Human Germline Mutation Analysis by Single Genome PCR: Application to Dynamic Mutations

ESTHER P. LEEFLANG Molecular Biology Program, University of Southern California, Los Angeles, California 90089

SIMON TAVARÉ Molecular Biology Program, University of Southern California, Los Angeles, California 90089; and Department of Mathematics, University of Southern California, Los Angeles, California 90089

PAUL MARJORAM Department of Mathematics, University of Southern California, Los Angeles, California 90089

RAJI GREWAL, CAROLYN O. S. NEAL, AND NORMAN ARNHEIM[1] Molecular Biology Program, University of Southern California, Los Angeles, California 90089

[1]To whom correspondence should be addressed.

I. THE DYNAMIC MUTATION PROCESS

Trinucleotide repeat disease alleles can undergo "dynamic" mutations in which repeat number can change when a disease gene is transmitted from an affected parent to an offspring (reviewed in [1]). The molecular basis of dynamic mutation is of great fundamental interest and stands in contrast to the static nature of classical nucleotide substitution mutations when transmitted through families. The discovery of trinucleotide repeat diseases has stimulated interest in studying length mutations in microsatellite repeats with the aim of understanding fundamental aspects of the expansion and contraction processes.

In the dynamic mutation process, repeat instability can be influenced by the number of repeats, the sex of the transmitting parent, and the "purity" of the repeat tract [1]. A role for DNA sequences linked and unlinked to the repeat region has been proposed [2–4]. Additional factors such as age and even environmental influences may also be involved.

Studies on model systems involving trinucleotide repeats cloned in *Escherichia coli,* yeast, and mice [5–10] have not yet been able to reproduce all the mutation properties typical of the transmission of many large human disease alleles. Thus, understanding the contribution of the different factors involved in germline instability of trinucleotide repeats in humans must be grounded in a description of the human mutation process.

II. ADVANTAGES OF USING SINGLE SPERM GENOMES FOR ANALYSIS OF GERMLINE MICROSATELLITE MUTATIONS: COMPARISON TO FAMILY STUDIES

The study of microsatellite mutations in humans is traditionally based on family studies and is subject to several limitations. Since only a small number of offspring are available from any one family, the number of available germline transmissions is relatively small and is insufficient to estimate accurately the mutation frequency in the parents. To obtain meaningful estimates, data must be pooled from many different families. Another potential problem is that family studies based on analysis of DNA derived from lymphoblastoid cell lines have dinucleotide repeat mutation frequencies higher than in untransformed lymphocytes [11].

The finding that PCR can be used to determine the genotype at a locus in a single human diploid or sperm cell using PCR [12] or a diluted molecule from sperm [13] led to alternative approaches to studying human germline recombination and mutation. Single genome analysis of gametes is a more precise way of measuring mutation frequencies. The large sample sizes afforded by single genome analysis make it possible to measure accurately the mutation frequency and size distribution of mutant sperm (mutation spectrum) at trinucleotide repeat loci in a single individual. Since the meiotic products themselves are studied directly, mutations occurring during the culture of lymphoblastoid lines and the effects of biological selection following gametogenesis or fertilization are eliminated. The ability to study mutations in a single individual makes it easier to design experiments to determine the role that different factors play in contributing to microsatellite repeat instability.

III. METHODS OF SINGLE GENOME STUDY

A germline mutation is defined in reference to the size of the allele inherited at the time of fertilization. In many of the trinucleotide repeat diseases this can easily be determined by analysis of the allele size in somatic tissues of that individual. In some of the diseases, however (e.g., Frax A and DM), this is not so easily accomplished since there can be extensive allele

size heterogeneity in somatic tissues [1]. Single genome analysis of trinucleotide repeat germline mutations has been carried out in a number of ways (see below). Which method to use will depend on many factors including any problems associated with amplifying the particular locus, the mutation frequency, and the size distribution of the mutations.

A. Single Sperm Analysis

One method involves isolation of single sperm cells (using flow cytometry, micromanipulation, or dilution) followed by lysis, PCR, and allele length analysis using electrophoresis. Single sperm typing has been carried out at the HD, SBMA, SCA-1, DM, SCA-3, and Frax A loci including normal, premutation, and disease alleles [14–20].

B. Small Pool PCR

An alternative approach is to use small pool PCR (SP-PCR) [21]. Dilutions of DNA purified from tens of millions of sperm are made such that each sample contains 5–10 genomes. In such a diluted sample, PCR product from a mutant genome will have an altered electrophoretic mobility compared to PCR product from the unmutated genomes. This method can more easily detect mutations involving the addition or deletion of many trinucleotide repeats, but one or two repeat changes in large alleles are more difficult to verify due to overlapping of the PCR "stutter" bands. SP-PCR has been applied to the germline analysis of Frax A premutation alleles [22] and both premutations and disease causing DM alleles [23].

C. Single Molecule Dilution

More extensive DNA dilutions, such that each sample carries on average a single or a fraction of a genome, can also be made. This approach has been used successfully to study dinucleotide repeat mutations in human tumor cells and tissues of mice lacking DNA mismatch repair [24–27] as well as trinucleotide repeat mutations in somatic tissues of patients with Huntington's disease [17]. If the mutation frequency is high, the sperm DNA should be diluted to well below single copy level to avoid samples containing two mutations. Compared to single sperm typing, a larger number of PCR assays are required since many aliquots will contain no template molecules. An advantage is that it requires no special technologies for sperm isolation.

D. Total Sperm PCR

Crude analysis of mutation spectra can be made in single individuals by carrying out PCR on total DNA samples purified from semen [28, 29]. Based on the extent of the smear of PCR products after electrophoresis, the degree of triplet repeat mutation can be evaluated on a gross level, with maximum, minimum, and modal values of allele size deduced from densitometric analysis. One advantage of this method is that the general level of stability at a locus can be estimated in a very large number of individuals with relatively little effort. Problems with this approach include the lack of detail in the mutation spectra which are critical for some analyses (see below). For example, faint signals derived from a small subset of the sperm population may be obscured by the inherent PCR stutter from more abundant allele sizes. Perhaps more serious is the possibility that data from amplification of total sperm DNA will be strongly biased toward smaller alleles due to their competitive advantage during PCR.

IV. *IN VITRO* PCR ARTIFACTS ARE MINIMAL

Single genome analysis requires many PCR cycles for detection of PCR product. Because of the possibility that PCR artifacts might result in *in vitro* generated mutations, specific control experiments have been designed to assess the likelihood of such artifacts.

A. Contamination

One possible artifact is that any observed mutation might result from contamination by DNA template derived from another individual or previously amplified PCR product. Experiments that yield positive results from no DNA template control samples should be viewed with caution. Contamination of the sample DNA with the genomic DNA from another individual can be examined in the case of single sperm typing by simultaneously amplifying tightly linked informative markers [17, 30].

B. PCR Stutter

Another possible artifact is that mutations may result from events that take place during PCR. Amplification of most di- and trinucleotide repeats using total genomic DNA as template typically results in the appearance of minor PCR products with altered repeat lengths. This

could be the consequence of somatic heterogeneity in the tissue from which the DNA was isolated or DNA replication slippage events during amplification. That PCR alone is capable of producing "stutter" is documented by the stutter patterns detected when using a single DNA molecule as template [17, 19, 20, 31]. Only PCR artifacts that occur in the first few cycles have a chance of producing detectable amounts of product which could be counted as a true *in vivo* mutation. Such false mutations are more likely to be observed as contractions than expansions since a smaller template generated by a PCR slippage error could increase in relative frequency due to a selective advantage during amplification.

