Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 165–169, January 1988
Genetics

Nonrandom distribution of structural mutants in ethylnitrosourea-treated cultured human lymphoblastoid cells

SAMIR M. HANASH*†, MICHAEL BOEHNKE‡, ERNEST H. Y. CHU§, JAMES V. NEEL§, AND RORK D. KUICK§

Departments of *Pediatrics, University of Michigan Medical School, ‡Biostatistics, School of Public Health, and §Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109-0618

Contributed by James V. Neel, August 24, 1987

ABSTRACT **Two-dimensional polyacrylamide gel electrophoresis has been used to detect somatic cell gene mutations altering protein structure, following ethylnitrosourea treatment of cultured human lymphoblastoid cells. A total of 267 polypeptides encoded by 263 loci were scored in a series of 1143 lymphoblastoid clones. Sixty-five electrophoretic mutants were detected at a total of 49 loci. Sixteen of the 65 mutations were phenotypically repeat mutations, occurring at 11 loci. Furthermore, structural mutations occurred more frequently at loci known to be polymorphic. These results provide evidence that the mutations that are detectable at the protein level by two-dimensional polyacrylamide gel electrophoresis do not occur at random and that their frequency is greater among polymorphic loci.**

The question of how random mutational events are distributed throughout the genome of higher eukaryotes is not only of broad evolutionary interest but also has practical relevance in experimental mammals and humans; also of broad interest is the question of what are the potential errors in extrapolating from mutation rates at a relatively few loci to the entire genome? Because two-dimensional polyacrylamide gel electrophoresis permits the clear visualization of hundreds of cellular polypeptides (1), to monitor human populations for changing mutation rates, we have been exploring the potential of two-dimensional PAGE coupled with silver staining for polypeptide localization (2). *N*-Ethyl-*N*-nitrosourea has been shown to be a very potent mutagen in the mouse specific-locus test, leading to its use as a model chemical mutagen (3). To determine the usefulness and limitations of two-dimensional PAGE for mutation studies, we have used a human diploid lymphoblastoid cell line (TK-6) mutagenized with *N*-ethyl-*N*-nitrosourea (4). Evidence is presented for the nonrandom distribution of structural gene mutations among the 263 scored loci encoding proteins.

MATERIALS AND METHODS

Experimental Design. The human lymphoblastoid cell line TK-6 was kindly provided by William Thilly (5). A subclone with a high plating efficiency of ≈70% was isolated. The subclone could be maintained in exponential growth for an indefinite period as a proliferating mass culture. The experimental design consisted of treating a mass culture with either *N*-ethyl-*N*-nitrosourea at 50 μg/ml for 40 min or five successive daily exposures to the mutagen at 10 μg/ml. The former treatment resulted in 95% cell killing, and the latter resulted in 60% cell killing. Single-cell clones were isolated at various times following treatment and individually cultured, until a population size of 8–12 × 10^6 cells was obtained. Harvested cells were either pelleted for two-dimensional PAGE analysis or stored frozen in liquid nitrogen for subsequent propagation and repeated analysis. Control clones were isolated from an untreated mass culture under conditions comparable to those for treated clones.

Two-Dimensional Electrophoresis. Cell pellets were solubilized by addition of 40 μl of a lysis buffer consisting of (per liter) 9 M urea, 40 ml of Nonidet P-40, 20 ml of ampholytes (pH 3.5–10.0), 20 ml of 2-mercaptoethanol, and 9.2 mM of phenylmethylsulfonyl fluoride in distilled deionized water. The pH was adjusted to 9.5. In most cases, a 30-μl aliquot of the solubilized cells (1.5 × 10^6 cells) was applied onto isoelectric focusing gels immediately following preparation. First-dimension gels contained 20 ml of ampholytes per liter (pH 3.0–10.0), and focusing was carried out at 1200 V for 16 hr and 1500 V for the last 2 hr. Twenty gels were focused simultaneously. For the second-dimension separation, an acrylamide gradient of 11.5–14.0 g/dl was used, in which there was 2.6% crosslinking with bisacrylamide. The polypeptides in the gels were visualized by the silver-staining technique of Merril *et al.* (6). The preparation that results from these procedures is illustrated in Fig. 1.

