Journal of Cellular Biochemistry 30:319–329 (1986)
Biochemical and Molecular Epidemiology of Cancer 99–109

Human Restriction Fragment Length Polymorphisms and Cancer Risk Assessment

Theodore G. Krontiris, Nancy A. DiMartino, Mark Colb, H. David Mitcheson, and David R. Parkinson

Division of Hematology/Oncology and Cancer Research Center (T.G.K., N.A.D., M.C., D.R.P.) and Department of Urology (H.D.M.), Tufts-New England Medical Center, and Department of Molecular Biology and Microbiology, Tufts University School of Medicine, (T.G.K.), Boston, Massachusetts 02111

The polymorphic restriction fragments of the human Ha-*ras* locus, produced by the variable tandem repetition (VTR) of a short consensus sequence, fall into three classes based on allelic frequencies. Alleles of the "rare" class (individual frequencies $<0.5\%$) have been detected only in white blood cell and tumor DNA of cancer patients. This phenomenon is independent of ethnic origin. No significant association of rare alleles with cancer patients has been demonstrated at an independent tandem repeat locus, VTR4.1. The results suggest that the Ha-*ras* restriction fragment length polymorphism is useful in cancer risk assessment.

Key words: **oncogenes, Ha-*ras*, restriction fragment length polymorphism, human genetics, cancer risk assessment**

The recent advances in molecular biology have had a major impact on our understanding of the pathogenesis of cancer. At the same time, concepts that have been applied to the diagnosis of genetic diseases hold promise for the analysis of inherited susceptibility to cancer. One of the most important of these concepts is the restriction fragment length polymorphism (RFLP). When restriction endonucleases were employed in the structural analysis of DNA, it was shown that hereditary variations such as deletions, insertions, and even point mutations could be reproducibly demonstrated by this approach [reviewed in 1]. Thus polymorphic changes in DNA sequence could result in altered restriction fragment length as detected by digesting DNA with the appropriate enzyme and then subjecting the products to agarose gel electrophoresis. These migration differences were first detected in cloned

Abbreviations used: RFLP, restriction fragment length polymorphism; WBC, white blood cell; VTR, variable tandem repeat; GU, genito-urinary.

Received September 3, 1985; revised and accepted November 11, 1985.

© **1986 Alan R. Liss, Inc.**

DNA; but, with the advent of Southern blotting, polymorphisms in genomic DNA also became accessible. The key maneuver in this type of analysis is choosing the proper restriction endonuclease. This is largely a matter of trial and error, although certain enzymes with CpG in the recognition sequence (MspI and TaqI, for example) more frequently reveal RFLPs [2].

DNA polymorphisms detected by restriction endonucleases fall into one of two categories. In the first, called *site polymorphism,* the recognition site for a given enzyme in a particular region of DNA either appears or disappears as the result of point mutation. Two phenotypes are therefore possible: the two different fragment sizes generated by the site-present and site-absent alleles. Such a polymorphism, based on the specific sequence of one restriction endonuclease, will be revealed only by that enzyme. In the second category, called *insertion/deletion polymorphisms,* variation in fragment length is the result of insertion or deletion of DNA sequences. This polymorphism will be detected by any restriction endonuclease that possesses recognition sites tightly spanning the region of sequence alteration. Since the insertions or deletions can assume a continuum of lengths, more than two alleles are possible.

RFLPs in human DNA were first described in the β-globin gene cluster [3,4]. These were site polymorphisms identified within the δ-globin gene [3] and near the β-globin gene [4]. The clinical significance of these results was immediately appreciated; the β-globin RFLP was associated with the sickle trait. Insertion/deletion polymorphism in human DNA was first demonstrated in a region with unknown function [5]. Since that time, insulin [6], ζ-globin [7], and Ha-*ras* [8–11] have all been linked to polymorphisms of the insertion/deletion type.

The list of DNA polymorphisms associated with disease loci continues to expand. These data currently provide the basis for many active investigations on the utility of RFLPs in both prenatal diagnosis and risk assessment. In addition to the diagnostic applications, RFLPs promise to revolutionize human genetic mapping. Botstein et al [12] have pointed out that the systematic collection of 150–400 independent probes for polymorphic regions of human DNA should provide a definite set of genetic markers with which to map the entire genome. A cooperative effort on the international level is now being organized [13].

