threshold models being implicated by one or more studies (Chung et al. 1986; Marazita et al. 1986, 1992; Hecht et al. 1991*b*). The association of transforming growth factor alpha (TGFA) and CL(P) is fascinating and important and may provide some important insights into the etiology of clefting. There are now three studies (Ardinger et al. 1989; Chenevix-Trench et al. 1991; Holder et al. 1992) confirming and one study (Qian et al. 1991) not confirming the association. However, while we agree that sample-size considerations are important in linkage studies, the results of both our linkage study and an independent linkage study from England have shown that TGFA is not linked to CL(P) in the tested multiplex families (Hecht et al. 1991*a*; Holder et al. 1992). It is also interesting that the same investigators (Vintiner et al. 1992) found an association between TGFA in a group of individuals with CL(P) but found no evidence of linkage when multiplex families were studied (Holder et al. 1992). In fact, their linkage results were strikingly similar to our study results. These linkage studies *do* exclude TGFA as a major gene in these tested families. Further, there is a possibility that TGFA may play an epistatic role in the development of clefting but that it is not the major gene (J. C. Murray, personal communication). This is the same conclusion that we found in our linkage study. Further, we specifically concluded that TGFA may be linked in other multiplex CL(P) families. Additional families are now being tested.

Among the 20 combined multiplex CL(P) families from both published studies (Hecht et al. 1991*a*; Holder et al. 1992), the C2 allele was identified in 4 families and did not segregate with the putative disease locus. It will be interesting to study additional multiplex CL(P) families with the C2 allele, to determine whether it is linked in those families. For now, the association- and linkage-study results suggest that the causes of CL(P) are heterogeneous. Time and future studies will explain the probable myriad of causes that contribute to and cause facial clefting.

JACQUELINE T. HECHT,* YAPING WANG,*
SUSAN H. BLANTON,† VIRGINIA V. MICHELS,‡
AND STEPHEN P. DAIGER*

**Health Sciences Center, University of Texas, Houston; †University of Virginia, Charlottesville; and ‡Mayo Clinic/Foundation, Rochester, MN*

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

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

Chung CS, Bixler D, Watanabe T, Koguchi H, Fogh-Andersen P (1986) Segregation analysis of cleft lip with or without cleft palate: a comparison of Danish and Japanese data. Am J Hum Genet 39:603–611

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, Farren B, Malcolm S, Winter RM (1992) Confirmation of an association between RFLPs at the transforming growth factor-alpha locus and non-syndromic cleft lip and palate. J Med Genet 29:390–392

Marazita ML, Goldstein AM, Smalley SL, Spence MA (1986) Cleft lip with or without cleft palate: reanalysis of a three generation family study in England. Genet Epidemiol 3:335–342

Marazita ML, Hu D-N, Spence MA, Liu Y-U, Melnick M (1992) Cleft lip with or without cleft palate in Shanghai, China: evidence for an autosomal major locus. Am J Hum Genet 51:648–653

Qian JF, Stoll C, Feingold J, Sauvage P, May M (1991) No association of genetic variation of the transforming growth factor alpha with cleft lip and palate. Am J Hum Genet Suppl 51:A347

Vintiner GM, Holder SE, Winter RM, Malcolm S (1992) No evidence of linkage between the transforming growth factor-alpha gene in families with apparently autosomal dominant inheritance of cleft lip and palate. J Med Genet 29:393–397

0002-9297/93/5202-0025$02.00

Am. J. Hum. Genet. 52:437–440, 1993

Forensic Population Genetics and the National Research Council (NRC)

To the Editor:

In response to calls from the scientific and legal communities, the Board on Biology of the National Research Council established a Committee on DNA Technology in Forensic Science. This committee has now issued a

useful report that generally favors the forensic uses of DNA and that seeks to give guidelines to defuse some of the controversy that has attended these uses. Much of the controversy has centered on population genetic issues, and the NRC report addresses these issues. There is a danger that the impact of the report will be diluted by several instances where population genetic and statistical issues are treated with less care than would be expected for a report from such a distinguished panel.

