Mutation Research, 212 (1989) 241–252
Elsevier

MTR 04757

Development of molecular approaches to estimating germinal mutation rates

I. Detection of insertion/deletion/rearrangement variants in the human genome

Harvey W. Mohrenweiser *, Robert D. Larsen ** and James V. Neel

Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109-0618 (U.S.A.)

(Received 8 August 1988)
(Revision received 7 December 1988)
(Accepted 18 January 1989)

Keywords: Germinal mutation rates, estimation; Restriction enzyme mapping strategy

Summary

DNA from 130 individuals was studied with up to 18 (primarily cDNA) probes for the frequency of variants in this initial experiment to determine the feasibility of this approach to screening for germinal gene mutations. This approach, a modification of the usual restriction enzyme mapping strategy, focuses on the detection of insertion/deletion/rearrangement (I/D/R) variants, because the DNA is digested with only two restriction enzymes before transfer to membranes and hybridization with an extensive series of unrelated probes. Some 4000 noncontiguous, independent DNA fragments ("loci"), functional loci, pseudogenes or anonymous fragments, (a total of ~ 77400 kb) were screened. 19 different classes and 31 copies of presumably I/D/R variants were detected while 4 different classes and 24 individuals exhibiting base substitution variants were observed. 18 of the 19 I/D/R classes were rare variants, that is, each were observed at a frequency, within this population, of less than 0.01; 3 of the base substitution classes existed at polymorphic frequencies and only 1 was a rare variant. 10 of the I/D/R classes, occurring in a total of 18 individuals, were detected with probes which are not known to be associated with repetitive elements. This is a variant frequency for I/D/R variants without known repetitive elements of 0.15 classes and 0.23 copies for each 1000 kb screened; this would extrapolate to 1600 such variant sites in the genome of each individual. Within the context of a mutation screening program, the rare variants, either with or without repetitive elements, would have a higher probability of being de novo mutations than would polymorphic variants; this former group would be the focus of family studies to test for the heritability of the allele (fragment pattern). Sufficient DNA probes are available to screen a significant portion of the human genome for genetic variation and de novo mutations of this type.

* Current address: Biomedical Sciences Division L-452, Lawrence Livermore National Laboratory, Livermore, CA 94550 (U.S.A.).

** Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI 48109 (U.S.A.).

Correspondence: Dr. H.W. Mohrenweiser, Biomedical Sciences Division, L-452, Lawrence Livermore National Laboratory, Livermore, CA 94550 (U.S.A.).

0027-5107/89/$03.50 © 1989 Elsevier Science Publishers B.V. (Biomedical Division)

Estimation of human genetic risk associated with exposure to known and suspected mutagenic agents has been limited by the scarcity of relevant human data on germinal (heritable) gene mutations (UNSCEAR, 1986). One of the impediments has been that — fortunately — the number of offspring available for study is relatively small and the parental exposure is relatively low, by experimental standards. For example, in the major evaluation to data of an exposed population, the study of the aftermath of the atomic bombings of Hiroshima and Nagasaki, the estimated number of children born to survivors receiving increased radiation in these two cities is approx. 24000 and the average conjoint gonadal radiation exposure is only 0.48 Sv (Neel et al., 1988b). The number of offspring in other studies, such as children born to individuals who received radio- or chemo-therapy for childhood malignancy (Mulvihill and Byrne, 1985) or who unsuccessfully attempted suicide (Czeizel, 1986) is even less. By contrast, the mutagen doses in typical murine radiation experiments are much higher than estimated for human exposure ($> 10 \times$ for many studies) and the number of offspring available for study is a controllable variable. Therefore, estimation of human mutation rates will require new technologies which maximize the information which can be obtained from each child of the exposed parents (Office of Technology Assessment, 1986).

The emerging DNA technologies appear to have promise in this respect (Delahanty et al., 1986), since, with the human haploid genome estimated at 3×10^9 nucleotides, it is, in principle, possible for a relatively small number of individuals to provide sufficient information to estimate possible changes in germinal mutation rates. For example, given a background nucleotide substitution mutation rate of 10^{-8} (Neel, 1983; Neel et al., 1988a; Stamatoyannapoulos and Nute, 1982) each proband will possess some 70 de novo point mutations.

The development of the new DNA based approaches can be envisioned as directed toward the detection of two different types of molecular events (Mohrenweiser and Branscomb, in press; Delahanty et al., 1986). Most of the proposed approaches have emphasized detection of nucleotide substitution mutations with methods where relatively small segments of the genome would be intensely examined, although more encompassing methods have been suggested (Delahanty et al., 1986). Obviously some molecular lesions other than nucleotide substitutions within these regions would also be detected. In contrast, the effort contemplated in this report will emphasize the detection of variation due to insertions, deletions and rearrangements (I/D/R) of DNA, the class of genetic damage that a large literature suggests is expected to be especially predominant following radiation exposure.

The development of molecular techniques has provided the tools to define the basis for many genetic diseases, each of which reflects a mutational event in a previous generation. These molecular studies of human genetic diseases (Carroll et al., 1985; den Dunnen et al., 1987; Spritz and Forget, 1983; Yang et al., 1984; Youssoufian et al., 1987), including heritable cancers (Cavenee et al., 1986; Fearon et al., 1987), de novo germinal mutations (Barsh et al., 1985; Kazazian et al., 1988; Vidaud et al., 1986; Yussoufian et al., 1988) and also the characterization of mutations in somatic cells (Ashman and Davidson, 1987) indicate that I/D/R events are the molecular basis for many genetic alterations. Yet, with the exception of the studies of rearrangements within tandem-repetitive sequences (Jefferys et al., 1985, 1986; Nakamura et al., 1987; Wong et al., 1987), limited data are available on the frequency, within human populations, of rearrangements involving or adjacent to DNA regions related to functional genes and not involving known repetitive elements. This latter group of variants may be of greater health significance than base substitutions as most of the mutations of the I/D/R class, if they involve functional loci, should result in the loss of a functional gene product. The repetitive region elements are suggested to be "hotspots" for recombinational mutations, relative to the remainder of the genome, with an estimated mutation rate, based upon very limited data, of 10^{-3}–10^{-4} (Jefferys, 1987; Kovacs et al., 1987). It is unknown at this time whether exposure to a mutagenic agent or specific class of mutagenic agents increases the mutation rate in the region of these tandem-repeat elements.