Because of the highly polymorphic nature of the human Ha-*ras* gene, we analyzed the distribution of allelic restriction fragments in cancer patients and cancer-free controls. Our results suggest the Ha-*ras* RFLP may prove useful in determining cancer risk. Here we report updated results from a continuing study [first described in 11].

MATERIALS AND METHODS

Study Population and Sample Collection

Unrelated whites, without a personal or first-degree family history of cancer, comprised the control or normal population referred to below. Cancer patients, again unrelated whites, were enlisted from the New England Medical Center Hospitals. These patients demonstrated a variety of tumors, including carcinomas of the head and neck, breast, lung, and lower gastrointestinal tract, sarcomas, melanomas, acute and chronic lymphocytic and nonlymphocytic leukemias, and Hodgkin's and non-Hodgkin's lymphomas.

WBC were separated from peripheral blood by dextran sedimentation and frozen at −20°C until further use. Discarded tumor tissue from surgical resection or biopsy was frozen in liquid nitrogen or dry ice-ethanol and stored at −70°C.

Southern Blotting

DNA was extracted from WBC or tumor tissue, digested with restriction endonuclease, fractionated on agarose gels, and transferred to nitrocellulose as previously described [11]. Radiolabeled probe, utilizing the human Ha-*ras* plasmid pEC [14], was prepared by nick-translation [15] and employed in filter hybridization [18] as described by Der et al [19].

Isolation of VTR4.1

The 990 bp MspI fragment of pEC, which contains the Ha-*ras* variable tandem repeat (VTR), was nick-translated, and the resulting probe was used to screen a human phage library [3] by standard methods of in situ hybridization [17]. Filters were hybridized at 50°C in 5× SET (1× SET = 0.15 M NaCl, 1 mM EDTA, 20 mM Tris pH 8) and washed several times at room temperature and 37°C in 2× SET. One phage of approximately 200,000 screened demonstrated a strong signal. The phage was isolated, after which the region of homology to the Ha-*ras* VTR was identified and subcloned. DNA sequencing of the subclone revealed the tandem repetition of a 35 bp consensus sequence. [M. Colb, B. Mermer, and T.G. Krontiris, in preparation]. The region, designated VTR4.1, has been used as probe in the survey of human DNAs as described in Results.

RESULTS

Approximately 1,000 bp downstream from the polyadenylation signal of the human Ha-*ras* gene is a region of tandemly repeated nucleotides [9]. A 28 bp consensus sequence is aligned head-to-tail from 30 to 100 times. This VTR is the basis for the Ha-*ras* polymorphism [8–11].

The enzyme combination MspI/HpaII has sites closely flanking the Ha-*ras* VTR; digestion occurs about 115 bp upstream and 50 bp downstream. The remainder of the Ha-*ras* sequences recognized by the probe are digested so extensively that they are not seen under the gel conditions we use. (The isoschizomers are employed together to reduce the effect of DNA methylation on digestion.) Southern blots of over 350 WBC and tumor DNA preparations have revealed at least 24 fragments with distinct gel migration patterns (Table I, Fig. 1).

Four of the fragments revealed by MspI/HpaII (MH) digestion were quite common, accounting for ~93% of the total number analyzed (Table I). The individual frequencies in this group ranged from 7% to 65%. Another five alleles occurred at intermediate frequencies, from 0.5% to 1.5% for each allele. Finally, we detected

TABLE I. Distribution of Ha-*ras* Alleles

Class	Members	Frequency
Common	a1, a2, a3, a4	0.07–0.65
Intermediate	a1.1, a1.2, a1.3, a4.1, a5	0.005–0.015
Rare	Total of 15	<0.005