Gel Analysis. A subset of 267 polypeptide spots was selected for scoring. The choice of spots was made by an investigator with no prior knowledge of their genetic variability. The selected spots covered a wide range of intensity and were sufficiently above the detection threshold that if mutation of one allele resulted in a new spot and in a reduced intensity of the spot representing the gene product of the common allele, both spots would be readily detected. Spots in crowded regions of the gel or those associated with streaking were not included in the subset. A further prerequisite was that the spot could be unambiguously scored in at least 95% of the gels. In practice, the scorability of the members of this battery of spots averaged 98%. Spots were visually scored for their presence or absence and for the occurrence in a clone of a new polypeptide associated with a reduced intensity of a neighboring polypeptide.

We have examined (7) the cellular polypeptide constituents of lymphocytes obtained from a group of Caucasoid children to detect genetically determined polypeptide variants by two-dimensional PAGE. For each child, gels from both parents were also examined to substantiate the genetic basis of any possible variants. Of the 106 polypeptides scored in 40 families, 14 were determined to be genetically polymorphic. Ten of these 14 polymorphic polypeptides were among the 267 spots scored for mutations in the TK-6 cell line. The individual from whom TK-6 cells were derived was heterozygous for two of these polypeptides. In addition, during the course of the study it became very probable, on the basis of spot characteristics, that, in the TK-6 cell line, two additional polypeptides were the products of alleles at an additional locus exhibiting heterozygosity. The family studies had not revealed (7) polymorphism at this locus. Thus 14 of the polypeptides scored were the products of 11 loci,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†To whom reprint requests should be addressed.