The approach employed in this study is a vari-

ation of the standard restriction enzyme mapping technique. In the usual RFLP mapping format, a DNA sample is digested with an extensive series of enzymes in an effort to detect variation (usually base substitutions) at a locus. The data reported here were obtained by digesting each sample with a single combination of two enzymes and the blots were probed with an extensive series of probes. Thus, a significant fraction of the genome was screened for insertion/deletion/rearrangement events while only a limited number of bases were scanned for substitutions. The data generated in this study are *not* intended to address the mutation rate question directly but will begin to provide answers to two question relevant to developing an efficient approach to mutation screening: (1) What is the frequency in human populations for the classes of events screened? (2) What is the potential for developing an efficient approach for screening for this class of mutational events?

Materials and methods

High molecular weight chromosomal DNA was prepared from leucocytes isolated from whole blood, cultured cells, or placental tissues using a modification of the techniques described by Maniatis et al. (1982). The DNA extraction procedure included digestion with proteinase K (Sigma Chemical), two phenol extractions, one chloroform/phenol (1 : 1 vol./vol.) extraction, two chloroform extractions, and ethanol precipitation. The final DNA pellet was resuspended in TE buffer (10 mM Tris, 1 mM EDTA; pH 8.0) and stored at 5 °C until needed for use in restriction enzyme digestions. Genomic DNA samples were digested with restriction endonucleases using conditions described by the suppliers (New England BioLabs; Bethesda Research Laboratories). For routine screening, genomic DNA samples (6 μg DNA/lane) were digested with the combination of EcoRI and BamHI.

Digested genomic DNA fragments were separated by agarose gel electrophoresis (Maniatis et al., 1982). HindIII digested λ-phase DNA fragments were used as standard molecular weight markers. DNA fragments in agarose gels were stained with ethidium bromide (Sigma Chemical), photographed, and transferred to Gene Screen Plus membranes (New England Nuclear) by the method of Southern (1975). Membranes were prehybridized at 42 °C for 6–14 h in 1 M NaCl, 40% (vol./vol.) formamide (Fluka), 0.1 M Tris (pH 7.5), 1% (wt./vol.) sodium dodecyl sulfate (SDS) (Sigma Chemicals), 5% Denhardt's (see Maniatis et al., 1982), 1 mM EDTA, and 100 μg/ml denatured sheared salmon sperm DNA (Sigma Chemicals). DNA probes were radioactively labeled by the "oligolabeling" technique of Feinberg and Vogelstein (1983) using α-32 P-dCTP (Amersham). After hybridization at 42 °C for 40–48 h, the following stringency washes were performed: (1) 2 × SSC [1 × SSC is 0.15 M NaCl, 0.015 M Na citrate (pH 7.0)], 1% SDS, at room temperature for 10 min; (2) 2 × SSC, 1% SDS at 65 °C for 30 min; and (3) 2 × SSC, 1% SDS at 65 °C for 30 min. Fragment patterns were visualized following autoradiography at −70 °C for 24–48 h with one Cronex Lightning Plus X-ray intensifying screen (DuPont). In preliminary experiments, the membranes were rewashed as in step 3, but with 0.1 × SSC plus 1% SDS, to insure the specificity of the hybridization pattern. This more stringent wash did not alter the pattern observed with any of the probes, relative to the standard washing protocol with 2 × SSC. Radioactive probes were stripped from membranes as described by the manufacturer (New England Nuclear) and the filters subsequently autoradiographed overnight as described above to insure complete probe removal. Stripped membranes were stored at −20 °C until used for additional hybridizations.

Variant DNA samples were further characterized by digestion with the following battery of restriction endonuclease: (1) EcoRI; (2) BamHI; (3) EcoRI plus BamHI; (4) HindIII; (5) PvuII; (6) SstI; and (7) XbaI. Additionally, a new DNA samples was isolated and the initial EcoRI/BamHI digestion repeated to check for completeness of the digestion, etc. as necessary to confirm the initial observation. The resolution between fragments was estimated to be sufficient to detect fragments which differ in size by ~2%, thus deletions or insertions involving less than 25–250 base pairs, depending upon the size of the original fragment, would not be detected. Each probe was used to screen DNA samples from 130 unrelated individuals except for two globin probes, γδ1.6BX

and γG IVSII, where only 51 samples were screened. The samples included 20 placental tissues, 12 somatic cell lines and 98 lymphocyte samples, the latter from students and laboratory personnel. The population was 95% Caucasian and 5% Oriential.

Results

Detection of variation. Table 1 lists the 18 probes employed in the screening of the genomic DNA samples from 130 unrelated individuals in a preliminary effort to estimate the extent of genetic variation associated withI/D/R events. Most of the probes are cDNAs and therefore derived from functional genes. These genes are localized to 9 different chromosomes. Many of the probes also hybridize to additional noncontiguous DNA fragments; these additional fragments are either duplicated genes (e.g. actin, aldolase, phosphoglycerate kinase) and/or pseudogenes (e.g. hypoxanthine phosphoribosyl transferase, triosephosphate isomerase) and often map to chromosomes other than the site of the functional gene, thus regions of almost every chromosome is represented in the screen. The number of noncontiguous DNA fragments being screened is not known exactly but, with the hybridization/washing stringency of these studies, it is estimated to be approx. 40. As both alleles are available for direct examination (except for sex-linked DNA fragments) the number of allelic fragments monitored would be 80 in females and an estimated 74 in males.

