

2

RECEIVED BY OSTI JUL 16 1985  
CONF-ES06137--5

UCRL-92242  
PREPRINT

PROSPECTS FOR DNA METHODS TO MEASURE HUMAN  
HERITABLE MUTATION RATES

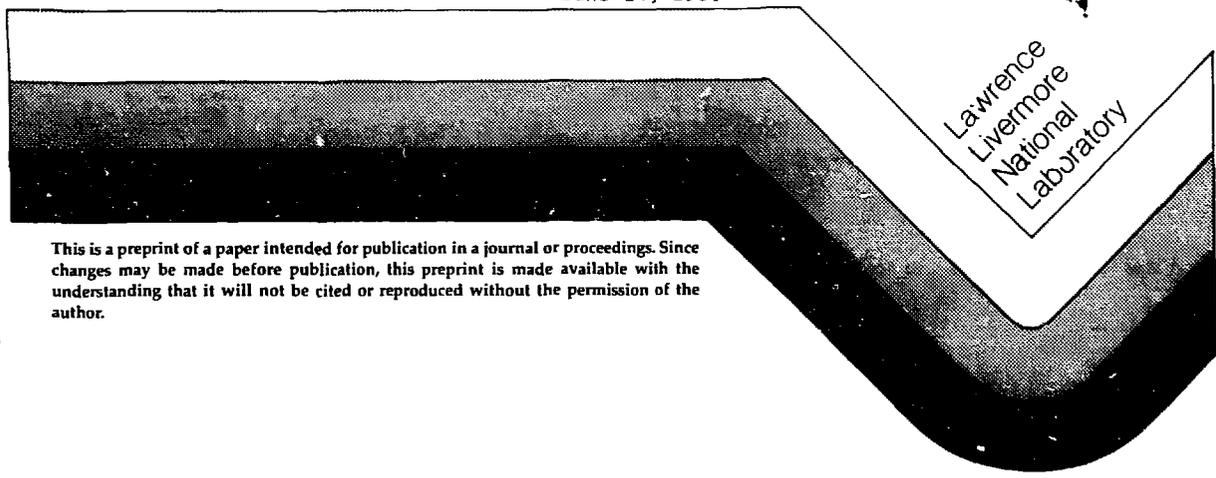
MORTIMER L. MENDELSON

BIOMEDICAL SCIENCES DIVISION

THIS PAPER WAS PREPARED FOR SUBMITTAL TO:  
"The Fourth International Conference on  
Environmental Mutagens"  
Stockholm, Sweden, June 24-28, 1985

June 14, 1985

11/14/85



This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

UCRL--92242

DE85 014555

### PROSPECTS FOR DNA METHODS TO MEASURE HUMAN HERITABLE MUTATION RATES

Mortimer L. Mendelsohn

Biomedical Sciences Division, Lawrence Livermore

National Laboratory, Livermore, California, 94550

### INTRODUCTION

A workshop cosponsored by ICPEMC and the US Dept. of Energy was held in Alta, Utah, USA Dec. 9-13, 1984 to examine the extent to which DNA-oriented methods might provide new approaches to the important but intractable problem of measuring mutation rates in control and exposed human populations. The definitive report of the workshop (Delehanty et al, 1986) is in the final stages of preparation. Meanwhile, I write this brief status report as a personal account of the sense of what took place. The workshop contributors were: D. Botstein, Massachusetts Inst. Technology; E. Branscomb, Lawrence Livermore Nat. Lab.; C. Cantor, Columbia U.; T. Caskey, Baylor Coll. of Med.; G. Church, U. Cal., San Francisco; J. Delehanty, editor, Burroughs Wellcome Co; L. Lerman, Genetics Inst.; M. Mendelsohn, administrative organizer, Lawrence Livermore Nat. Lab.; J. Mulvihill, Nat. Cancer Inst.; R. Myers, Harvard U.; J. Neel, U. Michigan; M. Olson, Wash. U.; E. Southern, Med. Res. Council, Edinburgh; S. Weissman, Yale U.; and R. White, scientific organizer, U. Utah.

### BACKGROUND

The mutation rate of human populations, control or otherwise, has been a much sought after but elusive property. It is of theoretical interest to those studying evolution, comparative biology and population genetics, and of highly practical interest to environmental mutagenicists and others concerned with the protection of the human genome and the minimization of induced genetic illness. The

complexity and variability of the human genome, the outbred and experimentally uncontrolled behavior of human breeding, and the rarity of induced mutations conspire to make the estimation of human mutation rate a very approximate and frustrating exercise. Mutation rates in other species are more tractable and better understood, although the depth of our knowledge tends to decrease (from single-cell to insect to mammal) with proximity to the human. Somatic cell mutation is another useful approximation, but there are important biological differences between somatic and germ cells as well as major inconsistencies between their mutabilities (Russell, this volume).