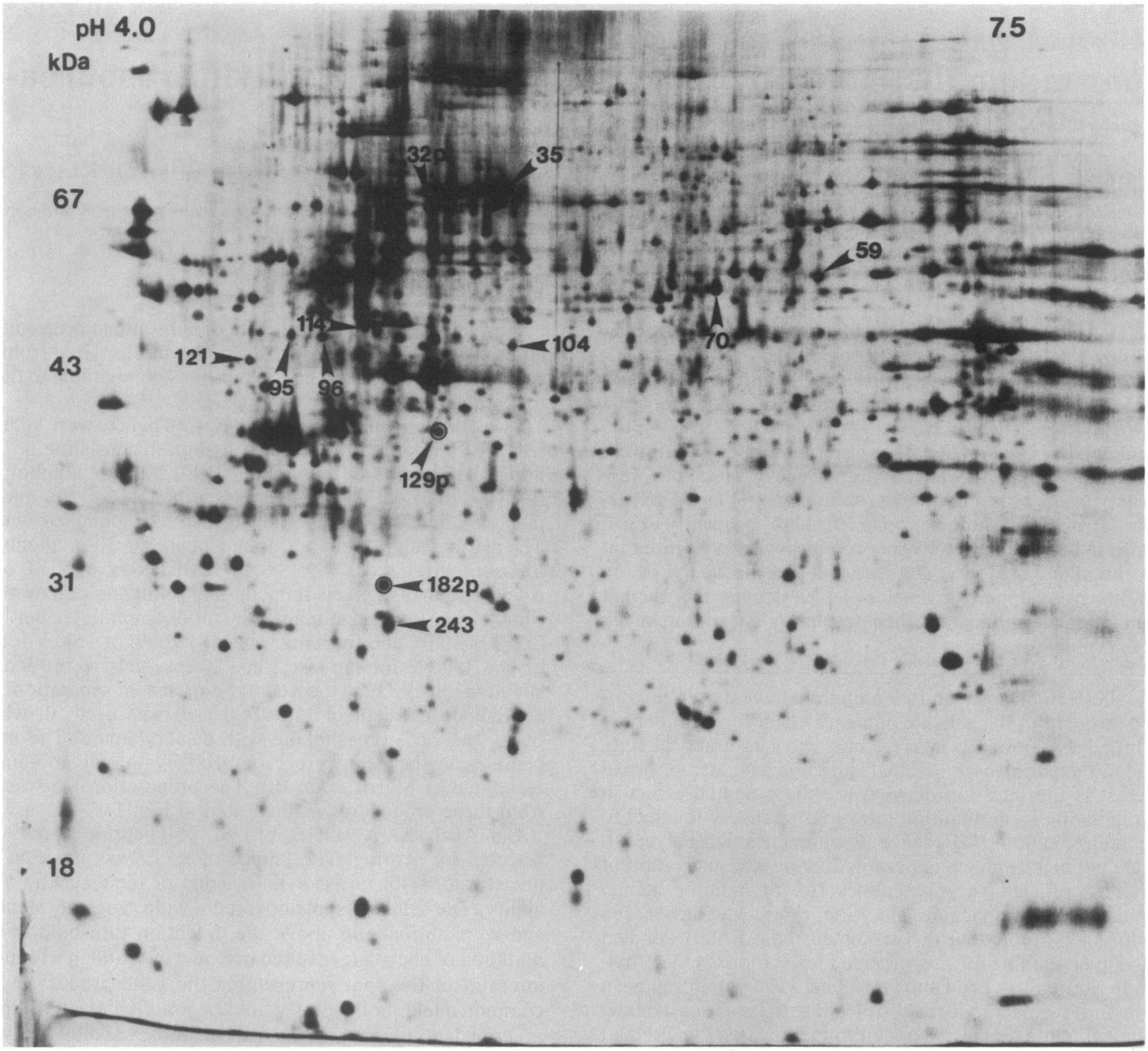

FIG. 1. Two-dimensional pattern of the silver-stained polypeptides from the TK-6 lymphocytoid cell line. The pH gradient of the preparation is indicated at the top of the figure and the molecular mass in kDa on the left vertical axis. Arrows designate polypeptides associated with more than one mutation. The circled spots designated with the letter p correspond to previously identified polymorphic loci that, in this study, have mutated more than once.

10 of which were known to be polymorphic on the basis of family studies (7).

Three clones in which a mutation was detected exhibited simultaneous alterations of the same pair of neighboring polypeptides. This phenomenon, of conjugated variation, is best explained on the basis of a primary gene product and a derivative product of the same protein. Thus the total of 267 spots scored corresponds to a maximum of 263 loci—this is the maximum because there may be other examples of unknown polymorphisms or derivative gene products among the spots scored.

Statistical Design. We consider two hypotheses of the randomness of the induced variants. The first is that mutation is equally likely to occur at any of the loci scored. The second is that mutant probability is proportional to molecular weight, this variable being chosen as the best available proxy for number of base pairs within exons of the corresponding gene.

These two hypotheses are examined in two ways. First, we consider all loci separately and ask whether the overall pattern of mutants is consistent with random mutation. We define a detected mutant as a "repeat" if the locus at which the mutant occurs is the site of a previously detected mutant. To test the general hypothesis of randomness of mutation, we count the number of repeat mutants and calculate the probability of observing as many or more repeat mutants assuming randomness. Second, we focus specifically on the 10 known-to-be-polymorphic loci and ask whether these loci are more mutable than the remaining 253 loci.

To test the hypothesis that mutation is equally likely at all loci, we make use of an occupancy problem described by Feller (8). The test for randomness is based on a recursion relation for calculating the probability $P(k,M;N)$ of k urns (loci) with zero balls (i.e., no mutants) given a total of M balls (mutants) placed at random in N urns (loci). The recursion relation is:

$$P(k,M;N) = P(k,M-1;N)[(N-k)/N] + P(k+1,M-1;N)(k+1)/N$$

Thus the probability of k empty urns given M balls equals the

probability of k empty urns given $M - 1$ balls times the probability the Mth ball occupies an already occupied urn plus the probability of $k + 1$ empty urns given $M - 1$ balls times the probability the Mth ball occupies a previously empty urn. Boundary conditions for the recursion are: $P(k,1;N) = 0$ ($k = 0, 1, 2, \ldots, N - 2$), and $P(N - 1,M;N) = (1/N)^{M-1}$ ($M = 1,2, \ldots$).