It should be noted that each genomic DNA fragment may yield several bands on a Southern

TABLE 1

LOCUS, FRAGMENT SIZE(S) IN EcoRI/BamHI DIGESTS, AND REFERENCE FOR EACH OF 18 PROBES

Locus	Probe	Fragment size(s) (kb)	kb per individual	Reference
β-actin	pA1	13.0, 11.5, 9.1 6.6, 5.9, 4.0, 2.6	52.7	Cleveland et al. (1980)
Adenosine deaminase	phADA-1	18.0, 2.3, 2.0, 0.8	23.1	Orkin et al. (1983)
Aldolase B	phL-413	5.5, 4.8	10.3	Rottmann et al. (1984)
Adenosine phosphoribosyl transferase	pT02	2.3	2.3	T. O'Toole, personal communication (1987)
Carbonic anhydrase II	H25-3.8	2.1, 1.1, 0.6	3.8	Lee et al. (1985)
Cytochrome C oxidase (subunit IV)	pCOX4.111	6.6, 6.0	12.6	Zeviani et al. (1987)
Glucose-6-phosphate dehydrogenase (related)	pGD-3	12.0	12.0	Toniolo et al. (1984)
Hypoxanthinephosphoribosyltransferase	mHPN-6	12.0, 8.5, 8.0, 5.6, 3.8 3.3, 2.2, 1.4, 0.6	45.4	Brennard et al. (1982)
Human c-*myc*	pHSR-1	14.8, 6.6, 4.4	25.8	Alitalo et al. (1983)
Phosphoglycerate kinase	phPGK-7e	8.8, 4.0, 3.6, 2.9 2.4, 2.3, 2.0	26.0	Michelson et al. (1983)
Human c-k-*ras*	p640	2.8	2.8	McCoy et al. (1983)
Triosephosphate isomerase	pHTPI-5a	8.1, 7.3, 6.5, 4.7, 4.3 3.9, 2.6, 2.0, 1.3, 1.0	41.9	Maquat et al. (1985)
α-globin	pR α-1	16.0	16.0	Higgs et al. (1986)
β-globin	pHB-1S	2.0, 1.1	3.1	Lawn et al. (1978)
γ-globin	JW-151	2.9, 2.5, 2.0	7.4	Wilson et al. (1978)
γ-globin/δ-globin intergene	γδ1.6BX	7.2	7.2	P. Henthron, personal communication (1986)
γ-globin intervening sequence II	GγIVSII	1.0	1.0	Vanin et al. (1983)
ζ-globin pseudogene	pHζp	15.5, 10.0, 2.5	*28.0*	Proudfoot et al. (1984)
Total			321.4	

blot because of internal restriction sites and that the 40 fragments are non-contiguous genome DNA segments, not the number of bands on a blot, the latter being 61 restriction fragments. Also, as differences in band intensity were not scored, a deletion which removed the totality of an allele, and therefore all of the target for a probe, would not be detected as a variant, except for the sex-linked DNA fragments. At reduced hybridization stringency, additional bands, especially for the actin probe, were observed but a conservative approach was utilized, resulting in the estimate that 40 noncontiguous DNA fragments were screened. As indicated above, no scored bands were lost when the washing stringency was increased beyond step 3. In total, these fragments account for 321 kb of DNA in the genome of the "standard" haploid chromosome set. Therefore, following an adjustment for fragments derived from the X chromosome because one half the samples were from male individuals, the average sample was screened for variation in approx. 600 kb of DNA.

As the probes hybridized to several noncontiguous fragments ("loci"), it was not always possible to establish allelic relationships between different bands. Thus, we have employed the terminology of number of *classes* for the different restriction digest variants which would be designated different alleles if they were restricted to a single locus and the number of *copies* of each variant for allele frequency.

Classification of variation. 2 groups of variants were identified (Table 2). 4 different variant classes had characteristics consistent with the existence of a base substitution, that is, variation was restricted

TABLE 2

FREQUENCY AND DISTRIBUTION OF VARIANTS

Locus	Insertion/Deletion/Rearrangement variants		Base substitution variants	
	Classes	Variants/class	Classes	Variants/class
β-actin	–	–	1	10
Adenosine deaminase	–	–	–	–
Aldolase B	–	–	–	–
Adenosine phosphoribosyl transferase	–	–	–	–
Carbonic anhydrase II	1	2	–	–
Cytochrome C oxidase (subunit IV)	–	–	–	–
Glucose-6-phosphase dehydrogenase related	–	–	–	–
Hypoxanthine phosphoribosyl transferase	–	–	1	4
Human-c-myc	2	1, 2	–	–
Phosphoglycerate kinase	–		–	
Human-c-ki-*ras*	–	–	–	–
Triosephosphate isomerase	–	–	2	8,2
α-globin	2	2, 3	–	–
β-globin	2	2, 2	–	–
γ-globin	3	2, 1, 1	–	–
γ-globin/δ-globin intergene	–	–	–	–
γg-globin intervening sequence II	–	–	–	–
ζ-globin pseudogene	9	1,1,1,1,1,2,2,2,2	–	–
Total	19	31	4	24

to only the BamHI or the EcoRI digest and was not discerned when the sample was digested with other restriction enzymes. Variant base substitution classes occuring in polymorphic frequency (defined as a frequency of a single class of greater than 0.01) were observed with the actin (10 individuals), HPRT (4 individuals) and the TPI (8 individuals) probes. Only one rare base substitution variant class was observed (TPI; 2 individuals). The HPRT and TPI variants appeared to involve a base substitution in the EcoRI recognition site as a normal pattern was observed in these DNA samples with each of the 5 other enzymes. The results of further analysis of the actin variants were not conclusive in differentiating between I/D/R variants and nucleotide substitutions as additional variation was observed in both normal and variant samples with several of the other enzymes.