The traditional epidemiologic clinical approaches to human mutation (ICPEMC, 1983) range from case-reporting of sporadic genetic abnormalities to systematic screening of populations for sentinel phenotypes (Mulvihill and Czeizel, 1983). New mutants appear per locus at rates on the order of one per million births. Even with the use of 20 or so sentinel phenotypes, the recording of sufficient events to estimate a change of rate would require huge numbers of exposed parents, many years and much careful clinical genetic screening. Such methods are unsuitable for small populations, and even when applied to large samples such as the children of a-bomb (Schull et al, 1981) or cancer therapy (Mulvihill, presented at Workshop) survivors, have not yielded sufficient results to differ significantly.

Induction of somatic chromosomal aberrations has been demonstrated repeatedly in exposed human populations, however the corresponding attempts to find induced aberrations in the children of exposed parents have been consistently unsuccessful.

The detection of induced protein variants (electromorphs) is the most recent method to be applied on a large-scale, and is best exemplified by Neel and colleagues (Schull et al, 1981; more recent data presented by Neel at the Workshop) in their efforts to find effects in the children of a-bomb survivors. Using one-dimensional gel electrophoresis for thirty red cell enzymes, they have surveyed 582,268 loci in 12,242 children of parents receiving an average gonadal dose of 43 rads of mixed gamma and neutron radiation, and have compared them to a similarly-sized control population receiving less than 1 rad. Six mutations have been found, three in controls and three in exposed, giving essentially identical frequencies and an

overall mutation rate for electromorphs of  $0.58 \times 10^{-5}$ .

For populations such as the a-bomb survivors, residual options are to extend the protein method to two-dimensional gel electrophoresis, to extend the methodology to the study of DNA, or to store appropriate samples until better methods are available. The Alta workshop was convened to explore specifically the potential for new DNA methods to measure heritable mutations.

#### DNA METHODS

RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP). David Botstein reviewed the status of this method which is now widely used in human linkage and other genetic studies (Botstein et al, 1980). For the basic sampling triad of a child and its two parents, the DNA of each is separately limit-digested by a restriction enzyme, electrophoresed on a sizing gel, Southern blotted, and marked with P32 labeled probe. The appearance in the child's DNA of a unique fragment length (unexplained by polymorphism or malpaternity) constitutes a mutation. Each probe and six-cutter enzyme examines 12 bases for loss of cutting site and on average another six bases for gain of cutting site per haploid genome. This constitutes 36 bases per diploid genome and when multiplied by 278 probes, 10 enzymes and 1000 children can test  $10^8$  base-pairs. Such an effort would require more than 20 technician years of effort and given a background rate of  $10^{-8}$  base-pair mutations per generation would detect one mutant.

The typical fragment size with a six-cutter is 4000 basepairs and has a size resolution of about 2% (80 base-pairs). Frame-shifts and deletions smaller than 80 base-pairs would go undetected unless they involve restriction sites directly. Larger DNA rearrangements would only be seen if they change the separation of the restriction sites (ie delete or add a significant amount of DNA within the fragment, or transfer one of the sites to a new location).

Lack of uniqueness of the child's DNA with ten highly polymorphic probes should reduce the probability of malpaternity to several orders of magnitude below the basepair mutation rate. Non-polymorphic probes are technically easier to use with RFLP, but might yield a lower mutation rate than polymorphic probes. Sufficient probes and enzymes are or will soon be available. RFLP has not

been used for anything of this magnitude, but should scale linearly and is sufficiently understood to not require confirmation in a model animal system.

**DIRECT SEQUENCING** George Church surveyed the current feasibility and cost of direct sequencing of genomic DNA using a recent modification of the Maxim-Gilbert approach (Church and Gilbert, 1984). The DNA from each member of the family triad is limit-digested with a single restriction enzyme. It is then partially cleaved with the four or five parallel base-specific hydrolytic reactions, and electrophoresed in the corresponding number of gel lanes. The fragments are transferred and cross-linked to nylon membrane, and are serially tested by annealing with P32 labeled probes. The probes are to the four strand-ending restriction sites at any available, unique sequence regions of the genome. For each probe and set of lanes, the sequence of autoradiographic bands translates into base sequences. Child to parent differences would be mutations. The efficiency of the method hinges on the direct resolution of base sequences and on the nylon-membrane method of re-utilizing the genomic DNA many times. Analysis of E. Coli is the current scale of this method, making uncertain the efficiency and feasibility of the method for human use. If successful, direct sequencing would, of course, give the definitive details of the base structure of mutational change. It would be silent about rearrangements encompassing analyzed fragments, and about any regions which lacked the appropriate probes. It would probably take greater than 200 technician years to analyze  $10^8$  basepairs.