The number of repeats R can be calculated directly from k, M, and N. In particular, $k = N - M + R$. That is, the number of empty cells equals the minimum number of empty cells ($N - M$) plus the number of repeats. Hence, the number of repeats $R = k - N + M$. Combining this fact with the calculated probabilities $P(k,M;N)$ results in the probability distribution for R, assuming each locus is equally likely to mutate. Calculating the probability that R is at least as large as the observed number of repeats gives the associated exact one-sided P value for this general test of randomness.

We tested the hypothesis that mutation probability is proportional to molecular weight by computer simulation. To do so, we divided the interval (0,1) into 263 subintervals. The length of the subintervals was chosen to be proportional to the approximate molecular weights of the proteins coded for by the 263 loci. We then generated sets of M random numbers on the interval (0,1). If a random number fell into the subinterval corresponding to a particular locus, a "mutation" is said to have occurred at that locus. Among the M "mutants," we then noted the number of repeat mutations that occurred. We repeated this process for 100,000 sets of 65 mutants. By doing so, we estimated the distribution of the number of repeat mutations assuming mutation probability proportional to molecular weight. By summing the probabilities corresponding to at least as many repeat mutations as actually observed, we obtain a good estimate of the one-tailed P value associated with the observed numbers of repeats for testing the hypothesis of mutation probability proportional to molecular weight.

We are also interested in the relative mutability of polymorphic and nonpolymorphic loci. Assuming mutation has an equiprobability of occurrence at all the loci whose products are scored, the probability a mutant occurs at a polymorphic locus is equal to the proportion of tested loci that are polymorphic. Assuming mutation probability proportional to molecular weight, the probability a mutant occurs at a polymorphic locus is equal to the fraction of the total molecular weight contributed by the polymorphic loci. In either case, under the additional assumption that the various mutants represent independent events, the number of mutants at polymorphic sites is distributed as binomial. To evaluate whether an excess number of mutants is observed at the polymorphic loci, we calculate the exact binomial tail probability corresponding to at least as many mutants at polymorphic loci as actually observed, resulting in a one-tailed P value.

RESULTS

In these experiments, a total of 1143 clones were scored for 267 spots corresponding at most to the products of 263 loci. The TK-6 line was derived from a male. If all the polypeptides were encoded by autosomal loci, there would have been ≈600,000 locus product tests. The true number would be somewhat less, in part because of the sex-linked inheritance of some polypeptides, in part because some spots represent derivative rather than primary gene products. A total of 65 mutants were identified by the appearance of an additional spot on the gel, differing from the presumed spot of origin by charge or molecular weight or both. The mutants were detected at a total of 49 loci. Three mutants were observed at each of five loci, two mutants at each of six loci, and one mutant at each of 38 loci. No mutants were observed at any of the remaining 214 loci. There were hence 16 repeats. Assuming random mutation and using our occupancy problem approach, we find that the probability of observing 16 or more repeats in this situation is <0.0005. Thus, we strongly reject the general hypothesis of randomness for the mutants detectable by two-dimensional PAGE.

First, to test whether the mutant frequency increased with increasing molecular weight, we divided the gel into two halves of equal size on the molecular weight dimension: 157 loci are found in the upper half of the gel (36–100 kDa), 106 loci in the lower half (10–35 kDa). Of the mutants, 50 fell in the upper half, 15 in the lower half. These data strongly argue that the higher molecular weight proteins are more mutable [χ^2 goodness-of-fit statistic: $\chi^2 = 8.02$; $df = 1$; $P < 0.005$ (where df is degrees of freedom)]. Assuming mutation probability proportional to molecular weight, only 279 of the 100,000 simulation replicates resulted in as many as 16 repeats. Thus, even taking into account molecular weight and hence approximate gene size, we still reject the hypothesis of randomness ($P < 0.003$).

Interestingly, of the 65 mutants, 8 occurred at the 10 loci known to be polymorphic, while 57 occurred at the remaining 253 loci. Fig. 2 represents a selection of the mutants detected at the polymorphic loci. Thus, mutants detectable by two-dimensional PAGE were detected ≈3.6 times as often at the loci known to be polymorphic as compared to the loci not known to be polymorphic. The corresponding binomial tail probability assuming mutants equally likely at each locus is <0.004. Assuming mutant probability proportional to molecular weight, the binomial-tail probability remains <0.004. Thus, there was also evidence that mutants detectable by two-dimensional PAGE occurred preferentially at polymorphic loci. It should be noted that some of the 253 remaining loci may be genetically polymorphic in human populations. However, any such misclassification should have biased the analysis toward the null hypothesis of randomness.