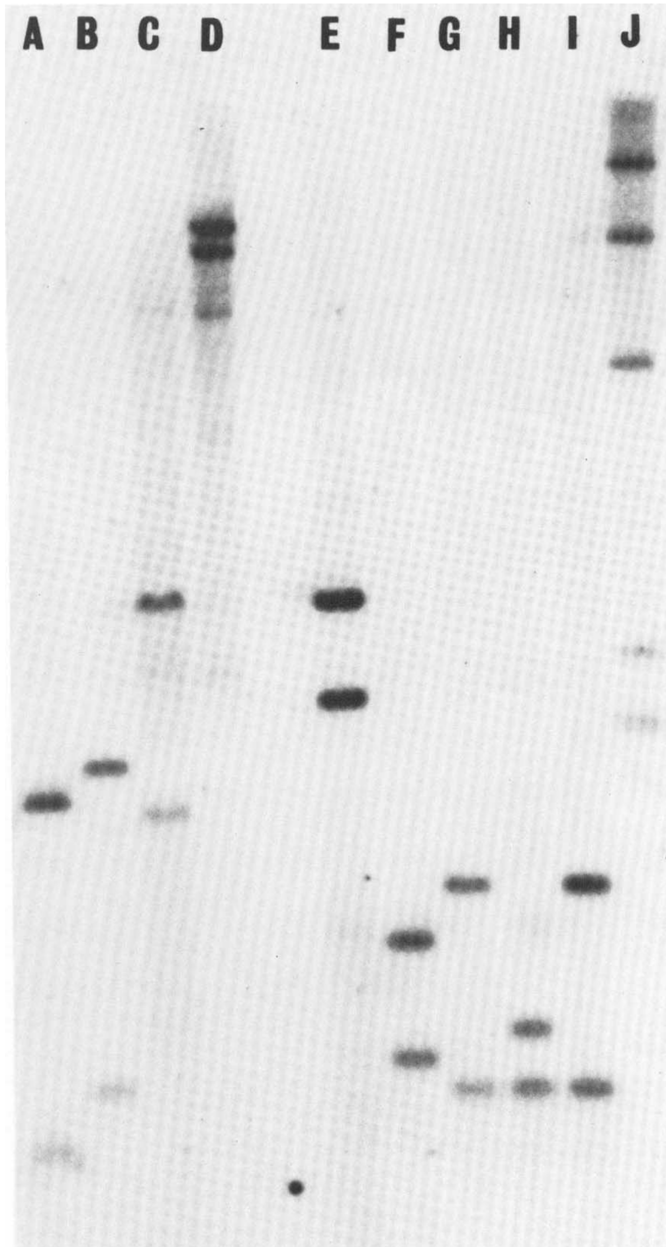

Fig. 1. Polymorphic fragments of the Ha-*ras* gene. WBC DNA was digested and blotted as described *in Materials and Methods. Probe was the nick-translated* 6.6 kb BamHI *fragment* of *pEC*. Lanes A–D represent the MspI/HpaII/PstI (A), MspI/HpaII (B), PstI (C), and BamHI (D) digests of a WBC DNA with a1 and a2.2 allelic fragments. Lanes E–I are the MspI/HpaII digests of a marker DNA with a3 and a4 fragments (E), a leukemia DNA with a1.1 and a1.4 fragments (F), a marker DNA with a1 and a2 fragments (G and I), and a leukemia DNA with a1 and a1.2 fragments (H). J is HindIII-digested and end-labeled λDNA; the 23 kb band is poorly visualized here.

a large class with at least 15 different members. Each of these fragments represented < 0.5% of the total.

Figure 1 shows examples of several alleles in the intermediate and rare categories. Lane E contains the MH digest of a DNA with the common a3 and a4 alleles. Lanes G and I have the MH digest of one with the common a1 and a2 alleles. The MH digest of DNA in lane B demonstrates the rare allele a2.2 as the upper band and common allele a1 as the lower band. Lanes F and H, also MH digests, demonstrate unusual alleles present in two acute myelogenous leukemia DNAs. Two unusual fragments in these lanes are in the intermediate class: a1.1 the lower band in F and a1.2 the upper band in H. The upper band in F is the rare allele a1.4.

When the MH digest reveals an unusual fragment, we repeat the digest with several enzymes to rule out partial digestion and MH site polymorphism. For example, lanes B–D in Figure 1 are the MH, PstI, and BamHI digests of the DNA with the a2.2 allele. Note that two prominent bands representing the Ha-*ras* gene are present in each digest.