C. Amplification Bias

If the PCR assay detects unmutated allele sizes more readily than mutated allele sizes a bias in the mutation frequency and mutation spectrum will result. This could occur if many mutants were expanded beyond a size possible for efficient amplification. A specific criterion can be used to test for such a bias. According to Mendel's laws, approximately 50% of single sperm genomes from an individual heterozygous for a normal allele and a disease allele would be expected to contain the normal allele. The disease allele and all mutations derived from it should make up the remaining 50%. The detection of a significantly large excess of the normal allele could suggest that not all mutations are detectable. In the case of HD [17] there does not appear to be any bias. This might not be the case for very large Frax A or DM alleles, which are more difficult to amplify. An alternative explanation for a deviation from Mendel's laws in sperm typing data is segregation distortion [18, 30]. Specific methods for detecting segregation distortion in sperm typing data have been developed [30].

D. Experimental Approaches to PCR Artifact Detection

An experimental strategy has been used [19, 32] to determine whether a single sperm identified as a mutant was misclassified due to an artifact during amplification or resulted from a true germline mutation event (Fig. 37-1). Single sperm are amplified for 6 PCR cycles. Half of the PCR product is removed from each sample and saved. The reaction is continued without any interruption with the remaining material and the final products analyzed to detect samples with a mutation. If a sperm with a mutation is identified, the corresponding PCR product that had been saved after 6 cycles can be analyzed. If the original sperm had a mutation, the molecules in the saved portion (half the duplexes produced after 6 cycles, or approximately 32 molecules maximally) should have the mutated allele. However, if the sample originally contained a normal molecule and a mutation occurred in the first PCR cycle (or later), a mixture of normal and mutant molecules would be present. Thus, to distinguish between *in vitro* and *in vivo* mutations, PCR is carried out on single molecule dilution aliquots from the saved sample that had undergone the first 6 PCR cycles. If normal molecules are present in these aliquots, the original sperm must have been normal. We have never found evidence for PCR artifacts contributing to misidentification of a sperm as being normal or mutant.

Another way of assessing the possible role of PCR artifacts in single genome analysis is to compare the size distribution of sperm mutations with the size distribution of mutations in somatic DNA from the same individual as determined by single molecule dilution experiments [17]. For example, analysis of HD sperm [17] from one donor with 51 CAG repeats in his somatic DNA showed a mutation frequency of 99% with a mean change of +21 repeats. Single molecule dilution analysis on somatic lymphoblastoid DNA from this donor was also carried out. Among the 30 disease-length molecules examined, seven events with a change of −1 repeats and three events with a change of +1 repeats were detected for a mean change of −0.13 repeat. Whether the variation in allele size in somatic DNA is due to PCR artifact, some variation in allele size among individual lymphoblast cells (possibly occurring during cell culture) or due to the inherent error in the exact measurement of molecular weight of the PCR product cannot be determined. Similar studies were carried out on sperm and somatic DNA from a donor with 36 repeats [17] which also showed that somatic variation was restricted compared to that of the germline. We conclude that the observed variation in single sperm allele sizes is due to germline mutation events.

V. SINGLE SPERM STUDIES ON SBMA

SBMA is a rare disease caused by CAG repeat expansions in the coding region of the X-linked human androgen receptor (AR) gene [33]. Clinically, SBMA is characterized by adult-onset proximal muscular atrophy with bulbar involvement and slow progression [34]. Affected individuals have more than 40 repeats.

A. Mutability of Non-Disease-Causing Alleles

The first studies to analyze trinucleotide repeat germline mutations using single genome analysis [19] focused

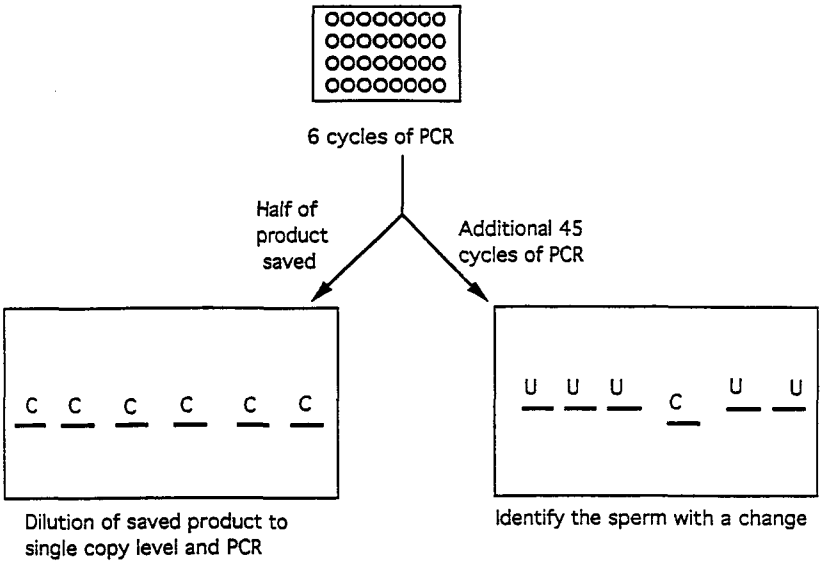

FIGURE 37-1 Outline of the control experiment showing that mutations detected by sperm typing do not result from PCR artifacts. Six cycles of PCR are carried out on individual sperm in a microtiter plate. Half the volume from each well of the microtiter plate is taken and saved. PCR on the remaining sample in each well is continued. Electrophoresis of PCR product (lower right) will reveal whether any sperm is unmutated (U) or underwent a mutation (C = contraction). The saved aliquot from the well exhibiting the contraction is diluted to single copy level and amplified. Following electrophoresis (lower left) only mutated PCR products are expected if the original sperm itself had been mutated *in vivo*. Reprinted by permission of Nature Publishing.

on how disease alleles might arise from normal alleles at the AR locus. The mutation frequency of normal alleles with the average repeat number in the population (20–22 repeats) was compared to the mutation frequency of alleles at the highest end of the normal range (28–31 repeats) but below the repeat number that causes disease.

Of 685 informative sperm from three individuals with alleles in the average size range (20–22 repeats), 1.33% were found to be mutated. Among 1253 informative sperm from four normal individuals with alleles in the 28–31 repeat range, the mutation frequency was 3.21%, a significantly larger proportion.

Most mutations were contractions. Contractions among the 28- to 31-repeat alleles occurred at a frequency of 2.9% compared to a frequency of 0.9% for the alleles of average size. This difference is also statistically significant. On the other hand, the expansion frequencies of the two allele size classes were much smaller and not significantly different from each other (0.31 and 0.43%, respectively). About half of the contractions and expansions observed involved one or two repeat changes.

It is interesting to note that the great excess of contractions over expansions (9:1) typical of the large (28–31 repeats) normal alleles at the AR locus would make it difficult for a normal allele to reach the repeat number required to become a disease allele. This may explain the lack of success in detecting alleles in the 33–39 repeat range in studies of normal populations. This would also explain the low prevalence of SBMA and why, compared to Huntington's disease, no new mutations have been documented for SBMA [35]. Additionally, the sexual dysfunction associated with SBMA may also limit the number of 33–39 repeat alleles introduced into the population by contractions of disease-causing alleles in affected individuals (see Section V.B).