There were 19 different classes of variants which had characteristics consistent with an I/D/R event. In contrast to the base substitution variants, the I/D/R variants were detectable following digestion of the sample with several different restriction enzymes. These events were detected with 6 different probes although 9 of the 19 variant classes were associated with the z-globin pseudogene probe. All of the variants occurred as only 1 or 2 copies per variant class with the exception of 3 copies for one class of variant at the α-globin locus, this being the only polymorphic variant (frequency of 0.011) in this group. There were a

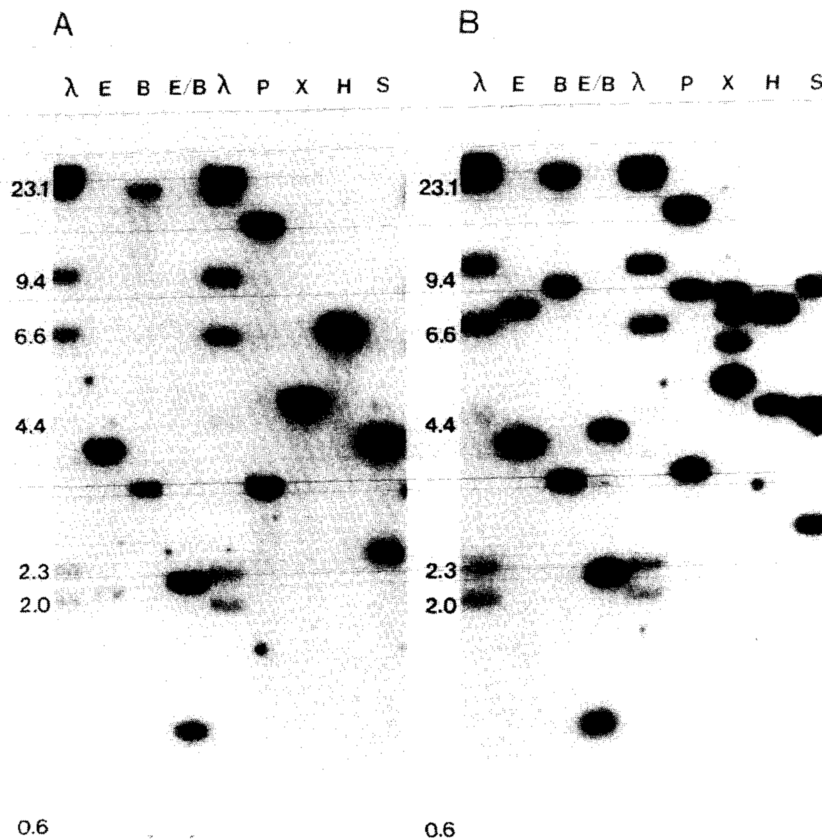

Fig. 1. Restriction enzyme pattern for an I/D/R variant at the carbonic anhydrase II locus. The common restriction enzyme pattern is in panel A while the variant pattern is in panel B. Lanes 1 and 5 in each panel are size markers (HindIII digested λ DNA). The remaining lanes are DNA from normal and variant individuals digested with the indicated enzymes. The enzymes are: E, EcoRI; B, BamHI; P, PstI; X, Xba, H, HindIII; S, SstI.

total of 31 I/D/R variants detected in the study. It should be noted that this is a minimal estimate, in that alterations involving a change in size of less than ~ 2% of any fragment would probably not be detected.

An example of the pattern of an I/D/R variant detected with a carbonic anhydrase II (CAII) probe, H25-3.8, is seen in Fig. 1. The normal pattern is shown in panel A while the variant, presumably a heterozygote individual as all of the bands corresponding to the normal pattern are also observed, is in Panel B. This hypothesis, that the new band(s) observed following digestion of the DNA with a series of restriction enzymes are variant bands in a heterozygous individual, is supported by the observation than the new band, especially in the EcoRI lane, is approximately half the intensity of the normal band. The existence of an altered pattern with each of six enzymes would be consistent with an I/D/R event rather than a base substitution. (It is assumed that only a single event is responsible for the altered pattern.) The variant bands should increase by a constant size if a small insertion, without an additional restriction site, is responsible for the variation. The existence of an internal restriction site in the insert would yield a variant pattern with two new fragments. Neither expectation was observed with this variant as the new EcoRI band is increased in size by approx. 3.1 kb while the new BamHI fragment is approx. 4.5 kb larger than any of the bands corresponding to bands observed in the comparable digest in Panel A. More importantly, based upon the pattern observed and the tentative restriction map of this region (Lee et al., 1985), a deletion is a more likely cause for the new bands. From a comparison of the fragment patterns, it is probable that at least one of the first two exons of the CAII locus has been deleted and it is possible that the deletion could involve 15–20 kb of DNA (P. Venta, personal communication). With the present data, it is not possible to determine if the deletion extends upstream or downstream from the first two exons. This variant pattern was observed in two samples, one a transformed lymphoblast cell line and the other, a lymphocyte sample from an unrelated individual. These individuals should be heterozygous for CAII deficiency, given that the presumed deletion involves one or two exons.

Estimation of genetic variation. The frequency with which rare variants exist (presumably inherited) can be expressed in several different manner. Assuming that the 18 probes surveyed 40 noncontiguous DNA fragments per haploid genome, then approx. 9950 independent or noncontiguous DNA fragments were screened (40 fragments × 2 "alleles" (except for sex-linked fragments) × 130 samples). Thus 1.9 different I/D/R variant classes and 3.1 total variants were identified for each 1000 fragments screened. If the data from the highly variable repetitive sequence detected with the z-globin pseudogene probe are excluded, the frequencies become 1.0 variant class and 1.8 variants per 1000 noncontiguous fragments screened. In each individual, approx. 600 kb of DNA were screened for variation (555 kb in males and 642 kb in females), thus, some 77400 kb of DNA were screened in this survey. I/D/R variant classes were detected with a frequency of 0.23 classes per 1000 kb screened and 0.40 total variants were observed per 1000 kb of DNA screened (or 0.23 variants if the globin pseudogene variants are excluded).