Figure 1 also demonstrates the resolution of our gels. From Ha-*ras* sequence data [9], we have determined that the MspI/HpaII site 5′ to the tandem repeat is 113 bp from the PstI site also on that side (itself 3 bp from the start of the tandem repeat). Lane A is an MspI/HpaII/PstI triple digest adjacent to the MH digest in lane B. This shows the migration difference produced by 115 bp in the region of a1 (lower pair of fragments in A and B) as well as in the region of a3 (upper pair of fragments). Since the Ha-*ras* tandem repeat monomer is 28 bp long, we are capable of discerning differences of two repeat units (lane F, lower fragment) and, occasionally, even of one repeat unit. Such resolution is diminished, of course, in higher-molecular-weight regions near a3 and a4.

When we analyzed the distribution of common, intermediate, and rare alleles in cancer patients and cancer-free controls, we found that rare alleles appeared only in cancer patients (Tables II and V). Intermediate alleles also occurred more frequently in cancer patients, although this association was not as strong as that observed for rare alleles.

We have observed rare restriction fragments in first-degree relatives of cancer patients. This indicates that unusual Ha-*ras* alleles are transmitted in a mendelian fashion and, at least for the cases we have observed, do not arise de novo in cancer patients. As an example, Figure 2 shows a pedigree from a large kindred with the Von Hippel Lindau syndrome. We have studied nearly 100 members of the family; the pedigree in Figure 2 is being presented because it illustrates the pattern of transmission of one rare allele, a2.2, and one intermediate allele, a1.2. Cross-hatched symbols indicate affected individuals. The index case, indicated by the arrow, was

TABLE II. Comparison of Rare Alleles in Control and Cancer DNAs*

DNA source	Common alleles	Intermediate alleles	Rare alleles	Probability
Control WBC	311	11	0	—
Tumors + pt. WBC	394	27	31	<0.001
Pt. WBC	241	19	24	<0.001
Tumors	153	8	7	<0.01

*Probability by χ^2 (2 df). Comparisons are control WBC vs Tumor + pt. (patient) WBC; control WBC vs pt. WBC; and control WBC vs tumor.

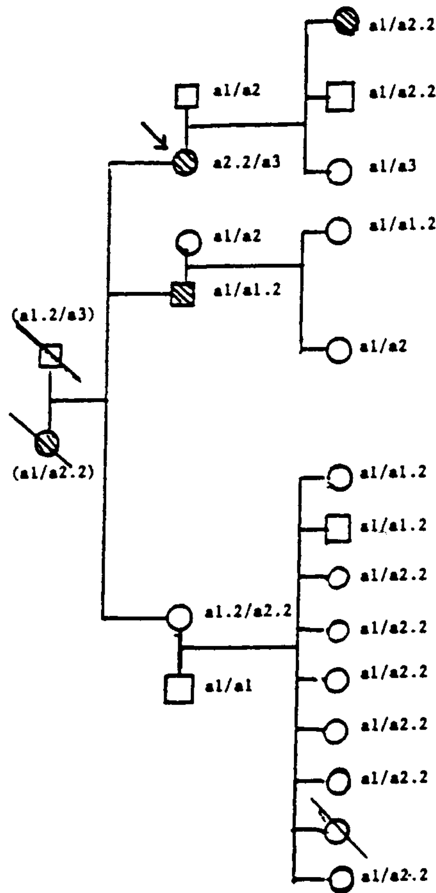

Fig. 2. Pedigree from a kindred with the Von Hippel Lindau syndrome. See text for details.

noted to have the rare allele a2.2. The genotypes in generation I were deduced both from sibs in the VHL kindred and from the progeny shown here. Although the number of genotypes is small, we can certainly conclude that both rare and intermediate alleles do not necessarily arise de novo in cancer patients. Both the a2.2 and a1.2 alleles are present in multiple sibs of generation II and were transmitted to progeny of generation III.

The ethnic composition of control and cancer populations could obviously be an important source of bias in our study. Data on the ethnic group/national origin of study entrants were collected and monitored to determine if ethnic imbalance materially influenced our results (Table III). The four grandparents of each entrant were classified by ethnic group/nationality. The number of grandparents in each ethnic category was determined and this figure expressed as a percentage of the total number of grandparents. Four additions were made to the "unknown" category when no ethnic data were available for a given individual. The distribution of rare alleles among these groups is given in Table III. In seven instances, an allele appeared in an individual with grandparents in two ethnic groups; in these cases, the alleles were