B. Instability of Disease-Causing Alleles

Studies on 1538 single sperm from two SBMA patients sperm ([20] and Grewal *et al.*, unpublished studies) revealed mutation frequencies of 80% (49 repeat individual) and 81% (47 repeat individual). The mutation spectrum for one individual [20] is shown in Fig. 37-2. The expansion mutation frequencies (66 versus 55%) and mean change in allele size for expansions (+2.7 versus +2.1 CAG repeats) were not significantly different between the two patients.

These sperm typing results are consistent with the limited data on transmissions of SBMA alleles pre-

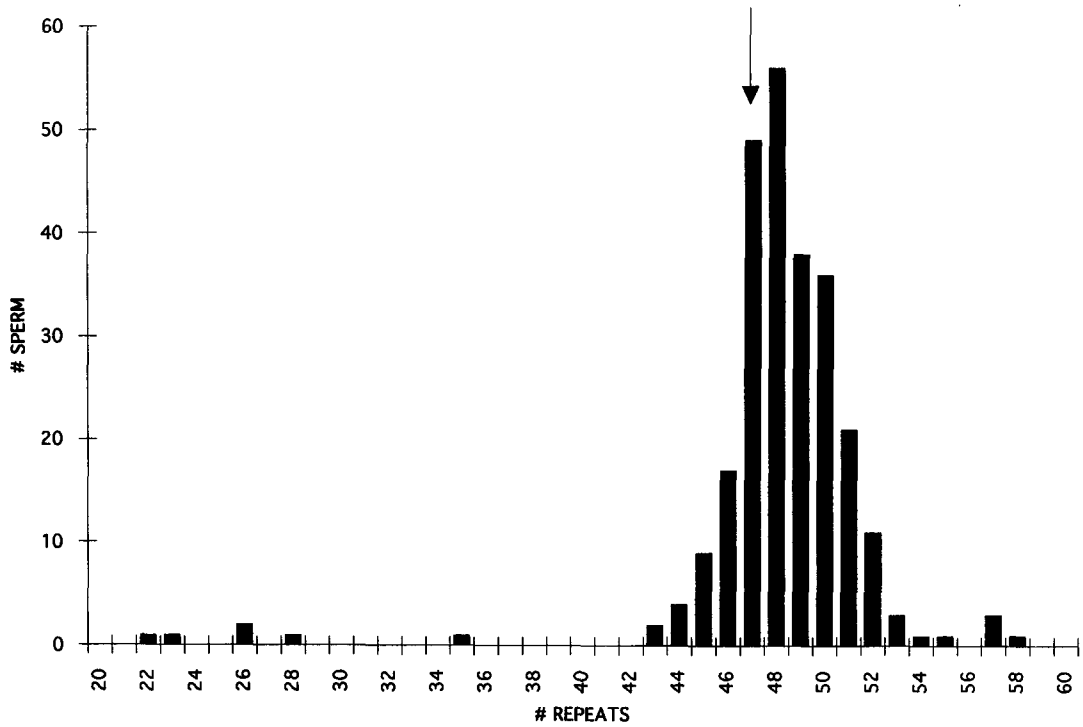

FIGURE 37-2 Distribution of allele size changes in single sperm from a patient with SBMA. The length of the donors SBMA allele in somatic cells was 47 repeats. Arrow denotes the length of the allele in somatic cells. Reprinted by permission of Nature Publishing.

viously reported in pedigrees [34, 36, 37]. In these families, only 17 transmissions via affected fathers are recorded. Eighteen percent showed no change, 12% had one repeat contractions, and 70% had expansions with an average gain of three repeats. Although the family data are generally consistent with the sperm typing data, a more detailed comparison can only be made if larger numbers of family studies with paternal transmissions become available. This will be difficult given the low incidence of the disease.

C. Comparison of the AR Locus Mutation Properties between Normal and Disease Alleles

The trinucleotide repeat mutation frequencies at the AR locus of normal individuals can be compared to those of the two SBMA patients. As the CAG repeat number increases from normal to the disease range, the average frequency of contraction increases about 20-fold. The average expansion frequency increases disproportionately by about 165-fold (from 0.37 to 61%). Since the contraction frequency does not decrease proportionally with the increase in the expansion frequency, the two mutational events show an independence that may be due to distinct molecular mechanisms [19]. For example, an additional mutation process may engage only when allele size exceeds a threshold (disease-causing alleles) and lead primarily, unlike the situation for normal alleles where contractions predominate, to expansions.

VI. SINGLE SPERM STUDIES ON DM

A study on three sperm donors with large normal alleles at the DM locus was carried out for comparison with the AR locus [19]. The donors were heterozygous for one allele at the high end of the normal range and the 5-repeat allele predominant in the general population. The single sperm were simultaneously amplified for a tightly linked polymorphic marker to distinguish which of the two DM alleles underwent a mutation event. Based on 788 observations on all three individuals, no mutations were confirmed in the 5-repeat allele class. The mutation frequency estimate of a 20-repeat DM allele (based on 249 observations) revealed contraction and expansion frequencies (0.8 and 0.4%, respectively), quite similar to the results for the 20- to 22-repeat allele size class at the AR locus. Analysis of a 27-repeat DM allele revealed a mutation behavior quite different from normal alleles of a similar size at the AR

locus. Based on 277 observations, the mutation frequency of the 27 repeat allele was estimated to be 6%. Even more surprising was the finding that expansions were 5.5 times more frequent than contractions. Finally, a 37-repeat allele that was expected to exhibit an even higher mutation frequency showed only 1% expansions. Unlike the 27-repeat allele, this larger allele underwent more contractions (2.6%) than expansions.

The observation that the individual with 27 repeats at the DM locus has an expansion frequency greater than the donor with 37 repeats was unexpected. We subsequently determined the DNA sequence of the 37-repeat allele. Unlike all other alleles reported at the DM locus we found that this allele had the structure $(CTG)_4(CCGCTG)_{16}CTG$ [31]. This structure could explain why the 37-repeat allele has a lower mutation frequency than the 27-repeat allele. Interrupted repeat tracts, as seen in normal alleles of SCA1 [38], SCA2 [39–41], and Frax A [14, 42, 43], play an important role in reducing instability.

VII. SINGLE SPERM STUDIES ON HD

Huntington's disease (HD) is a dominant hereditary neurodegenerative disorder, usually with an onset in middle age, and associated with progressive disordered movements, decline in cognitive function, and emotional disturbance. A large, worldwide study [44] showed that the repeat number at the HD locus in normal individuals ranges from 10 to 29 repeats. New disease alleles presumably arise as a result of a gradual increase of normal alleles into a repeat number range (29–36 repeats) referred to herein as intermediate size alleles or reduced penetrance alleles [45–48]. These alleles can, in a single generation, increase in size so as to cause clinical symptoms. Disease alleles can be very much more unstable, especially in paternal transmissions.

A. Instability of Normal HD Alleles

Data on 475 sperm from five normal HD alleles [17] in the 15–18 repeat range estimate a mutation frequency of 0.6% resulting from one- or two-repeat contractions, similar to a frequency estimate based on pedigree analysis (0.2%) [49]. Both estimates are uncertain considering the small number of mutations observed. Note also that sperm typing estimates measure the mutation frequency of specifically chosen allele sizes. Measuring the mutation frequency for a particular allele size based on pooling data from many different families segregating for the same size allele will almost always lack sufficient numbers for an accurate estimate. If the data for different allele sizes are pooled, the estimate of the mutation frequency is not allele size specific but represents an average mutation frequency based on the average of the allele sizes represented in the families.