We can also test the general hypothesis of randomness of mutants within the polymorphic and not-known-to-be polymorphic loci. At the 10 loci known to be polymorphic, we detected 8 mutants. Three mutations were detected at each of 2 loci, and 2 mutations were detected at a third locus; thus there were a total of 5 repeats. The probability of 5 or more mutant repeats under the hypothesis of mutation equally likely at each locus is <0.008. Assuming mutant probability proportional to molecular weight, only 1297 of 100,000 replicates resulted in as many as 5 repeat mutations. Even taking into account molecular weight, we still reject the hypothesis of randomness ($P < 0.02$). Thus, we have evidence that also among the polymorphic loci, there appears to be nonrandomness of mutants as detected by two-dimension-

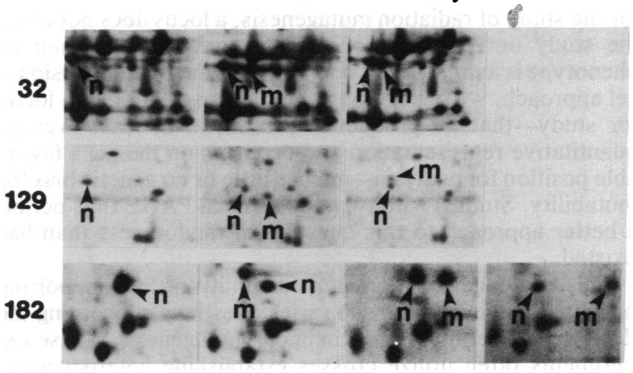

FIG. 2. Illustration of the multiple mutations of polymorphic polypeptides 32, 129, and 182. n, Peptide in the normal position; m, mutants. The differences between the individual mutations are highly reproducible.

al PAGE. At the remaining 253 loci not known to be polymorphic, we detected 57 mutants. Three mutants were detected at each of 3 loci, and 2 mutants were detected at each of 5 loci, for a total of 11 repeats. The probability of 11 or more repeats assuming mutants equally likely at each locus is <0.02. The estimated probability based on computer simulation of 11 or more repeats assuming mutant probability proportional to molecular weight is <0.05. Thus, again we have the suggestion of a departure from randomness. Since some of these 253 loci may actually be polymorphic, it is possible that our conclusion of a nonrandom distribution of mutants among the loci not known to be polymorphic could be due to misclassification of some polymorphic loci.

DISCUSSION

The data indicating a nonrandom distribution of electrophoretic mutants can be interpreted by either of two quite different hypotheses. The first is that the mutational process is nonrandom with respect to the loci investigated. The second is that mutation rates are more or less uniform across the loci under consideration, but the protein phenotypes resulting from this process are subject to differential selection, presumably predominantly negative but in principle also possibly positive. Whereas the latter hypothesis cannot be excluded, it requires a degree of selection for or against a heterozygous protein variant that to us seems inherently unlikely. In addition, since the selection would presumably be predominantly negative, the *N*-ethyl-*N*-nitrosourea-induced rate of mutation would have to be considerably higher than the already quite high rate these data imply (unpublished data). We prefer the first hypothesis.