TABLE III. Distribution of Rare Alleles by Ethnic Group/Nationality

	Percent of grandparents		Rare alleles	
Group	Cancer	Control	Cancer	Control
Unknown	35.8	27.4	11	0
Irish	14.0	27.3	3	0
Italian	14.3	7.7	4	0
English	5.7	7.6	3	0
Jewish	7.6	9.2	5	0
German	3.7	3.9	2	0
Scottish	3.6	2.0	2	0
French	3.0	2.2	0	0
Polish	1.9	4.0	1	0
Portuguese	2.4	0.9	1	0
Russian	0.2	1.2	0	0
Estonian	0.0	0.5	0	0
Ukranian	0.4	0.0	1	0
Lithuanian	0.9	0.9	0	0
Swedish	1.0	0.5	1	0
Norwegian	0.0	0.2	0	0
Danish	0.2	0.5	1	0
Finnish	0.2	0.5	0	0
Austrian	0.0	0.5	0	0
Swiss	0.0	0.2	0	0
Greek	1.3	0.5	0	0
Albanian	0.2	0.0	0	0
Lebanese	0.4	0.0	0	0
Armenian	0.4	0.0	1	0
Syrian	0.4	0.0	1	0
AmIndian	0.0	0.7	0	0
Canadian	1.1	0.9	0	0
Spanish	0.0	0.1	0	0
Hungarian	0.0	0.5	0	0
Dutch	0.4	0.1	1	0

listed under both ethnic groups. Several considerations support the conclusion that allele distribution is unaffected by ethnic composition.

First, the number of rare alleles in each ethnic group, including "unknown" is proportional to the representation of that ethnic group in the "cancer" sample. No single ethnic group, including "unknown," is contributing a disproportionate number of rare alleles. Therefore, even a great disparity in the ethnic composition of "cancer" and "control" would not be expected to bias the results. Corroborating evidence for this conclusion is found in the "Irish" category. There, an excess of Irish entries has recently accumulated in the "control" group. Despite this difference, rare alleles still predominate in the "cancer" group. So, when the sample is skewed in the "Irish" category by 1:2, "cancer" to "control," the alleles are still 3:0 in the opposite direction. Similarly, there is a slight excess of Italian patients in the "cancer" group, but, once again, the number of unusual alleles is proportional to the number of Italian grandparents. Second, when rare alleles appear more than once, they appear in more than one ethnic group. The allele a2.2, for example, has appeared in a patient with four Irish grandparents and a patient with four Armenian grandparents.

Bladder carcinoma is potentially a useful tumor model for the interaction of environmental and host susceptibility factors. Ha-*ras* gene activation by mutation is prominent in this tumor type. Also, environmental carcinogens play a prominent etiologic role. The mutation of Ha-*ras* by ultimate carcinogens has been observed [20]. Finally, bladder cancer patients are at greater risk for developing a second primary tumor in a different tissue. We have been blotting WBC DNA from patients with bladder cancer, nonbladder GU malignancies (mostly prostate and renal), and nonmalignant GU disease (mostly infertility and benign prostatic hypertrophy). As is shown in Table IV, there is a much greater prevalence of unusual Ha-*ras* alleles in patients with bladder malignancy.

These results are particularly interesting because of the nature of the alleles detected. Eight of 13 are within one to three tandem repeat monomers of the a1 allele (monomer length 28 bp). The others migrate very near a2 (one), and a3 (two), or a4 (two). Thus, in contrast to most rare alleles, which are quite distinct from the common alleles (see again lane B in Fig. 1), the unusual alleles associated with bladder carcinoma are grouped around specific deviations from the common alleles. Most of them appear identical to the intermediate alleles a1.1 and a1.2. Because of this clustering, we have considered the possibility that "bladder" alleles did not arise from changes in tandem repeat length but rather from clustered point mutations resulting in MH site polymorphisms. Such an outcome would, of course, imply hypermutability of the Ha-*ras* gene near its VTR. We are now conducting detailed restriction mapping of the WBC DNAs displaying these alleles to distinguish MH site polymorphism from minor VTR changes.