B. Instability of Intermediate Size HD Alleles

Alleles with as few as 29 repeats have been found in fathers of offspring clinically affected with HD from families with no previous history of the disease. It is clear, however, that the definition of a new mutation depends upon whether the father does or does not exhibit a clinical phenotype (for a discussion see [16, 45]).

The instability of an intermediate allele carried by a father can be examined by sperm typing. Among 163 sperm analyzed from an unaffected individual with a 36-repeat allele and in his fifth decade [17], 16% of the gametes derived from the 36-repeat allele repeat (Fig. 37-3A) had expanded into the allele size range expected to give a clinical phenotype (defined for this purpose as a ≥38-repeat allele). The chance of having a clinically affected child is, for this man, 8%. Recent sperm typing studies [16] on an individual heterozygous for a 35-repeat allele indicate a comparable risk of 2.3%. Studies on two other individuals with 29 and 30 repeats [50] showed the risk of having a clinically affected child as 0.1 and 0.3%, respectively.

The risk of a male carrier of an intermediate allele having a child with HD may vary among individuals as a consequence of *cis*-acting DNA sequences, related to their haplotype [51], that affect mutation frequency (see [16]). Regardless, the risk to any particular father carrying an intermediate allele of having an affected child can be determined directly using single molecule analysis.

C. Mutation Frequency of Disease-Causing Alleles

Studies on three sperm donors heterozygous for alleles with repeat numbers >39 have been published [17]. Among these individuals, 316 normal alleles were detected, while 287 HD alleles were observed; this is not statistically different from the expected 1:1 segregation. Thus, both large and small alleles can be amplified with almost equal efficiency.

Among the 287 sperm carrying disease causing alleles from these three individuals, 96% differed in size from the donors' somatic HD allele; 93% were expansions and 3% were contractions. Two-thirds of all the samples were studied to distinguish between alter-

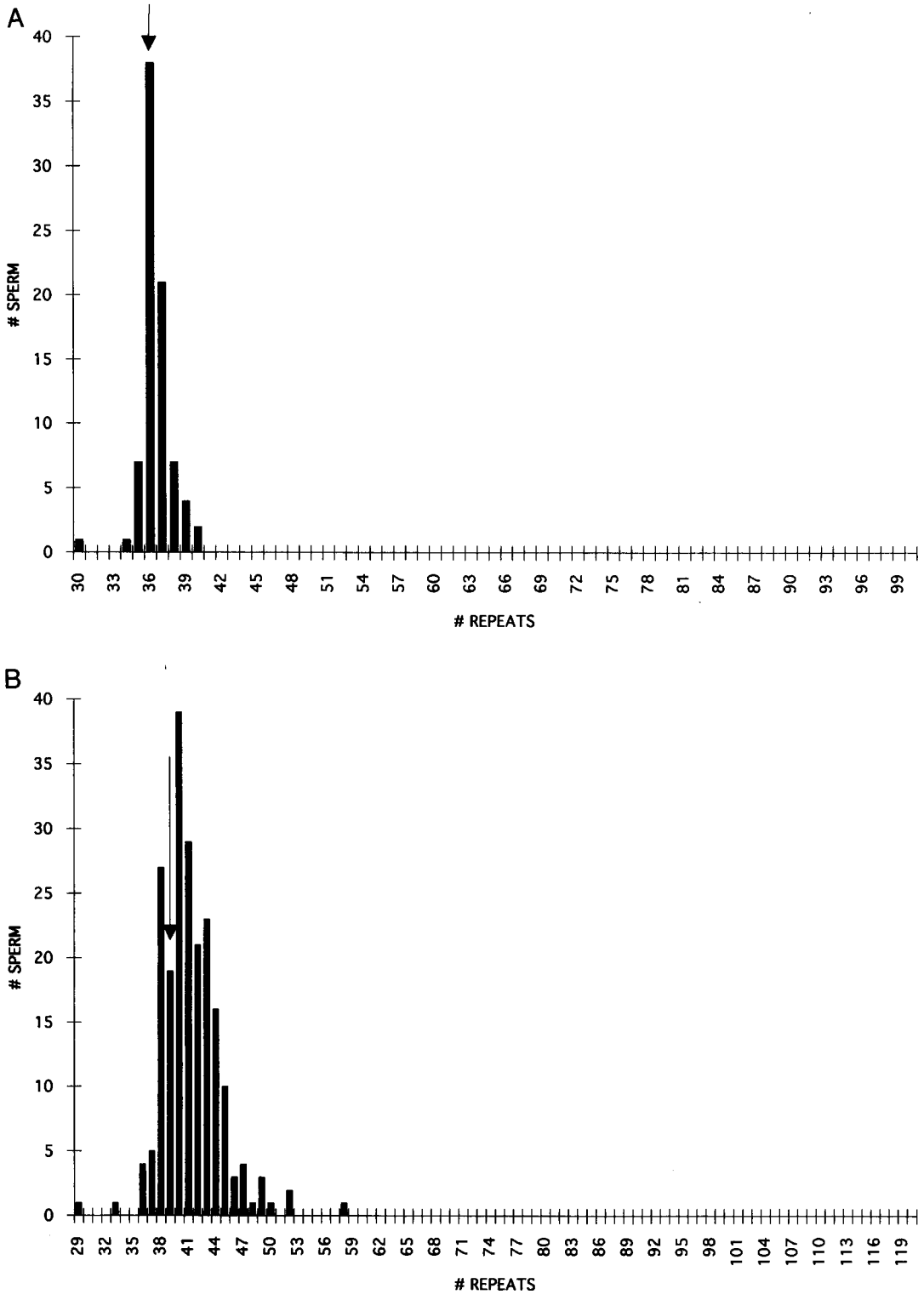

FIGURE 37-3 Mutation spectra of sperm donated by four individuals heterozygous for a 36 (A), 39 (B), and 49 (C) and 51 (D) repeat HD allele. A sample from the individual with 39 repeats but taken 2 years later is also shown (E). Arrows denote the length of the donor's HD allele in somatic cells determined by analysis of lymphoblast DNA. Figures A–D reproduced from *Hum. Mol. Genet.* **4,** 1519–1526, 1995, by permission of Oxford University Press.