The frequency of different rare base substitution variant classes was 0.1 per 1000 noncontiguous fragments while 0.2 total rare variants per 1000 noncontiguous fragments were detected. The 3 base substitution variants occurring in polymorphic frequency are not included in this latter calculation but their inclusion would increase the numbers to 0.4 classes and 2.4 total variants/1000 fragments. Rare base substitution variants occurred with a frequency of 0.012 classes and 0.082 total rare variants per 1000 kb screened. It should be noted that in reality only some 230 sites (4 for each band on an average Southern blot; 61 bands × 2 ends × 2 alleles (minus the number of X chromosome bands) or 1380 bases (as both enzymes have 6-base recognition sites) were being screened in each sample for base substitutions associated with loss of a restriction site. In addition, one half that number of nucleotides (690), would be screened for base substitutions associated with the potential gain of a restriction site in each individual (rather than the total of 600 kb). Therefore, the rare base substitution variant frequency is 0.0035 (1 variant class/(2070 bases × 130 individuals)) classes and 0.007 [2 variant

copies/(2070 bases × 130 individuals)] total rare variants per 1000 bases actually screened. The total (rare plus polymorphic) base substitution variant frequency is 0.015 classes and 0.09 variant copies per 1000 nucleotides screened.

Discussion

Since family studies were not performed in this exploratory investigation, the genetic basis for the observed variation was not confirmed. Several points, however, would suggest that the variants have a genetic basis comparable to that observed in many other examples of these types of variation. Each variant pattern was confirmed when DNA was extracted and reanalyzed from a new sample. With the exception of genes on the X chromosome in males, a restriction pattern consistent with codominant inheritance, that is, only the gain of bands, was observed. The variants were distributed among the samples as expected for rare variants, rather than clustered in a few individuals as might be expected for variation of a nongenetic nature. Obviously, an important next step is family studies to confirm the genetic nature and heritability of these variants. Such family studies would be necessary before any of the detected variation could possibly be considered to be due to de novo germinal mutation.

The significant degree of variation involving insertions, deletions and/or rearrangements of DNA in the human genome is becoming increasingly apparent from two types of studies. First, the studies of Jefferys et al. (Jefferys et al., 1986; Wong et al., 1987) and White and colleagues (Nakamura et al., 1987), among others, have identified significant variation among individuals associated with differences in the number of tandem repeat elements in an array in a DNA fragment identifiable by probing genomic DNA with probes which include various "core" sequences. It should be noted that in the current survey, approx. 50% of the I/D/R variants were detected with the z-globin pseudogene probe. The z-globin gene is known to be the site of a tandem-repetitive core sequence and to exhibit extensive allelic variation associated with the existence of variable numbers of tandem-repeats (Proudfoot et al., 1984). None of the other I/D/R variant classes are associated with known tandem-repeat series or other types of structural features which, with a high frequency, alter the relative spacing of a series of restriction enzyme sites. It is obvious that the other probes detect neither the number of variant classes nor variant individuals that are observed with the z-globin pseudogene probe. This class of variation is suggested to be generated via recombinational mechanism(s). The mutations/variants in these regions are probably not of physiological significance and therefore of somewhat limited usefulness for estimating disease burden. Given the apparent absence of functional significance, these segments presumably involve flanking, inter-gene or at least neither coding nor regulatory regions.

Additional effort will be necessary to determine if the alterations detected with the other probes, be they insertions, deletions or rearrangements, involve the coding, intron or flanking regions for any of the other genes or if they are generally restricted to the regions between functional genes or for example, involve the pseudogenes rather than the functional loci.

The second type of study has focused on characterizing genes of functional significance. For example, several heritable tumors are associated with a predisposing event involving the apparent absence of one allele at a locus (Cavenee et al., 1986). Additionally, molecular studies of a series of genetic diseases have ascertained several diseases associated with I/D/R events (Antonarakis et al., 1985; den Dunnen et al., 1987; Spritz and Forget, 1983; Yang et al., 1984) and a number of functionally significant de novo mutations have been identified where the molecular lesion is an insertion, deletion or rearrangement of DNA (Barsh et al., 1985; Kazazian et al., 1988; Vidaud et al., 1986; Yang et al., 1984; Youssoufian et al., 1988).

Previous surveys for rare genetic variation and potential mutation have generally screened plasma and erythrocyte proteins for electrophoretic mobility variants (Mohrenweiser et al., 1987; Neel et al., 1988a,b). Some more limited studies have screened for variation associated with the absence of a functional gene product as detected by a 50% reduction in erythrocyte enzyme activity (Mohrenweiser, 1987; Satoh et al., 1983). The electrophoretic mobility variant class would almost always