There is already considerable published evidence for the nonrandomness of both spontaneous and induced somatic and germinal mutation in mammals, both within and between genetic loci (e.g., refs. 9–14). Johnson and Lewis (13) studying chemically induced germinal mutation in mice by electrophoretic techniques, reported that over half of the mutants they detected occurred at only 3 of the 21 loci screened, and Siciliano *et al.* (14) observed differential mutational sensitivity at 44 loci encoding for isozymes in Chinese hamster ovary cells exposed to various doses of ethyl methanesulfonate or ultraviolet radiation. Since the polymorphic nature of the loci scored in these studies was not established, an association between mutability and polymorphism could not be made. A difficulty in putting this question of mutational nonrandomness into perspective has been that biases may easily have inadvertently crept into the selection of the loci whose mutability has been studied in the past. Thus, in such a standard approach as the multiple locus test for visibles of Russell (10), and Carter *et al.* (11) developed for the study of radiation mutagenesis, a locus does not enter the study until at least one readily differentiated mutant phenotype is available. By contrast with the two-dimensional gel approach, we believe our criteria for selection of a locus for study—that its product is present in at least average quantitative representation and occupies on the gel a favorable position for analysis—impose little or no genetic bias for mutability. Studies with two-dimensional PAGE thus permit a better approach to this question of randomness than has existed.

This evidence for a greater mutability of loci supporting protein polymorphisms also raises questions concerning the design of experiments on chemical mutagenesis. These experiments often utilize crosses establishing heterozygosity for protein polymorphisms, the polymorphism having been detected in the course of population surveys (e.g., ref. 13). The data of the present paper suggest there may have been inadvertent selection for more mutable loci.

Repeat mutations scored at a single locus may be phenotypically the same or phenotypically different. Phenotypically different mutants or mutants detected in separate experiments clearly represent independent mutational events. Phenotypically identical repeat mutations detected in the same experiment could represent mutant cells related by descent from only a single mutational event. While it seems unlikely that this is the case in these experiments, given the large number of cells in the original culture, it is worth considering the impact of such a possibility on this analysis. Six of the repeats detected, one among the known-to-be-polymorphic loci and five among the not-known-to-be-polymorphic loci, could be descendants by these criteria. Even if we exclude these possible descendants, we still have evidence that induced mutation occurs at a higher rate at the known-to-be-polymorphic loci ($P < 0.007$ for mutation equally likely at all loci; $P < 0.009$ for mutation probability proportional to molecular weight) and that mutation is not at random within these loci ($P < 0.03$ for mutation equally likely at all loci; $P < 0.02$ for mutation probability proportional to molecular weight). However, evidence against the hypothesis that mutation is at random among all loci ($0.06 > P > 0.05$ for a mutation equally likely at all loci; $P > 0.12$ for a mutation probability proportional to molecular weight) or among the not-known-to-be-polymorphic loci ($P > 0.30$ for a mutation equally likely at all loci, $P > 0.50$ for a mutation probability proportional to molecular weight) is attenuated.

In mutagenesis pilot studies using two-dimensional PAGE, Zeindl *et al.* (15) detected two structural mutations in 26 clones derived from Chinese hamster ovary cells treated with methylnitrosourea. The total number of locus tests in that study was estimated to be 33,800. Klose (16) detected with two-dimensional PAGE two mutant electromorphs in liver preparations of 463 mice whose male parent was treated with methylnitrosourea. There were no mutants in preparations from 312 progeny of controls. The number of spots scored is not given. Marshall *et al.* (17) detected two structural mutations in two-dimensional preparations of the livers of the offspring of *N*-ethyl-*N*-nitrosourea-treated male mice. Sixty-seven offspring were examined, resulting in a total of ≈46,000 locus tests. The small number of mutations detected in these studies and the lack of polymorphism data pertaining to the polypeptides scored did not permit an evaluation of differential locus sensitivity to mutation.

The presence of three heterozygous loci in this series provided the additional opportunity to search for mutations resulting in the disappearance of a gene product. Nine such occurrences were noted. The interpretation of this phenomenon was complicated when the density of the remaining spot was quantified (18) and found in five instances to have significantly increased in intensity—to an average of ≈168% over the intensity of the corresponding spot in the heterozygote (unpublished results). This finding, on the basis of present knowledge, can most readily be explained by mitotic recombination between homologous chromosomes (combined with random variation in spot intensities). Evidence that this phenomenon may occur in mammals (humans, mice, and Chinese hamster cells) has been forthcoming (19–24). We would thus prefer not to speculate at this time concerning the relative frequencies of mutation resulting in electromorphs and mutation resulting in loss of gene product. The pertinent point in the present context is that the data we are presenting on the greater frequency with which induced variants are observed at loci supporting genetic polymorphisms suggest that even when the ultimate basis for these loss-of-gene-product variants has been sorted out, caution must be exercised in extrapolating to the entire genome from loss-of-product findings at polymorphic loci.