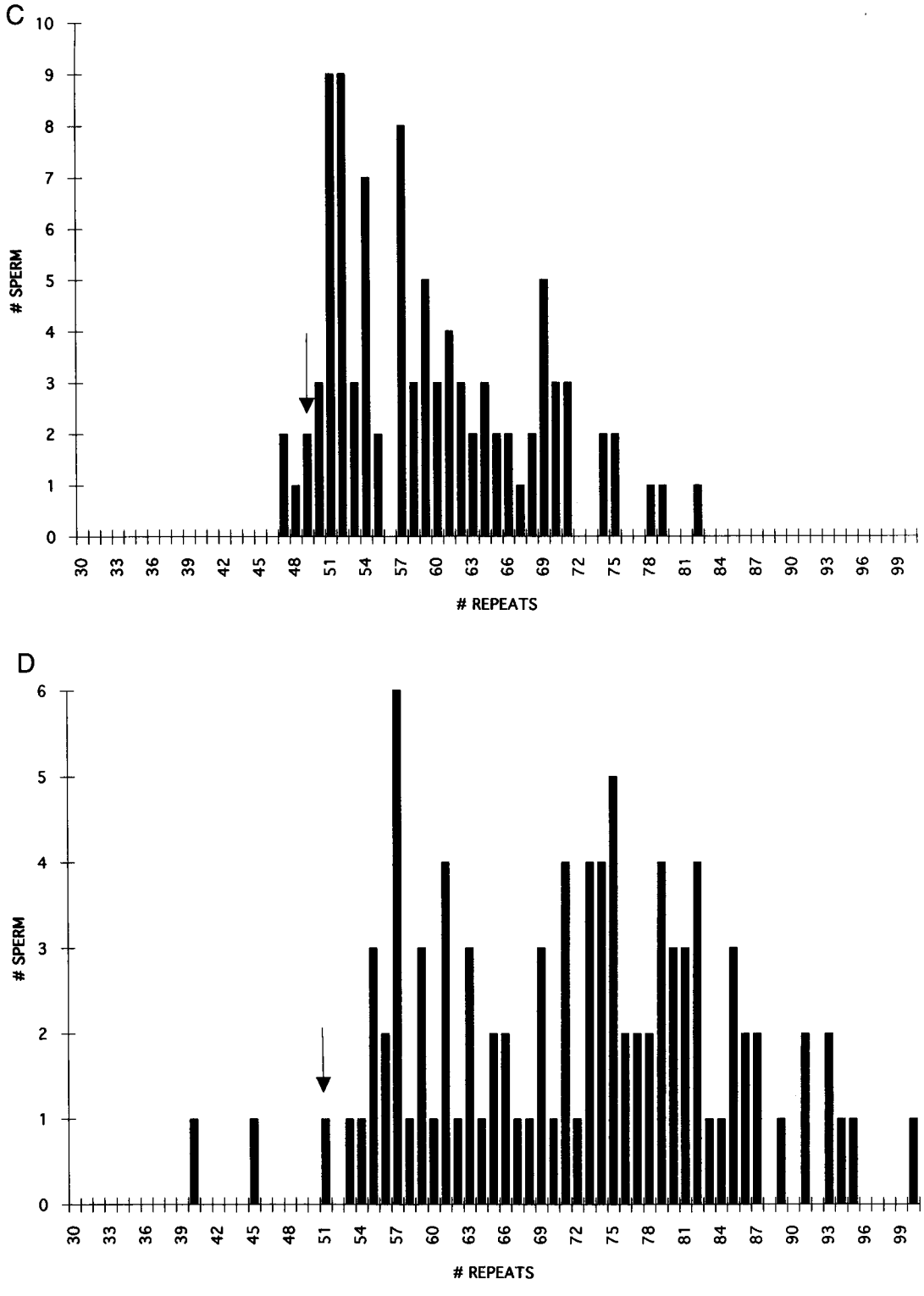

FIGURE 37-3 (*continued*)

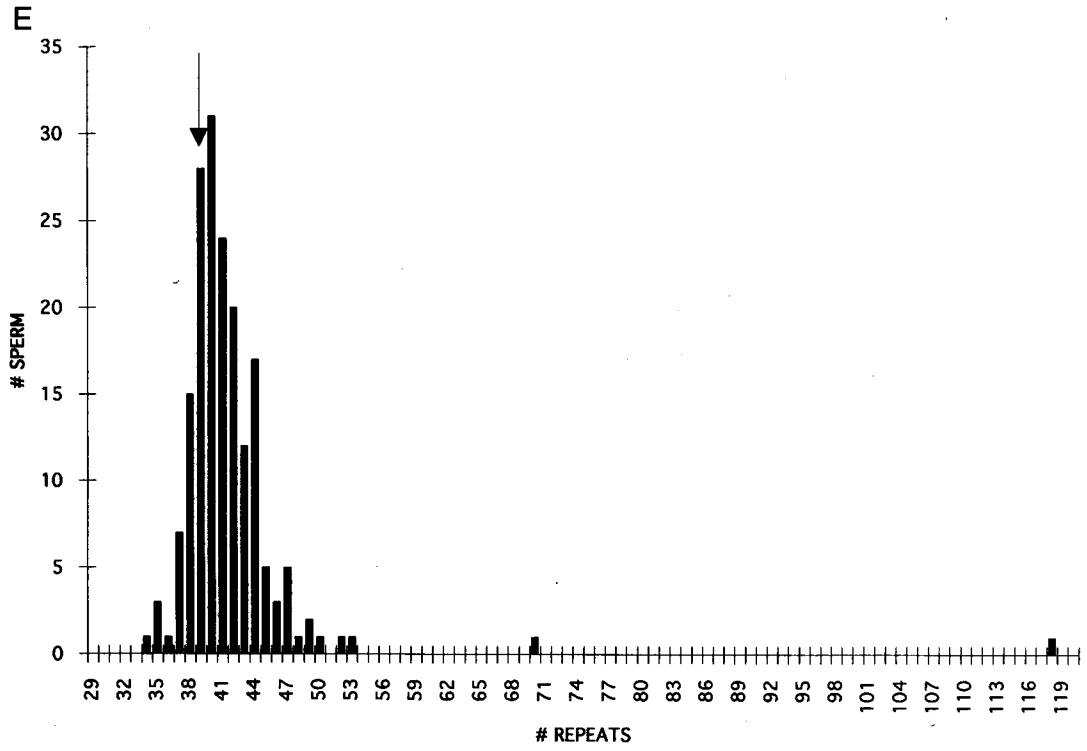

FIGURE 37-3 *(continued)*

ations in the number of CAG repeats and in the less polymorphic and immediately adjacent CCG repeats [52, 53]. No mutations in CCG repeat number were detected.

D. The Size Distribution of Mutations Derived from Disease-Causing Alleles

The CAG repeat number distribution in single sperm donors with 39, 49, and 51 repeat tracts are shown in Figs. 37-3B–37-3D, respectively [17]. There is a significant change in the size distribution of the expansion mutations with increasing allele size. It can be seen to progress from an apparently normal distribution around the somatic DNA size in the case of the 36-repeat allele (Fig. 37-3A), to a markedly more uniform distribution of expanded alleles up to twice the size of the somatic DNA in the case of the larger alleles. Contraction sizes were generally limited to 6 repeats or less.

E. Mutation Spectra of Two Samples Taken from the Same Individual

We were also able to estimate the constancy of HD allele size distributions on a time scale of 2 years by analyzing sperm samples taken at age 63 (Fig. 37-3B) and age 65 (Fig. 37-3E) [54] from the donor with 39 repeats. The mutation spectra show virtually identical patterns except for two sperm in the older sample that had allele sizes considerably larger than those seen in the younger sample.

F. Effects on Mutation of *Cis* Polymorphisms and Parental Origin

Recent analysis of instability in MJD families reported that heterozygosity for a single nucleotide polymorphism adjacent to the repeated region influences the instability of the disease-causing allele [2]. A tightly linked polymorphism exists adjacent to the HD CAG tract [52, 53]. Based on data from 26 HD sperm donors [54] we divided our samples into those that were homozygous for the $(CCG)_7$ allele (7 donors) and those that were heterozygous for $(CCG)_7/(CCG)_{10}$ alleles (19 donors). We found no significant difference between the two groups with respect to mean change in repeat number following mutation (unpublished). Presumably neither the CCG polymorphism, nor any other polymorphisms in linkage disequilibrium with it, contribute significantly to the instability properties of the CAG repeat tracts.