result from the base substitutions detected in this and other DNA screening efforts, specifically nucleotide substitutions which alter a codon and results in the interchange of amino acids of non-identical charge. The null variants could result from deletions/insertions as well as rearrangements which disrupt the normal coding sequence and certain base substitutions (Mohrenweiser, 1983). The frequency of rare electrophoretic mobility variants in a Caucasian sample expected to be similar to that of this study was 2.0 heterozygotes/1000 loci screened (range of 1.3–2.2) (Mohrenweiser et al., 1987) while the frequency of null variant heterozygotes was 3.1/1000 loci (Mohrenweiser, 1987). Assuming an average protein is encoded by a mRNA of 1000 nucleotides, the estimated rare base substitution variant frequency, extrapolated from the frequency of electrophoretic mobility variants would be 0.003 variants per 1000 nucleotides (2 variants times 3 additional nucleotide substitutions not expressed as altered mobility of the protein (Neel, 1984)/(2 alleles × 1000 nucleotides). This calculated frequency of nucleotide substitutions is in good agreement with the estimate derived from the limited comparable data of this study (0.007 rare base substitutions/1000 nucleotides) although the comparison is obviously based upon many assumptions, including comparability of sequences monitored and/or equal average variant frequencies throughout the genome. The total base substitution variant frequency (0.09 variants/1000 nucleotides) identified with these restriction enzymes and probes (predominantly cDNA probes) is consistent with the frequency usually observed with the enzymes BamH1 and EcoR1 (Feder et al., 1985) although it is obviously less than the variation proposed as the average frequency of base substitutions in the genome (Feder et al., 1985; Cooper and Schmidte, 1984). Neither the BamHI nor EcoRI recognition site includes the CpG dinucleotide which is proposed to account for the high frequency of polymorphic sites detected with TaqI and MspI, the two enzymes which contribute most significantly to the estimated level of nucleotide variation derived primarily from restriction enzyme mapping studies (Feder et al., 1985; Cooper and Schmidte, 1984).

The comparison of the relative frequency of null variants detected by loss of functional gene product with the direct DNA analysis data is much more tenuous as null variants can be associated with a spectrum of molecular lesions including base substitutions as well as I/D/R events. Also, it is obviously not clear what fraction of the I/D/R variants have functional relevance at the level of the gene product. But, assuming that 20–50% of the null variants are associated with I/D/R events which could be detected by the restriction enzyme mapping techniques of this study (Yang et al., 1985; Youssoufian et al., 1987) then the predicted frequency of detectable I/D/R alleles, based upon the measured enzyme deficiency allele frequency, would be 0.5 alleles/1000 fragments [3.1 deficiency variants × 30% due to I/D/R variants/(1000 loci × 2 alleles)]. Assuming that the "average" enzyme structural locus is equivalent to the "average" noncontiguous fragment monitored in this screen, then the frequency of detected I/D/R variants not known to be associated with repetitive elements, which is 1.8 alleles/1000 fragments, is in satisfactory agreement with the value of 0.5 alleles/1000 fragments, "estimated" for I/D/R variants by extrapolating from the null variant frequency, with the assumptions stated above. The discrepancy between the estimates would be reduced if a disproportionate fraction of the I/D/R variants (in addition to the repetitive sequence) did not involve the coding region for a functional gene, which would seem to be a reasonable assumption.

The data obtained in this study have implications for both the development of germinal mutation monitoring protocols as well as efforts to sequence the human genome. For example, given the human haploid genome is 3×10^9 base pairs and the frequency of I/D/R variants is 0.23 variants/1000 kb, then any chromosome set will contain an average of 1600 insertions, deletions or rearrangements involving unique sequence regions of the genome. Additional variation associated with the tandem-repetitive elements is even more prevalent (Jefferys et al., 1986; Nakamura et al., 1987; Wong et al., 1987).

A study of mutation based on I/D/R events appears feasible although the data are not yet available for addressing several questions, including the best analytical strategy, the sensitivity of

the approach, and the cost of a monitoring effort. Given the large number of probes becoming available from effort to identify functional genes and also efforts to construct genetic maps, a study of mutation based on I/D/R variants will not be limited by the availability of probes. For example, 200 probes at 30 kb scanned per probe would screen 0.2% of the genome for mutations and the number of probes could be easily increased, although development of an efficient screening program will require that new methods for RFLP analysis continue to be developed. But, assuming that 6000 noncontiguous fragments/proband can be screened and the detectable background mutation frequency for I/D/R events is 10^{-5}/noncontiguous fragment, (a mutation rate similar to that estimated from sentinel phenotype data (Mulvihill and Czeizel, 1983) and 10–100-fold less than estimated with the tandem-repeat probes (Jefferys, 1987; Kovacs et al., 1987) then 6 mutations would be detected in 1000 probands. From this information, the sample sizes necessary to demonstrate (or exclude) various increases in the mutation rate can be readily calculated. In addition, several other DNA-based techniques have been proposed for detecting nucleotide substitutions (Office of Technology Assessment, 1986) and two-dimensional gel electrophoresis techniques for studying cellular proteins continue to be developed (Neel et al., 1984). In point of fact, an effective screening effort for estimating the total human germinal gene mutation rate will necessitate the inclusion of other techniques which focus more on base substitution events, in addition to a I/D/R screening effort. But, the DNA-based methodologies have an inherent advantage over other approaches in that, potentially, all of the genome is available for screening, and if this potential can be exploited, then each proband yields a significant data base. A combined approach could make efforts to estimate human germinal mutation rates a reality.

Acknowledgments

The assistance of K.H. Wurzinger, S. Fielek and S. Dembinski is acknowledged as is the cooperation of the many investigators who facilitated this effort by making the probes available. This work was supported by U.S. Department of Energy contract DE-AC02-82ER-60089.

References

Alitalo, K., M. Schwab, C.C. Lin, H.E. Varmus and J.M. Bishop (1983) Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma, Proc. Natl. Acad. Sci. (U.S.A.), 80, 1707–1711.

Antonarakis, S.E., H.H. Kazazian Jr. and S.H. Orkin (1985) DNA polymorphism and molecular pathology of the human globin gene clusters, Hum. Genet., 69, 1–14.

Ashman, C.R., and R.L. Davidson (1987) Sequence analyses of spontaneous mutations in a shuttle vector gene integrated into mammalian chromosomal DNA, Proc. Natl. Acad. Sci. (U.S.A.), 84, 3354–3358.

Barsh, G.S., C.L. Roush, J. Bonadio, P.H. Byers and R.E. Gelinas (1985) Intron-mediated recombination may cause a deletion in an α I type I collagen chain in a lethal form of osteogenesis imperfecta, Proc. Natl. Acad. Sci. (U.S.A.), 82, 2870–2874.