Approximately 40% of patients from families with prominent histories of cancer demonstrate rare alleles. However, our preliminary linkage data indicate that Ha-*ras* is not identical to, nor closely linked to, the disease locus in several cancer syndromes. Although we must be careful about generalizing from the cancer family results, the emerging picture is that Ha-*ras* is not the primary determinant of disease in patients with rare alleles. Therefore, the two broadest alternatives for the association phenomenon are 1) that the Ha-*ras* gene, its VTR, or a closely linked gene tangentially participates in the pathogenesis of a given tumor in a patient with a rare fragment—say, through the intervention of a germline mutant Ha-*ras* gene during tumor progression, or 2) the that Ha-*ras* locus is irrelevant to pathogenesis, the VTR being merely a marker for genomic instability—say, increased sister chromatid exchange—in those people destined to get cancer. The investigation of the former possibility requires cloning and characterizing rare allelic fragments. To pursue the latter possibility, we have begun to isolate clones representing other tandem repeat loci and to perform population studies with them. In this way, we can compare the population genetics of VTRs to determine if the Ha-*ras* phenomenon is unique or global. If the phenomenon is global, we would systematically collect and characterize VTRs for their use as a battery of predictive markers.

TABLE IV. Unusual Ha-*ras* Alleles in Bladder Carcinoma

	Unusual alleles/total patients
Bladder cancer	13/31
Nonbladder GU cancer	2/20
Benign GU disease	1/25

We have cloned another tandem repeat using the Ha-*ras* VTR as probe. The clone, designated VTR4.1, recognizes a locus independent from Ha-*ras*. The enzyme HaeIII just spans the repeat region, which consists of head-to-tail arrays of a 35 bp consensus sequence. Therefore, we have used this enzyme to digest and Southern blot nearly all of the DNAs screened in the Ha-*ras* population study. (The VTR4.1 survey employed about 20 DNA samples not screened in the Ha-*ras* survey, and vice versa.) Six of the seven fragments detected with this probe are shown in Figure 3. The seventh (a7) is an apparent deletion of a large segment of the VTR in a cancer-free individual. The fragment counts and frequencies (we have not yet formally demonstrated allelism with all fragments of VTR4.1) are listed in Table V; fragments with an asterisk (*) have been detected only in cancer patients. Unlike that of the Ha-*ras* gene, polymorphic variation of VTR4.1 is relatively restricted. There is no significant accumulation of rare fragments in cancer patients. Therefore, this locus does not demonstrate the same population behavior.

DISCUSSION

Our survey of Ha-*ras* restriction fragments in cancer patients and cancer-free controls has revealed a marked inequality in the distribution of fragments with allelic

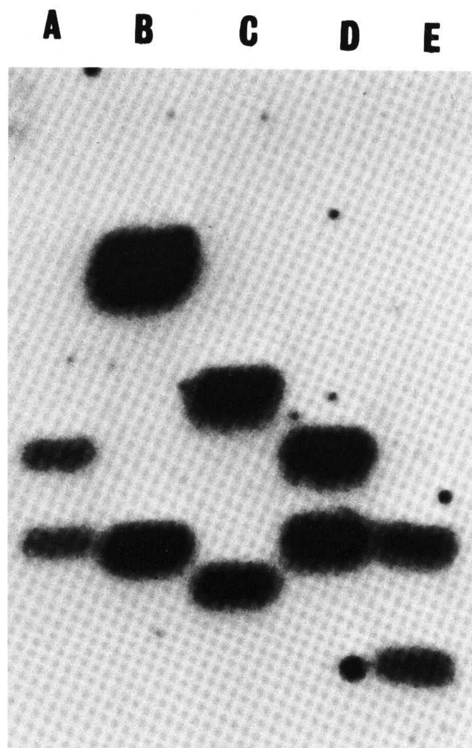

Fig. 3. Polymorphic fragments of the VTR4.1 locus. Five WBC DNAs were digested with HaeIII and blotted. Probe was the nick-translated VTR4.1 tandem repeat. DNAs possessing fragments corresponding to alleles a1 and a4 (lanes A and D), a1 and a5 (B), a2 and a3 (C), and a1 and a6 (E) are shown. a3 is the upper band of lane C.