We thank Dr. Robert Wolfe for directing our attention to ref. 8. This work was supported in part by Public Health Service Grants CA-26803 and HL-16008.

1. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250,** 4007–4016.
2. Neel, J. V., Rosenblum, B. B., Sing, C. F., Skolnick, M. M., Hanash, S. M. & Sternberg, S. (1984) in *Methods and Applications of Two Dimensional Gels Electrophoresis of Proteins*, eds. Celis, J. E. & Bravo, R. (Academic, New York), pp. 315–340.
3. Russell, W. L., Kelly, E. M., Hunsicker, P. R., Bangham, J. W., Maddox, S. C. & Phipps, E. L. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5818–5819.
4. Hanash, S. M., Chu, E. H. Y., Kuick, R., Skolnick, M. M., Neel, J. V., Strahler, J., Pivirotto, S. & Niezgoda, W. (1987) *Proteins* **2,** 13–19.
5. Thilly, W. G., DeLuca, J. G., Furth, E. E., Hoppe, H., Kaden, D. A., Krolewski, J. J., Liber, H. L., Skopek, T. R. & Slapikoff, S. A. (1980) *Chem. Mutagens* **6,** 331–364.
6. Merril, C. R., Goldman, D., Sedman, S. A. & Ebert, M. H. (1981) *Science* **211,** 1437–1438.
7. Hanash, S. M., Baier, L. J., Kuick, R., Galteau, M. & Welch, D. (1986) *Am. J. Hum. Genet.* **39,** 317–328.
8. Feller, W. (1968) *An Introduction to Probability Theory and Its Applications* (Wiley, New York), 3rd Ed., Vol. 1, pp. 31–40.
9. Green, E. L., Schlager, G. & Dickie, M. M. (1965) *Mutat. Res.* **2,** 457–465.
10. Russell, W. L. (1951) *Cold Spring Harbor Symp. Quant. Biol.* **16,** 327–336.
11. Carter, T. C., Lyon, M. F. & Phillips, R. J. (1956) *Br. J. Radiol.* **29,** 106–117.
12. Ehling, U. H. (1978) *Chem. Mutagens* **5,** 233–256.
13. Johnson, F. M. & Lewis, S. E. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3138–3141.
14. Siciliano, M. J., Stallings, R. L., Humphrey, R. M. & Adair, G. M. (1986) *Chem. Mutagens* **10,** 509–531.
15. Zeindl, E., Sperling, K. & Klose, J. (1982) *Mutat. Res.* **97,** 67–78.
16. Klose, J. (1977) *Arch. Toxicol.* **38,** 53–60.
17. Marshall, R. R., Raj, H. S., Grant, F. J. & Heddle, J. A. (1983) *Can. J. Genet. Cytol.* **25,** 457–466.
18. Kuick, R. D., Hanash, S. M., Chu, E. H. Y. & Strahler, J. R. (1987) *Electrophoresis* **8,** 199–204.
19. Cavenee, W. K., Dryja, T. P., Phillips, R. A., Benedict, W. F., Godbout, R., Gallie, B. L., Murphree, A. L., Strong, L. C. & White, R. L. (1983) *Nature (London)* **305,** 779–784.
20. Wasmuth, J. J. & Vock Hall, L. (1984) *Cell* **36,** 697–707.
21. Rajan, T. V., Halay, E. D., Potter, T. A., Evans, G. A., Seidman, J. G. & Margulies, D. H. (1983) *EMBO J.* **2,** 1537–1542.
22. Kipps, T. J. & Herzenberg, L. A. (1986) *EMBO J.* **5,** 263–268.
23. Potter, T. A., Zeff, R. A., Frankel, W. & Rajan, T. V. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1634–1637.
24. Langlois, R. G., Bigbee, W. L., Kyoizumi, S., Nakamura, N., Bean, M. A., Akiyama, M. & Jensen, R. H. (1987) *Science* **236,** 445–448.