An analysis of the mean change in repeat number based on grouping the HD alleles according to whether they were paternally (7 donors) or maternally (18 donors) inherited was also made. The results (unpublished) reveal that there was no major effect of parental imprinting on instability.

G. Comparison of HD Sperm Typing Data to Family Transmissions

The published single sperm typing data on the HD alleles [17] can be compared with HD paternal transmissions studied in families. In five reports [46, 47, 55, 56] exact repeat number changes can be determined in 141 paternal transmissions. Of these, 72% showed a contraction or expansion mutation. The single sperm data [17] gave a 96% average mutation frequency. Expansions account for 97% of the sperm mutations and 91% (range: 83–100%) of the mutations detected in paternal family transmissions. An average of 5.6 repeats was added for each expansion observed in the pedigrees (range: 3.0 to 9.0) while 12.1 repeats was the average size of an expansion detected in sperm. The differences between the sperm and family data could result from the small number of HD sperm donors studied and the fact that two of the three had an HD allele size significantly larger than the average paternal HD allele size in the families. There may also be differences among the populations studied. Some of the differences between the sperm and family data could also reflect selection against sperm with very large alleles in terms of survival, the ability to carry out fertilization, or postzygotic selection.

H. Comparison of SBMA and HD Disease Allele Instability

Single sperm data from the two SBMA individuals with 47 and 49 CAG repeats showed an 81% average mutation frequency (61% expansions and 20% contractions). The change in the number of repeats averaged 2.4. The largest expansion seen was +11. The sperm data on the HD allele closest in size (49 CAG repeats) showed a 95% expansion and a 3% contraction frequency with an average of 10.8 repeats added. The largest expansion was +33 repeats. Even donors with a smaller number of HD repeats can have a higher mutation frequency, a greater average change in repeat number per mutation, and a broader distribution in mutant allele size than the 47- and 49-repeat SBMA alleles.

The difference in the instability at these two loci is just one more demonstration that repeat number is not the only important factor that contributes to the measured triplet expansion frequency (see [1]). The differences in mutation behavior between the loci remain to be explained.

VIII. ANALYSIS OF SPERM TYPING DATA

A. Elucidating the Contribution of Different Variables to the Trinucleotide Repeat Mutation Frequency

The ability to estimate accurately the mutation frequency for a given trinucleotide repeat allele requires a large sample and can usually be achieved only by single genome methods such as sperm typing. Up to the present, data on only a small number of sperm donors have been published for any one disease. However, if a large sample of individuals could be studied, then the factors that influence the trinucleotide repeat mutation frequency can be examined using routine epidemiological designs. Standard statistical methods could be used to study the relative effects on germline mutation frequency of repeat number, *cis* DNA sequences, unlinked or linked polymorphisms, parental origin of the examined allele, or indeed any other variable. For example, a study of men from the large Venezuelan HD cohort [54] would effectively eliminate any influence of *cis* DNA sequences as a variable since all of the HD alleles derive from the same founder chromosome [57].

B. Understanding the Trinucleotide Repeat Mutation Process

The sperm typing data may also be used to gain insight into the molecular details of the trinucleotide repeat mutation process. One approach is to create a realistic probabilistic model of the mutation process based on current biochemical information [17]. The success of any model can be judged on the basis of how well it describes the actual sperm typing data. Such models can assist in formulating hypotheses to be tested by biochemical, biophysical, and genetic studies. One important question is how the number of repeats influences the trinucleotide repeat mutation rate. For example, mutations in large disease alleles show greater changes in repeat number than smaller disease alleles. A simple explanation is that larger alleles have more triplets and, given a constant mutation rate per triplet during DNA replication, the likelihood of mutation is greater. A number of factors that must be considered when trying to estimate the trinucleotide repeat mutation rate follow.

C. Definition of the Mutation Rate

Estimates of the nucleotide substitution mutation rate for cells in culture using standard methods are defined in terms of mutations per cell division. In humans, mutation rates are often defined in terms of mutations per generation. Unlike "classical" mutations, not all sperm carrying a trinucleotide repeat mutation have necessarily undergone the same number of mutation events. Thus, a sperm which has gained 10 repeats compared to somatic DNA could have undergone the expansion due to a single mutation event during meiosis or have experienced a mixture of multiple, but smaller, expansion and contraction events occurring during mitosis over many germline cell divisions. This uncertainty complicates any estimate of a mutation rate. Indeed, the definition of mutation rate in this case is ambiguous. It could refer to the probability of a new allele being transmitted per human generation, the probability that a new allele will be generated per cell division, or the probability that as each triplet is being replicated a change in repeat number will occur. These ambiguities require new analytical methods to be devised.

D. Molecular Mechanism of Mutation

Both contraction and expansion events in a repeated sequence can result from unequal reciprocal recombination between homologs or sister chromatids. In either case, equal numbers of contractions and expansions should be observed, an expectation that is not fulfilled by the trinucleotide repeat data. Nonreciprocal recombination mechanisms leading to changes in repeat number [58] cannot be excluded, although data from yeast suggest that any role for recombination is unlikely (see [59]). Mutations due to replication slippage are a more likely possibility [60] and recent experimental evidence on microsatellite repeat mutations in yeast and studies on human colon cancer also support this type of mechanism [6, 7, 61–65].

E. Details of the Molecular Mechanism of a DNA Replication Slippage Event

The proximal cause of a slippage event, the number of repeats added or deleted per slippage event, and the number of slippage events that take place as the polymerase traverses the repeated region is unknown. In the case of trinucleotide repeat sequences, mutations may be promoted by the ability of the repeated DNA to form secondary structures (reviewed in [5, 66]). Slippage could result from formation of an intrastrand secondary structure on the replication template. If the replication complex were to "skip over" the folded region, the number of repeats would be reduced. Alternatively, elongation of the nascent strand might be impeded by such a structure, leading to backwards slippage of the nascent strand, additional triplet incorporation, and an expansion mutation.

Slippage mutations could occur when either the leading or lagging strand serves as DNA replication template. Arguments supporting the importance of slippage on the lagging strand include the observation that significant instability at trinucleotide repeat containing loci is first manifested when the length of the repeated region approaches the size of an average mammalian Okazaki fragment [67]. Data on the strand preference of some mutation processes in *E. coli* also support a lagging strand model (reviewed in [43]).

Once DNA replication is completed, mutations can be subject to DNA repair. Yeast and human tumor cells (reviewed in [65]) as well as mice [24–26, 68] that are deficient in DNA mismatch repair have increased dinucleotide repeat mutation frequencies. How DNA repair specifically influences trinucleotide repeat mutations in mammals is not known.

F. Meiotic or Mitotic Origin of Mutations

Trinucleotide repeat mutations could occur during meiosis or throughout the mitotic divisions of the germline. If mutations occur mitotically, the number of cell divisions over which a mutation event may have occurred needs to be estimated. One approach is to consider the age of the donor at the time the sperm sample was collected. Before spermatogenesis begins at puberty (assumed to begin at age 13), the spermatogonial stem cells have undergone an estimated 34 divisions since formation of the zygote. After puberty the stem cells divide approximately 23 times a year [69].

