, equally important consideration of type II errors in the analysis of complex traits (for recent comment, see Clerget-Darpoux and Bonaïti-Pellié 1992).

To a certain extent our objections to the Hecht et al. (1991*a*) paper are ideological and hinge on the evaluation of the merits of two alternative strategies. In the first strategy, sufficient multiplex families are collected and analyzed with appropriate methods, to have a good chance of detecting linkage with a sample size computed with consideration of epidemiological data. This has the merit that the type II error is minimized, but it is likely to need large numbers of families. In the second strategy, relatively few anecdotal families that "fit" a simple Mendelian model are tested for linkage, in the hope that a major fraction of the families will be "monogenic." If each family includes a substantial number of affected individuals, then concerns over etiologi-

cal heterogeneity are reduced. Such families will have sufficient power to accept or reject linkage by themselves and to allow formal tests of heterogeneity. The informative families analyzed by Hecht et al. (1991*a*) were not sufficiently large to fulfill these conditions.

The "multiplex monogenic" strategy is splendid when it generates a true positive result (e.g., Alzheimer disease and chromosome 21); however, we anticipate that apparent exclusions will be frequently overinterpreted. This will lead to erroneous exclusion of important susceptibility loci for complex traits after linkage analysis using overly simple genetic models.

In conclusion, the nature of the association between TGFA polymorphisms and CL/P should be explored in a linkage study under an appropriate genetic model. The genotypic association data show that TGFA is a susceptibility locus of modest effect, and this should be allowed for in the specification of genetic models for linkage analysis. Analysis under an inappropriate model may result in erroneous exclusion of a candidate susceptibility locus. Although it is impossible to know the minimum number of families to collect to ensure a good chance of detecting linkage, the maximum number of families *can* be calculated. If studies are designed in this way, then the paradox of inability to confirm linkage in the face of overwhelming evidence of association can be avoided.

MARTIN FARRALL,* KENNETH H. BUETOW,† AND JEFFREY C. MURRAY‡
**MRC Molecular Medicine Group, Royal Postgraduate Medical School, London; †Fox Chase Cancer Center, Philadelphia; and ‡Department of Pediatrics, University of Iowa Hospitals, Iowa City*

References

Ardinger HH, Buetow KH, Bell GI, Bardach J, VanDemark DR, Murray JC (1989) Association of genetic variation of the transforming growth factor-alpha gene with cleft lip and palate. Am J Hum Genet 45:348–353

Carter CO, Evans K, Coffrey R, Fraser Roberts JA, Buck A, Fraser Roberts M (1982) A three generation study of cleft lip with or without cleft palate. J Med Genet 19:246–261

Chenevix-Trench G, Jones K, Green A, Martin N (1991) Further evidence for an association between genetic variation in transforming growth factor alpha and cleft lip and palate. Am J Hum Genet 48:1012–1013

Chotai J (1984) On the lod score method in linkage analysis. Ann Hum Genet 48:359–378

Clerget-Darpoux F, Bonaïti-Pellié C (1992) Strategies based on marker information for the study of human diseases. Ann Hum Genet 56:145–153

Fogh-Andersen P (1942) Inheritance of harelip and cleft palate. Munksgaard, Copenhagen

Hecht JT, Wang Y, Blanton SH, Michels VV, Daiger SP (1991*a*) Cleft lip and palate: no evidence of linkage to transforming growth factor alpha. Am J Hum Genet 49:682–686

Hecht JT, Yang P, Michels VV, Buetow KH (1991*b*) Complex segregation analysis of nonsyndromic cleft lip and palate. Am J Hum Genet 49:674–681

Holder SE, Vintiner GM, Malcolm S, Winter R (1991) Nonsyndromic cleft lip and palate: association and linkage studies with the transforming growth factor-alpha gene. Human Gene Mapping 11 (1991). Cytogenet Cell Genet 58:1869

James JW (1971) Frequency in relatives for an all-or-none trait. Ann Hum Genet 35:47–49

Mitchell LE, Risch N (1992) Mode of inheritance of nonsyndromic cleft lip with or without cleft palate: a reanalysis. Am J Hum Genet 51:323–332

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318

Risch N (1987) Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 40:1–14

——— (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229–241

Shiang R, Lidral AC, Ardinger HH, Murray JC, Buetow KH (1991) Association of TGFA DNA variants with cleft lip and palate (OFC2). Cytogenet Cell Genet 58:1872

Stoll C, Qian JF, Feingold J, Sauvage P, May E (1992) Genetic variation in transforming growth factor alpha: possible association of *Bam*HI with bilateral sporadic cleft lip and palate. Am J Hum Genet 50:870–871

0002-9297/93/5202-0024$02.00

Am. J. Hum. Genet. 52:436–437, 1993

Reply to Farrall et al.

To the Editor:

Farrall et al. point out the dilemma that investigators working on cleft lip with or without cleft palate (CL(P)) have been facing; that is, the heritability of clefting does not fit a straightforward pattern of inheritance. Complex segregation analyses have yielded varying results with major-gene locus, mixed, and multifactorial/