Brennand, J., A.C. Chinault, D.S. Konecki, D.W. Melton and C.T. Caskey (1982) Cloned cDNA sequences of the hypoxanthine/guanine phosphoribosyltransferase gene from a mouse neuroblastoma cell line found to have amplified genomic sequences, Proc. Natl. Acad. Sci. (U.S.A.), 79, 1950–1964.

Carroll, M.C., A. Palsdottir, K.T. Belt and R.R. Porter (1985) Deletion of complement C4 and steroid 21-hydroxylase genes in the HLA class III region, EMBO J., 4, 2547–2552.

Cavenee, W.K., A. Koufas and M.F. Hansen (1986) Recessive mutant genes predisposing to human cancer, Mutation Res., 168, 3–14.

Cleveland, D.W., M.A. Lopata, R.J. MacDonald, N.J. Cowan, W.J. Rutter and M.W. Kirschner (1980) Number and evolutionary conservation of α- and β-tubulin and cytoplasmic β- and γ-actin genes using specific cloned cDNa probes, Cell, 20, 95–105.

Cooper, D.N., and J. Schmidtke (1984) DNA restriction fragment length polymorphisms and heterozygosity in the human genome, Hum. Genet., 66, 1–16.

Czeizel, A. (1986) Self-poisoning as a model for the study of mutagenicity and teratogenicity of chemicals in human beings, Prog. Clin. Biol. Res., 209B, 237–244.

Delehanty, J., R.L. White and M.L. Mendelsohn (1986) Approaches to determining mutation rates in human DNA, Mutation Res., 167, 215–232.

den Dunnen, J.T., E. Bakker, G.K. Breteler, P.L. Pearson and G.J.B. van Ommen (1987) Direct detection of more than 50% of the Duchenne muscular mutations by field inversion gels, Nature (London) 329, 640–642.

Fearon, E.R., S.R. Hamilton and B. Vogelstein (1987) Clonal analysis of human colorectal tumors, Science, 238, 193–197.

Feder, J., L. Yen, E. Wysman, L. Wang, L. Wilkins, J. Schroder, N. Spurr, H. Cann, M. Blumenberg and L.L. Cavalli-Sforza

(1985) A systematic approach for detecting high frequency restriction fragment length polymorphisms using large genomic probes, Am. J. Hum. Genet., 37, 635–649.

Feinberg, A.P., and B. Vogelstein (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity, Anal. Biochem., 132, 6–13.

Higgs, D.R., J.S. Wainscoat, J. Flint, A.V.S. Hill, S.L. Thein, R.D. Nichols, H. Teal, H. Ayyub, T.E.A. Peto, A.G. Falusi, A.P. Jarman, J.B. Clegg and D.J. Weatherall (1986) Analysis of the human α-globin gene cluster reveals a highly informative genetic locus, Proc. Natl. Acad. Sci. (U.S.A.), 83, 5165–5169.

Jefferys, A.J. (1987) Highly variable minisatellites and DNA fingerprinting. Biochem. Soc. Trans., 15, 309–317.

Jefferys, A.J., V. Wilson and S.L. Thein (1985) Hypervariable "minisatellite" regions in human DNA, Nature (London), 314, 67–73.

Jefferys, A.J., V. Wilson, S.L. Thein, D.J. Weatherall and A.J. Ponder (1986) DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees, Am. J. Hum. Genet., 39, 11–24.

Kazazian, Jr. H.H., C. Wong, H. Youssoufian, A.F. Scott, D.G. Phillips and S.E. Antonarakis (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man, Nature (London) 332, 164–166.

Kovacs, B.W., F.O. Sarinana, R.B. Wallace and D.E. Comings (1987) Spontaneous germinal mutations in hypervariable regions of the human genome, Am. J. Hum. Genet., 41, A223.

Lawn, R.M., E.F. Fritsch, R.C. Parker, G. Blake and T. Maniatis (1978) The isolation and characterization of linked δ- and β-globin genes from a cloned library of human DNA, Cell, 15, 1157–1174.

Lee, B.L., P.J. Venta and R.E. Tashian (1985) DNA polymorphism in the 5′ flanking region of the human carbonic anhydrase II gene on chromosome 8, Hum. Genet., 69, 337–339.

Maniatis, T., E.F. Fritsch and J. Sambrook (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory.

Maquat, L.E., R. Chilcote and P.E. Ryan (1985) Human triosephosphate isomerase cDNA and protein structure, J. Biol. Chem., 260, 3748–3753.

McCoy, M.S., J.J. Toole, J.M. Cunningham, E.H. Chang, D.R. Lowy and R.A. Weinberg (9183) Characterization of a human colon/lung carcinoma oncogene, Nature (London), 302, 79–81.

Michelson, A.M., A.F. Markham and S.H. Orkin (1983) Isolation and DNA sequence of a full-length cDNA clone for human X chromosome encoded phosphoglycerate kinase, Proc. Natl. Acad. Sci. (U.S.A.), 80, 472–476.

Mohrenweiser, H.W. (1983) Enzyme deficiency variants: frequency and potential significance in human populations, isozymes, Curr. Top. Biol. Med. Res., 10, 51–68.

Mohrenweiser, H.W. (1987) Functional hemizygosity in the human genome: Direct estimate from twelve erythrocyte enzyme loci, Hum. Genet., 77, 241–245.

Mohrenweiser, H.W., and E. Branscomb (1989) Molecular approaches to the detection of germinal mutations in mammalian organisms, including man, New Trends in Genetic Risk Assessment, in press.

Mohrenweiser, H.W., K.H. Wurzinger and J.V. Neel (1987) Frequency and distribution of rare electrophoretic mobility variants in a human population, Ann. Hum. Genet., 51, 303–316.