G. Mathematical Modeling

Probabilistic models of molecular mutation mechanisms may be used in conjunction with observed mutation spectra to study the details of the mutation process. Any such model includes a number of unknown parameters, for example those defining the number of triplets added or lost during an expansion or contraction event, and the rate of contraction or expansion events during replication. Comparison of the observed mutation spectra with the spectra predicted by the model can be used to estimate the parameter values that best match the model to the data. Many different numerical methods

can be used to accomplish this (for an example that uses maximum likelihood estimation, see [17]). For a given mutation model, even the "best-fitting" one may not provide a good description of the underlying data. To assess this issue, it is usual to compare the best-fitting spectrum with the observed spectrum, both qualitatively and quantitatively, and look for gross inconsistencies. Such inconsistencies indicate the inappropriateness of one or more assumptions in the model, or omission of a key biological feature. Simulation of synthetic data from the fitted model often provides valuable insight into issues involving adequacy of fit, especially when dealing with discrete events such as repeat number, where particular allele sizes may be represented rather infrequently in the data.

It is important to emphasize that in modeling of this sort there is always a trade-off between simplification and the (putative) biological realities. It is clearly not useful to fit a model with 50 parameters to (say) 20 data points. Thus parsimonious description of the biological issues is required. It is also worth emphasizing that recent developments in computing power and in computational statistics have to some extent freed us from many of the drawbacks of models chosen for analytical or mathematical convenience.

Once a variety of different models of mutation have been fitted to the data, it is often possible to distinguish among them statistically. For example, in another report [17] we fitted two models that assumed mutations occur at the last mitotic division before meiosis. The models differed with respect to the magnitude of change in repeat number that accompanied a single mutation event. By comparing the variance in repeat number predicted under the model with the variance observed in the data we found that small incremental changes are more consistent with the data than very large changes.

In recent unpublished work [54], we have been investigating models that address the issue of whether a single mutation process, occurring during the mitotic divisions of spermatogenesis, can explain mutation spectra taken from a large sample of HD patients. Such a model has to follow the cell division history of a repeat region as it evolves from fertilization through the spermatogonial stem cell cycle to meiosis. The details of the mutation mechanism itself also have to be specified; we assumed a slippage mechanism that might be different on the two template strands. We described the probabilistic mechanism of repeat change during replication of a triplet in terms of three parameters: one gives the chance of a ± 1-repeat slippage event, another the chance of a large expansion, and a third that describes the features of such an expansion. Coupled with the estimates discussed above for the number of mitotic divisions in the history of a typical sperm, we can use the statistical methods outlined previously to estimate the three parameters and assess the adequacy of the fit.

We compared this model to one in which all mutations arise at a single mitotic division prior to meiosis. Do the data allow us to distinguish between these two models? From a statistical perspective, we found that both models provide an adequate fit to the data; statistical considerations alone could not separate the models. However, biological considerations support a model with a multigenerational mutation process rather than the single generation process. This example emphasizes the time-honored observation that statistical adequacy of a model does not imply biological adequacy.

Once a model has been found that adequately summarizes a complex data set such as the HD mutation spectra, it could be used for prediction. For example, the model that allows mutations to arise during any mitotic division could be used to estimate the mutation spectra of a given individual at different ages. The same sort of consideration can be used, in theory at least, to provide an experimental approach to testing such a model: take sperm samples from a given individual over a number of years, and compare the observed spectra with those predicted by the model. As yet, such data are not available.

Finally, we note that a model might also suggest other biological features that warrant further investigation. For example, our recent HD work points to the possible role of two different DNA repair mechanisms that correct "small" and "large" slippage loops.

References

1. Ashley, C. T., and Warren, S. T. (1995). Trinucleotide repeat expansion and human disease. *Annu. Rev. Genet.* **29,** 703–728.
2. Igarashi, S., Takiyama, Y., Cancel, G., Rogaeva, E. A., Sasaki, H., Wakisaka, A., Zhou, Y. X., Takano, H., Endo, K., *et al.* (1996). Intergenerational instability of the CAG repeat of the gene for Machado-Joseph Disease (MJD1) is affected by the genotype of the normal chromosome: implications for the molecular mechanisms of the instability of the CAG repeat. *Hum. Mol. Genet.* **5,** 923–932.
3. Nolin, S. L., Lewis, F. A., Ye, L. L., Houck, G. E., Glicksman, A. E., Limprasert, P., Li, S. Y., Zhong, N., Ashley, A. E., *et al.* (1996). Familial transmission of the FMR1 CGG repeat. *Am. J. Hum. Genet.* **59,** 1252–1261.
4. Murray, A., Macpherson, J. N., Pound, M. C., Sharrock, A., Youings, S. A., Dennis, N. R., McKechnie, N., Linehan, P., Morton, N. E., and Jacobs, P. A. (1997). The role of size, sequence and haplotype in the stability of FraxA and FraxE alleles during transmission. *Hum. Mol. Genet.* **6,** 173–184.
5. Wells, R. D. (1996). Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem.* **271,** 2875–2878.
6. Schweitzer, J. K., and Livingston, D. M. (1997). Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. *Hum. Mol. Genet.* **6,** 349–355.