#### **GRADIENT DENATURATION GELS AND HETERODUPLEX DNA.**

Leonard Lerman described how electrophoresis of DNA into increasing gradients of formamide or urea can be used to detect base-mismatches in double-stranded DNA (Lerman et al, 1984). Melting of DNA by temperature or in such a gradient occurs in steps as one domain in the DNA succumbs after another. In the gradient denaturation gel, detectable bands are produced as each domain melts and retards the progress of the DNA. All but the most resistant domains can be examined in this way in a double-stranded DNA fragment, and even the most resistant domain can be seen by adding a GC-rich, late-melting, terminal sequence or a hairpin piece of DNA to the fragment. Base-pair mismatches dramatically reduce the melting temperature.

**One-Dimensional Gel Application.** Rick Myers described

how a gradient denaturing gel can be used in a practical method to find base-pair mismatches (Myers et al, 1985). Each of the triad genomes is divided into fragments of 400 base-pairs or less. A P32 labelled, 400 base-pair, unique sequence probe is added in excess to each set of DNA. The mixes are melted and reannealed to form heteroduplexes, the uneven ends are trimmed with S1 nuclease, and the fragments are ligated to a looped-end 19-mer. The triad of heteroduplexes are then run in parallel on gradient denaturing gels, looking for instances where one of the child's heteroduplexes melts at a different position than any of the heteroduplexes of the parents. Since each probe examines 400 bases per haploid genome, the combination of 125 probes and 1000 children would examine  $10^8$  base-pairs. Such an effort is estimated to require 20 technician years. This method would not detect DNA rearrangements, has only been used in model experiments, and requires major scale-up.

**Two-Dimensional Gel Application.** Lerman introduced another possible way that gradient denaturation might identify mutational changes in DNA. Two mixtures are run in parallel: one containing the DNAs of both parents; the other containing the DNAs of all three members of the triad. Each mixture is limit-digested with an appropriate restriction endonuclease. The fragments are separated by length and transferred to the horizontal starting lane of a two-dimensional gel and then electrophoresed vertically down the gradient. The DNA fragments from two or three genomes are thus distributed by size and earliest melting domain over the two dimensions. A thin horizontal slice cut from the gel would then represent iso-melting DNA across a range of fragment sizes. Such a slice with the size discrimination preserved would be melted, reannealed and reapplied to the origin of a fresh gel where it would again be put through gradient electrophoresis. Since the fragments have had the opportunity to become heteroduplexed across the four or six involved haploid genomes, those that are perfect matches should melt and arrest at the same position the slice was taken from. Those that contain a single polymorphic base-pair difference should end up in a region immediately above the slice location. Those that contain two base-pair changes (double polymorphisms) will be in an even broader band above and overlapping the single polymorphisms, and so on. At some point in this progression, probably in upper region of the double events, the density of DNAs will become sufficiently sparse to allow the resolution of bands. Unmutated DNA fragments from the child should heteroduplex

to form the same spots as parental DNA alone, whereas fragments with single base-pair changes due to mutation should behave as if additionally polymorphic and hence shift away from the original slice location and into regions where they might form a new spot. Very likely it will require the chance interaction of one polymorphism and one mutation to produce such an effect, and it is this event that the method is designed to detect. The parallel processing of several percent of the triad DNA in each slice is the major virtue of this method; its major deficiency is the requirement for chance association of mutation and polymorphism with the resulting inefficiency and possible bias toward hypermutable parts of the genome. Any mutational events larger than isolated basechanges will alter the melting behavior too dramatically to be useful. Estimates of efficiency and demonstration of feasibility are not yet available.

**RNASE A.** Rick Myers suggested another possible method based on the ability of RNASE A to cleave G:A mismatches in RNA:DNA heteroduplexes. The cleavage is complete and highly selective for this and only this mismatch. The approach is closely parallel to his gradient gel method, except that in this case the probe is RNA, a hair-pin terminus is unnecessary, and the final steps are RNASE A digestion and search for unique, small (labelled) RNA fragments. Using RNAs of both polarity, this method can recognize one-sixth of the twelve possible base changes. Its efficiency and feasibility are similar to Myer's gradient gel method.