Mulvihill, J.J., and J. Byrne (1985) Offspring of long-time survivors of childhood cancer, Childhood cancer: Late effects, Clin. Oncol., 4, 333–343.

Mulvihill, J.J., and P. Czeizel (1983) Perspectives in mutation epidemiology 6: A 1983 view of sentinel phenotypes, Mutation Res., 123, 345–361.

Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fukimoto, M. Hoff, E. Kumlin and R. White (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping, Science, 235, 1616–1622.

Neel, J.V. (1983) Frequency of spontaneous and induced "point" mutations in higher eukaryotes, J. Hered., 74, 2–15.

Neel, J.V. (1984) A revised estimate of the amount of genetic variation in human proteins: Implications for the distribution of DNA polymorphisms, Am. J. Hum. Genet., 36, 1135–1148.

Neel, J.V., B.B. Rosenblum, C.F. Sing, M.M. Skolnick, S.M. Hanash and S. Steinberg (1984) Adapting two-dimensional gel electrophoresis to the study of human germ-line mutation rates, in: J.E. Celis and R. Bravo (Eds.), Two-dimensional Gel Electrophoresis of Proteins, Academic Press, New York, pp. 259–306.

Neel, J.V., C. Satoh, K. Goriki, J. Asakawa, M. Fujita, N. Takahashi, T. Kageoka and T. Hazana (1988a) Search for mutations altering protein charge and/or function in children of atomic bomb survivors: Final report, Am. J. Hum. Genet., 42, 663–676.

Neel, J.V., H.W. Mohrenweiser and H. Gershowitz (1988b) A pilot study of the use of placental cord blood samples in monitoring for mutational events, Mutation Res., 204, 265–377.

Office of Technology Assessment (1986) Technologies for detecting heritable mutations in human beings, OTA-H-298, U.S. Government Printing Office, Washington DC.

Orkin, S.H., P.E. Daddona, D.S. Sheewach, A.F. Markham, G.A. Bruns, S.C. Goff and W.N. Kelley (1983) Molecular cloning of human adenosine deaminase gene sequences, J. Biol. Chem., 258, 12753–12756.

Proudfoot, N.J., A. Gil and T. Maniatis (1984) The structure of the human ζ-globin gene and a closely linked, nearly identical pseudogene, Cell, 31, 553–563.

Rottmann, W.H., D.R. Tolan and E.E. Penhoet (1984) Complete amino acid sequence for human aldolase B derived from cDNA and genomic clones, Proc. Natl. Acad. Sci. (U.S.A.), 81, 2738–2742.

Satoh, C., J.V. Neel, A. Yamashita, K. Goriki, M. Frijita and H.B. Hamilton (1983) The frequency among Japanese of heterozygotes for deficiency variants of 11 enzymes, Am. J. Hum. Genet., 35, 656–674.

Southern, E. (1975) Detection of specific sequences among

DNA fragments separated by gel electrophoresis, J. Mol. Biol., 98, 503–517.

Spritz, R.A., and B.G. Forget (1983) The thalassemias: Molecular mechanisms of human genetic disease, Am. J. Hum. Genet., 35, 333–361.

Stamatoyannapoulos, G., and P.E. Nute (1982) De novo mutations producing unstable HbS and HbM II: Direct estimates of minimum nucleotide mutation rates in man, Hum. Genet., 60, 181–188.

Toniolo, D., M. d'Urso, G. Martini, M. Persico, V. Tufano, G. Battistuzzi and L. Luzzatto (1984) Specific methylation pattern at the 3′ end of the human housekeeping gene for glucose 6-phosphate dehydrogenase, EMBO J., 3, 1987–1995.

UNSCEAR (1986) Genetic and somatic effect of ionizing radiation, Report of the United Nations Scientific Committee on the Effects of Atomic Radiation, United Nation sales publication E.86.1X9, New York.

Vanin, E.F., P.S. Henthorn, D. Kioussis, F. Grosveld and O. Smithies (1983) Unexpected relationships between four large deletions in the human β-globin gene cluster, Cell, 35, 701–709.

Vidaud, M., C. Chabret, C. Gazengel, L. Grunebaum, J.P. Cazenavè and M. Goossens (1986) A de novo intragenic deletion of the potential EGF domain of the factor IX gene in a family with severe hemophilia B, Blood, 68, 961–963.

Wilson, J.T., L.B. Wilson, J.K. deRiel, L. Villa-Komaroff, A. Efstratiadis, B.G. Forget and S.M. Weissman (1978) Insertion of synthetic copies of human globin genes into bacterial plasmids, Nucl. Acid. Res., 5, 564b1581.

Wong, Z., V. Wilson, I. Patel. S. Povey and J. Jeffreys (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA, Annu. Hum. Genet., 51, 269–288.

Yang, T.P., P.I. Patel, A.C. Chinault, J.T. Stout, L.G. Jackson, B.M. Hildebrand and C.T. Caskey (1984) Molecular evidence for new mutation at the hprt locus in Lesch–Nyhan patients, Nature (London), 310, 412–414.

Youssoufian, H., S.E. Antonarakis, S. Aronis, G. Tsiftis, D.G. Phillips and H.H. Kazazian Jr. (9187a) Characterization of five partial deletions of the factor VIII gene, Proc. Natl. Acad. Sci. (U.S.A.), 84, 3772–3776.

Youssoufian, H., S.E. Antonarakis, C.K. Kasper, D.G. Phillips and H.H. Kazazian Jr. (1987b) The spectrum and origin of mutations in hemophilia A, Am. J. Hum. Genet., 41, A249.

Zeviani, M., M. Nakagawa, J. Herbert, M.I. Lomax, L.I. Grossman, A.A. Sherbany, A.F. Miranda, S. DiMauro and E.A. Schon (1987) Isolation of a cDNA clone encoding subunit IV of human cytochrome oxidase, Gene, 55, 205–217.