TABLE V. Comparison of Allele Distribution at Two Distinct VTR Loci*

Ha-*ras*		4.1	
Allele	No.	Allele	No.
a1	493	a1	651
a2	86	a2	89
a3	70	a3	26
a4	56	a4	2
a1.2	15	a5*	2
a1.1	8	a6*	1
a4.1	7	a7	1
a5	4		
a1.3	4		
a2.2*	3		
a3.2*	3		
a0.1*	2		
a2.01*	2		
a1.4*	2		
a3.1*	2		
a2.02*	1		
a2.1*	1		
a2.11*	1		
a2.12*	1		
a4.2*	1		
a3.3*	1		
a3.4*	1		
a2.015*	1		
a2.025*	1		
a>1*	2		
a>2*	1		
a>3*	4		
a>4*	1		
Total	774		772

*The asterisk (*) indicates alleles detected only in DNAs from tumors or patient WBC. a > 1 designates unusual fragments migrating between a1 and a2; a > 2, between a2 and a3; a > 3, between a3 and a4; and a > 4, above a4. These fragments do not appear to comigrate with intermediate alleles but have not yet been further characterized.

frequencies below 0.5%. These rare alleles have thus far been observed only in cancer patients. From studies with first-degree relatives and kindreds, as summarized above, it is likely that the mode of transmission for rare alleles is truly mendelian. It is quite possible that these alleles arise over much shorter time periods than one would ordinarily expect genetic variation to occur. Jeffreys et al [17] reported the de novo appearance of a new allele at another tandem repeat locus. Perhaps such repeats are more unstable in high-risk groups. In any event, once a new *ras* fragment appears, its mode of inheritance is apparently uncomplicated.

We have also shown that the ethnic distribution in our study populations is unlikely to affect our conclusions. First, no ethnic group with a demonstrably low frequency of rare alleles is overrepresented in the "control" group. Second, no ethnic group with a demonstrably high frequency of rare alleles is overrepresented in the "cancer" group.

The high degree of polymorphism and the association of many rare alleles with cancer patients was not observed with another tandem repeat locus, VTR4.1. At the present time, we cannot ascribe this difference to some intrinsic feature of Ha-*ras* (thereby favoring a direct role of this region in pathogenesis) or to the unsuitability of VTR4.1. We will continue to search for an independent VTR locus as polymorphic as Ha-*ras* to repeat this type of comparison. A population analysis with insulin, which does have as many allelic fragments as *ras*, is being contemplated. We would prefer, however, to conduct the initial studies with unlinked markers (Insulin, like Ha-*ras*, is on 11p).

We have recently obtained molecular clones of the common (a1) and unique (a2.1) allelic fragments from the WBC DNA of a patient with familial melanoma. Studies are now in progress to determine if phenotypic differences between the clones can be detected. This approach, together with continuing population studies, may lead to an understanding of the molecular genetic basis of cancer risk.

ACKNOWLEDGMENTS

This work was supported by the American Institute for Cancer Research and by NIH grant CA37866 as well as by the Laub Memorial Fund of the New England Medical Center Hospitals. N.A.D. and M.C. were supported by postdoctoral training grants HL07437 and CA09429, respectively. T.G.K. is a fellow of the John A. and George L. Hartford Foundation.

REFERENCES

1. Nathans D, Smith H: Annu Rev Biochem 44:273, 1975.
2. Barker D, et al: Cell 36:131, 1984.
3. Lawn RM, et al: Cell 15:1157, 1978.
4. Kan YW, Dozy AM: Proc Natl Acad Sci USA 75:5631, 1978
5. Wyman AR, White R: Proc Natl Acad Sci USA 77:6754, 1980.
6. Bell GI, et al: Nature 295:31, 1982.
7. Goodbourn SEY, et al: Proc Natl Acad Sci USA 80:5022, 1983.
8. Goldfarb M, et al: Nature 296:404, 1982.
9. Capon DJ, et al: Nature 302:33, 1983.
10. White R, et al: Nature 313:101, 1985.
11. Krontiris TG, et al: Nature 313:369, 1985.
12. Botstein D, et al: Am J Hum Genet 32:314, 1980.
13. Marx JL: Science 229:150, 1985.
14. Chang EH, et al: Nature 297:479, 1982.
15. Maniatis T, et al: Proc Natl Acad Sci USA 72:1184, 1975.
16. Maniatis T, et al: "Molecular Cloning". Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1982.
17. Jeffreys AJ, et al: Nature 314, 67, 1985.
18. Southern EM: J Mol Biol 98:503, 1975.
19. Der CJ, et al: Proc Natl Acad Sci USA 79:3637, 1982.
20. Marshall C, et al: Nature 310:586, 1984.