**SUBSTRACTIVE HYBRIDIZATION WITH SYNTHETIC OLIGONUCLEOTIDES.** In a method he conceived at the workshop, George Church suggested a way to parallel process an entire genome for any mutant sequences that might be present. The method is based on three concepts: that a library approximating all possible 18-mer oligodesoxynucleotides can be synthesized; that DNA strands larger than circa 40 bases, either single stranded, double stranded or heteroduplexed with an 18-mer, can be fully separated from singlet or doublet 18-mers; and that melting conditions with tetramethylammonium chloride (TMAC, a solvent that eliminates the melting differences of AT-GC pairs) will highly discriminate perfect complements from singlebase or greater mismatches. The plan is to combine DNA from both parents, reduce it to fragments of 40-200 basepairs by physical, chemical or enzymatic means, and add it in excess to a sample of the random sequence 18-mers. The mixture is melted in TMAC and allowed to reanneal at just below the  $T_m$  for perfect matches. After

sufficient time for equilibrium, all 18-mers unassociated with parental DNA are separated. These should represent sequences that are not present in either parent's DNA. The process is then repeated by mixing the similarly fragmented child's DNA with the leftover 18-mers, this time looking for heteroduplexes (of genomic and synthetic DNA) which are perfectly annealed. These should contain 18-mers that match perfectly to the child but to neither parent. Such presumptive mutant markers are counted or analyzed in detail, with the expectation of finding 18 complementary pairs of 18-mers for each altered base in the child's genome.

This method has the potential to identify any genomic changes within 18 base-pairs of a unique sequence, with the exception of the roughly 3% of changes that are by coincidence already represented elsewhere in the triad genomes. As with any of the DNA methods, such events could be due to somatic mutation or physiologically unstable genes (ie immunoglobulin genes), and can be minimized by requiring multiple samples particularly of different tissues and by removing or avoiding genes known to be physiologically modulated. Although never tested in any form and hence unknown in feasibility, this method could be orders of magnitude more efficient than any of the others.

#### DISCUSSION

The major problem facing DNA-based (or any other) methods of detecting heritable mutations is the rarity of mutant events. The predicted background rate of  $10^{-8}$  base-pair changes per generation, converts to 60 potential events in the child's entire genome. Assuming that all such events were detectable, the yield with one exposed and one control triad is just enough to detect a 2.5 fold change in mutation rate with 90% likelihood at a significance level of 5%. Clearly those methods which require multiple technician-years to detect a single event are of little practical use under such circumstances.

The DNA methods tend also to be limited by the types of DNA change that they are sensitive to. A specific basepair change or even all single basepair changes will not be representative of all DNA changes. For some mutagens, such as ionizing radiation, and perhaps for some endpoints, such as carcinogenesis (in somatic mutagenesis), base changes may be relatively unimportant. DNA rearrangements are poorly detected by current DNA and protein methods, and without new

approaches there will remain a large unresolved region bounded below by the 1- to 10- kilobase level of presently useful DNA fragments and above by the megabase lowest resolution of chromosome aberrations.

#### CONCLUSIONS.

The Workshop identified and analyzed six DNA methods for detection of human heritable mutation, including several created at the meeting, and concluded that none of the methods combine sufficient feasibility and efficiency to be recommended for general application. An increasing flow of innovative, exciting developments in DNA technology is expected to improve rapidly the prospects for practical methods of mutation measurement, and gives further emphasis to the virtue of not rushing to premature application. Significant resources will be required on an international scale for miniaturization, quantitation and automation of DNA technology, and for the conception, development and eventual application of the mutation methods. Resolution of mutation at the DNA level should bring many new challenges to genetics, and may, for the first time, provide a sound basis for health and environmental protection of human DNA.

Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract number W-7405-ENG-48.

#### REFERENCES

- Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331.
- Delehanty J, White RL, Mendelsohn ML (1986). Approaches to determining mutation rates in human DNA. *Mutat Res* (submitted).
- ICPEMC Committee 5 (1983). Mutation epidemiology: Review and recommendations. *Mutat Res* 123:1-11.
- Lerman LS, Fischer SG, Hurley I, Silverstein K and Lumelsky N (1984). Sequence-determined DNA separations. *Ann Rev Biophys Bioeng* 13:393-423.
- Mulvihill JJ and Czeizel A (1983). Perspectives in mutation epidemiology: A 1983 view of sentinel phenotypes. *Mutat Res* 123:345-361.
- Myers RM, Lumelsky N, Lerman LS and Maniatis T (1985). Detection of single base substitutions in total genomic DNA. *Nature* (in press).
- Russell WL (1985). This symposium
- Schull WJ, Otake M and Vael JV (1981). Genetic effects of the atomic bombs: A reappraisal. *Science* 213:1220-1227.