On page 13 the report states: "Some legal commentators have pointed out that frequencies should be based on the population of possible perpetrators, rather than on the population to which a particular suspect belongs. Although that argument is formally correct, practicalities often preclude use of that approach." It is not clear what these "practicalities" might be, but it is clear that calculations should be based on the population of perpetrators. The point has been made verbally by Weir and Evett (1992) and formally by Evett and Weir (1992). A simple informal demonstration will be helpful here: suppose that E is the event that both a suspect and some crime-scene material have DNA profile A. There are two possibilities: C, that the crime-scene material came from the suspect, and $\bar{C}$, that it came from someone else. In the first case, $\Pr(E|C) = \Pr_s(A) \times 1$, where $\Pr_s(A)$ is the probability that a person from the population to which the suspect belongs has profile A. In the second case, $\Pr(E|\bar{C}) = \Pr_s(A) \times \Pr_p(A)$, where $\Pr_p(A)$ is the probability that a person in the population to which the perpetrator belongs has profile A. It is assumed here that, under $\bar{C}$, the perpetrator and the suspect are unrelated. The likelihood ratio $\Pr(E|C)/\Pr(E|\bar{C})$ is just the reciprocal of $\Pr_p(A)$. Note that the population, or ethnic background, of the suspect has cancelled out and so is quite irrelevant. This ratio is a very convenient way of expressing the results. A jury could be told, for example, that "the evidence is a million times more likely to have arisen if the crime scene material was left by the defendant than if it was left by some unrelated person."

Not only is the last statement convenient but also it is easy to understand. On page 62 the report states:

> Some have advocated that testing laboratories, instead of using a match criterion, should report a likelihood ratio—the ratio of the probability that the measurements would have arisen if the samples came from the same person to the probability that they would have arisen if they came from different persons. No testing laboratories in the United States now use that approach. The committee recognizes its intellectual appeal, but recommends against it. Accuracy with it requires detailed information about the joint distribution of fragment positions, and it is not clear that information about a match could be understood easily by lay persons.

The committee has blurred the distinction between the simple argument of the last paragraph and the use of analyses based on continuous fragment lengths. The committee also neglects to mention at this point that such likelihood ratios are used by all U.S. paternity-testing laboratories, under the name of "paternity index," and are routinely presented to laypersons in jury trials of paternity suits.

Use of the likelihood ratio also has the advantage that the quantity $\Pr(E|\bar{C})$ can be modified easily for the case when the perpetrator has a specified relationship to the suspect. Similar calculations are performed in the paternity setting. On page 87 of the NRC report a result is given for full sibs: "Roughly speaking, the probability of a match at k loci will be approximately $(0.25 + 0.5p + 2p^2)^k$ in the general population, where p is the average chance that two alleles will match (i.e., the apparent homozygosity rate.)" Evidently this formula contains a typographical error, as the coefficient of p^2 was meant to be .25. There is no reason, however, for the report not to give the correct formula. If a locus has alleles A with frequencies p_A, then the probability that two full sibs have the same genotype at that locus is $[1 + 2\Sigma_A p_A^2 + 2(\Sigma_A p_A^2)^2 - \Sigma_A p_A^4]/4$.

In several places the NRC report discusses confidence limits on frequencies. On page 76, it incorrectly states: "If the pattern occurred in 1 of 100 samples, the estimated frequency would be 1%, with an upper confidence limit of 4.7%. (The upper bound cited is the traditional 95% confidence limit, whose use implies that the true value has only a 5% chance of exceeding the upper bound.)" The correct confidence limit is $.01+1.645\sqrt{.01 \times .99/100} = .026$, and the NRC value does not obtain even if a two-sided confidence interval is used. It is not correct to speak of the chance with which the true value exceeds a limit. It is the varying limit that has a 95% chance of exceeding the fixed true value. Although it is not clear why the report obtains the wrong value in this case, on page 92 the wrong value in the formula "$p+1.96\sqrt{p(1-p)/N}$" comes from using the standard normal value 1.96 for a two-sided 95% confidence interval, instead of the value 1.645 required for a one-sided 95% confidence interval. The need for forensic calculations to be conservative means that one-sided confidence intervals are required. Using normal approximations for binomial proportions away from .5 is less of a problem than the implication of the table on page 92, that the confidence limit for a product is the product of the confidence limits of the com-

ponent parts. Confidence limits are useful, since they reflect the effect of sample size on the precision of estimates, but it is not an easy task to arrive at an analytical expression for the confidence limit for a product of binomial variates, some of which come from a single sample. The simplest solution is probably to use numerical resampling (e.g., see Efron 1982).

The major population genetic recommendation of the report is in use of the "ceiling principle." On page 92 the report states: "For each allele, a modified ceiling frequency should be determined by (1) calculating the 95% upper confidence limit for the allele frequency in each of the existing population samples and (2) using the largest of these values or 10%, whichever is larger." The rationale for this approach follows from recognizing that the product over alleles j of the frequencies p_{ij} in some population i is bounded by the product of the maximum over populations of these frequencies:

$$\prod_j p_{ij} \leq \prod_j \max_i(p_{ij}) .$$

It is also true, however, that the product $\prod_j p_{ij}$ is bounded by the maximum over populations of the products:

$$\prod_j p_{ij} \leq \max_i(\prod_j p_{ij}) .$$

There are several advantages to the second version of the ceiling principle. In the first place, it recognizes that a DNA profile is contributed by one person and not by several different people in different populations. Second, it recognizes the empirical findings (e.g., see Weir 1992*a*, 1992*b*) that differences between allele frequencies in different populations are diminished when collections of several different alleles are used. It cannot be true that all alleles at one locus in one population are more frequent than all those in another population. Over loci, it is observed that some alleles are more frequent in the first of two populations, while others are less frequent. Under the NRC version of the ceiling principle, a defendant could very well conduct a global search for samples in which each allele in the matching profile had a very high frequency and then amalgamate those frequencies, even if the *profile* had comparable frequencies in each of those samples.

The ceiling principle is based either on arbitrary bounds of 10% "designed to address a remaining concern that populations might be substructured in unknown ways with unknown effect" (p. 92) or on upper 95% confidence limits providing "a pragmatic approach to recognize the uncertainties in current population sampling" (p. 92). This care for the effects of sampling is removed on page 93, with the assertion that "the availability of data based on a more rigorous sampling scheme will make it unnecessary to take an upper 95% confidence limit for each allele frequency." No amount of rigor is going to remove sampling variability, although it will be appropriate to place confidence limits on the profile frequencies, rather than on the allele frequencies.

The NRC report is correct to require independence of allele frequencies (p. 91), "provided . . . statistical evaluation of Hardy-Weinberg equilibrium and linkage equilibrium has been carried out . . . and no significant deviations [have been] seen," although careful writing would have made the distinction between linkage disequilibrium and the dependence of the frequencies of all alleles in a profile (e.g., see Weir and Cockerham 1989). There seems to be some inconsistency on page 83: "it [the ceiling principle] is believed to be conservative even if there are correlations among alleles because of population substructure." Cohen (1992) has demonstrated that the presence of linkage disequilibrium can cause the ceiling principle to underestimate a profile frequency. The same happens with Hardy-Weinberg disequilibrium. Suppose that a population consists of two equal subpopulations in which the frequencies of allele A are p_1 and p_2, with $p_1 > p_2$. Then, if each of the populations has the same degree f of departure from Hardy-Weinberg, so that the frequencies of individuals heterozygous for allele A are $2p_1(1-p_1)(1-f)$ and $2p_2(1-p_2)(1-f)$, the ceiling principle estimates the population-wide heterozygous frequency as $2p_1(1-p_2)$. This is an underestimate of the true frequency whenever $f < -(p_1-p_2)(1+p_1-p_2)/[p_1(1-p_1)+p_2(1-p_2)]$. The trouble is that the NRC report advocates that data be collected from several populations with samples of size 100. Such sizes may be too small to allow adequate testing of independence of frequencies within each population. A noncentral χ^2 approximation (Weir 1990) shows that f^2 must be $>.1051$ in order for a Hardy-Weinberg test on a sample of size 100 to be significant at the 5% level with 90% probability. In other words, the amount of disequilibrium required to cause the ceiling principle to underestimate the true frequency may not be detected in a sample of size 100. Testing in the actual population to which an unknown perpetrator belongs, when a defendant has denied responsibility for a crime, will not be possible at all if that population is not known or sampled. On the other hand, it must be pointed out that the amount of underestimation is not

likely to be of practical importance with the actual frequencies found in forensic data bases (Weir 1992*a*, 1992*b*).

Whenever departure from independence is found for any pair of alleles, within or between loci, only one of those alleles should be used in the estimation of profile frequencies. Under the ceiling principle as advocated by the NRC report, such departures would require the omission of alleles in all data bases, even if the disequilibrium was seen in only one data base. This drastic loss of information would be prevented by determining profile frequencies within each data base before looking for maximum values.

The concern of the NRC report that DNA profile frequencies be estimated conservatively would appear to be met in a defensible way by collecting large samples from several populations. Each sample should be assessed for independence of allele frequencies, and only alleles with independent frequencies should be used in products to estimate profile frequencies. The maximum profile frequency over populations would furnish a conservative estimate of the frequency of the profile of interest. A measure of precision of this estimate would be provided by using bootstrapping to generate the distribution of profile frequency estimates within each data base.

B. S. WEIR

Program in Statistical Genetics
Department of Statistics
North Carolina State University, Raleigh

Acknowledgments

I had useful correspondence with Joel Cohen, Ian Evett, and Newton Morton. This work was supported in part by NIH grant GM45344.

References

Cohen JE (1992) The ceiling principle is not always conservative in assigning genotype frequencies for forensic DNA testing. Am J Hum Genet 51:1165–1168

Efron B (1982) The jackknife, the bootstrap and other resampling plans. CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38. SIAM, Philadelphia

Evett IW, Weir BS (1992) Flawed reasoning in court. Chance 4:19–21

Weir BS (1990) Genetic data analysis. Sinauer, Sunderland, MA

——— (1992*a*) Independence of VNTR alleles defined as fixed bins. Genetics 130:873–887

——— (1992*b*) Independence of VNTR alleles defined as floating bins. Am J Hum Genet 51:992–997

Weir BS, Cockerham CC (1989) Complete characterization of disequilibrium at two loci. In: Feldman MW (ed) Mathematical evolutionary theory. Princeton University Press, Princeton, pp 86–110

Weir BS, Evett IW (1992) Whose DNA? Am J Hum Genet 50:869

0002-9297/93/5202-0026$02.00

Am. J. Hum. Genet. 52:440–441, 1993

Errors in Human Gene Mapping II

To the Editor:

The catalogs of DNA polymorphisms, published by the Human Gene Mapping (HGM) workshops, are referred to extensively by molecular geneticists: the HGM 10 listing (Kidd et al. 1989) has more than 180 citations. While using the more recent HGM 11 listing (table 4 of Williamson et al. 1991), I have encountered a disturbing number of avoidable errors, which I believe merit wider attention.

The most serious error is that a number of previously listed loci seem to have been accidentally erased. These include 10 of 370 reference markers that are defined as "the best marker(s) available for a given chromosome interval" (table 3 of Williamson et al. 1991). The missing loci (D1S47, D3S16, D5S43, D7S21, D7S22, D17S5, D18S18, D19S20, DXYS14, and DXS255) include some very well-known polymorphic DNA probes, such as M27β and several of the hypervariable minisatellites from Alec Jeffreys' laboratory. I have counted 20 additional discrepancies between tables 3 and 4 of Williamson et al. (1991), in the listing of the reference markers. Erasures of nonreference loci have also occurred, but I have not quantified these.

Allele frequency data on individual loci are helpful in assessing the loci's informativeness, expressed either as a simple heterozygosity (h) or as the PIC; the values of h and PIC are similar when $h \sim 1$ but diverge at low h (Botstein et al. 1980). Table 4 of Williamson et al. (1991) generally quotes PIC values, but these are sometimes misleadingly high and hence may overestimate the utility of a particular probe. Such errors arise when (1) h, rather than PIC, has actually been calculated or (2) the allele frequencies add up to less than 1, as a result of the omission of alleles, the ignoring of null alleles, or the introduction of rounding errors.