7. Maurer, D. J., Ocallaghan, B. L., and Livingston, D. M. (1996). Orientation dependence of trinucleotide CAG repeat instability in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **16,** 6617–6622.
8. Monckton, D. G., Coolbaugh, M. I., Ashizawa, K. T., Siciliano, M. J., and Caskey, C. T. (1997). Hypermutable myotonic dystrophy CTG repeats in transgenic mice. *Nature Genet.* **15,** 193–196.
9. Gourdon, G., Radvanyi, F., Lia, A. S., Duros, C., Blanche, M., Abitbol, M., Junien, C., and Hofmannradvanyi, H. (1997). Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice. *Nature Genet.* **15,** 190–192.
10. Mangiarini, L., Sathasivam, K., Mahal, A., Mott, R., Seller, M., and Bates, G. P. (1997). Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nature Genet.* **15,** 197–200.
11. Weber, J., and Wong, C. (1993). Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2,** 1123–1128.
12. Li, H., Gyllensten, U., Cui, X., Saiki, R., Erlich, H., and Arnheim, N. (1988). Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* **335,** 414–417.
13. Jeffreys, A. J., Wilson, V., Neumann, R., and Keyte, J. (1988). Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res.* **16,** 10953–10971.
14. Kunst, C., Leeflang, E., Iber, J., Arnheim, N., and Warren, S. (1997). The effect of FMR1 CCG repeat interruptions on mutation frequency as measured by sperm typing. *J. Med. Genet.* **34,** 627–631.
15. Chong, S. S., McCall, A. E., Cota, J., Subramony, S. H., Orr, H. T., Hughes, M. R., and Zoghbi, H. Y. (1995). Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type-1. *Nature Genet.* **10,** 344–350.
16. Chong, S. S., Almqvist, E., Telenius, H., Latray, L., Nichol, K., Bourdelatparks, B., Goldberg, Y. P., Haddad, B. R., Richards, F., Sillence, D., Greenberg, C. R., Ives, E., Van den Engh, G., Hughes, M. R., and Hayden, M. R. (1997). Contribution of DNA Sequence and CAG Size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analyses. *Hum. Mol. Genet.* **6,** 301–309.
17. Leeflang, E. P., Zhang, L., Tavare, S., Hubert, R., Srinidhi, J., Macdonald, M. E., Myers, R. H., Deyoung, M., Wexler, N. S., *et al.* (1995). Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. *Hum. Mol. Genet.* **4,** 1519–1526.
18. Takiyama, Y., Sakoe, K., Soutome, M., Namekawa, M., Ogawa, T., Nakano, I., Igarashi, S., Oyake, M., Tanaka, H., Tsuji, S., and Nishizawa, M. (1997). Single sperm analysis of the CAG repeats in the gene for Machado-Joseph disease (MJD1): evidence for non-Mendelian transmission of the MJD1 gene and for the effect of the intragenic cgg/ggg polymorphism on intergenerational instability. *Hum. Mol. Genet.* **6,** 1063–1068.
19. Zhang, L., Leeflang, E. P., Yu, J., and Arnheim, N. (1994). Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene. *Nature Genet.* **7,** 531–535.
20. Zhang, L., Fischbeck, K. H., and Arnheim, N. (1995). CAG repeat length variation in sperm from a patient with Kennedy's disease. *Hum. Mol. Genet.* **4,** 303–305.
21. Monckton, D., Neumann, R., Guram, T., Fretwell, N., Tamaki, K., Macleod, A., and Jefferys, A. (1994). Minisatellite mutation-rate variation associated with a flanking DNA-sequence polymorphism. *Nature Genet.* **8,** 162–170.
22. Mornet, E., Chateau, C., Hirst, M. C., Thepot, F., Taillandier, A. C.-O., and Serre, J. L. (1996). Analysis of germline variation at the FMR1 CGG repeat shows variation in the normal-premutated borderline range. *Hum. Mol. Genet.* **5,** 821–825.
23. Monckton, D. G., Wong, L.-J. C. A. T., and Casky, C. T. (1995). Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analysis. *Hum. Mol. Genet.* **4,** 1–8.
24. Baker, S. M., Bronner, C. E., Zhang, L., Plug, A. W., Robatzek, M., Warren, G., Elliott, E. A., Yu, J., Ashley, T., Arnheim, N., *et al.* (1995). Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. *Cell* **82,** 309–319.
25. Baker, S. M., Plug, A. W., Prolla, T. A., Bronner, C. E., Harris, A. C., Yao, X., Christie, D. M., Monell, C., Arnheim, N., Bradley, A., Ashley, T., and Liskay, R. M. (1996). Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over. *Nature Genet.* **13,** 336–342.
26. Edelmann, W., Cohen, P. E., Kane, M., Lau, K., Morrow, B., Bennett, S., Umar, A., Kunkel, T., Cattoretti, G., Chaganti, R., Pollard, J. W., Kolodner, R. D., and Kucherlapati, R. (1996). Meiotic pachytene arrest in MLH1-deficient mice. *Cell* **85,** 1125–1134.
27. Shibata, D., Navidi, W., Salovaara, R., Li, Z. H., and Aaltonen, L. A. (1996). Somatic microsatellite mutations as molecular tumor clocks. *Nature Med.* **2,** 676–681.
28. MacDonald, M. E., Barnes, G., Srinidhi, J., Duyao, M. P., Ambrose, C. M., Myers, R. H., Gray, J., Conneally, P. M., Young, A., Penney, J., Shoulson, I., Hollingsworth, Z., Koroshetz, W., Bird, E., Vonsattel, J. P., Bonilla, E., Moscowitz, C., Penchaszadeh, G., Brzustowicz, L., Alvir, J., Bickham Conde, J., Cha, J.-H., Dure, L., Gomez, F., Ramos-Arroyo, M., Sanchez-Ramos, J., Snodgrass, S. R., de Young, M., Wexler, N. S., MacFarlane, H., Anderson, M. A., Jenkins, B., and Gusella, J. F. (1993). Gametic but not somatic instability of CAG repeat length in Huntington's disease. *J. Med. Genet.* **30,** 982–986.
29. Telenius, H., Kremer, B., Goldberg, Y. P., Theilmann, J., Andrew, S. E., Zeisler, J., Adam, S., Greenberg, C., Ives, E. J., Clarke, L. A., *et al.* (1994). Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm *Nature Genet.* **6,** 409–414. [Published erratum appears in *Nature Genet.* **7**(1) 113, 1994]
30. Leeflang, E., McPeek, M., and Arnheim, N. (1996). Analysis of meiotic segregation using single-sperm typing: meiotic drive at the myotonic dystrophy locus. *Am. J. Hum. Genet.* **59,** 896–904.
31. Leeflang, E., and Arnheim, N. (1995). A novel repeat structure at the Myotonic Dystrophy locus in a 37 repeat allele with unexpectedly high stability. *Hum. Mol. Genet.* **4,** 135–136.
32. Cortopassi, G., and Arnheim, N. (1990). Detection of a specific mitochondrial DNA deletion in tissues of older humans. *Nucleic Acids Res.* **18,** 6927–6933.
33. La Spada, A., Wilson, E., Luban, D., Harding, A., and Fischbeck, K. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352,** 77–79.
34. Shimada, N., Sobue, G., Doyu, M., Yamamoto K, Yasuda, T., Mukai, E., Kachi, T., and Mitsuma, T. (1995). X-linked recessive bulbospinal neuronopathy: clinical phenotypes and CAG repeat size in androgen receptor gene. *Muscle Nerve* **18,** 1378–1384.
35. Tanaka, F., Doyu, M., Ito, Y., Matsumoto, M., Mitsuma, T., Abe, K., Aoki, M., Itoyama, Y., Fischbeck, K., and Sobue, G. (1996). Founder effect in spinal and bulbar muscular atropy (SBMA). *Hum. Mol. Genet.* **9,** 1253–1257.
36. La Spada, A., Roling, D., Harding, A., Warner, C., Spiegel, R., Hausmanowa-Petrusewicz, I., Yee, W.-C., and Fischbeck, K. (1992). Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nature Genet.* **2,** 301–304.

37. Biancalana, V., Serville, F., Pommier, J., Julien, J., Hanauer, A., and Mandel, J. (1992). Moderate instability of the trinucleotide repeat in spinobulbar muscular atrophy. *Hum. Mol. Genet.* **4,** 255–258.
38. Chung, M. Y., Ranum, L. P. W., Duvick, L. A., Servadio, A., Zoghbi, H. Y., and Orr, H. T. (1993). Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type-I. *Nature Genet.* **5,** 254–258.
39. Sanpei, K., Takano, H., Igarashi, S., Sato, T., Oyake, M., Sasaki, H., Wakisaka, A., Tashiro, K., Ishida, Y., *et al.* (1996). Identification of the spinocerebellar ataxia type-2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nature Genet.* **14,** 277–284.
40. Pulst, S. M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X. N., Lopescendes, I., Pearlman, S., Starkman, S., Orozcodiaz, G., *et al.* (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type-2. *Nature Genet.* **14,** 269–276.
41. Imbert, G., Saudou, F., Yvert, G., Devys, D., Trottier, Y., Garnier, J. M., Weber, C., Mandel, J. L., Cancel, G., *et al.* (1996). Cloning of the gene for spinocerebellar ataxia-2 reveals a locus with high-sensitivity to expanded CAG/glutamine repeats. *Nature Genet.* **14,** 285–291.
42. Eichler, E. E., Holden, J. J. A., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A., and Nelson, D. L. (1994). Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature Genet.* **8,** 88–94.
43. Kunst, C. B., and Warren, S. T. (1994). Cryptic and polar variation of the fragile-X repeat could result in predisposing normal alleles. *Cell* **77,** 853–861.
44. Kremer, B., Goldberg, P., Andrew, S. E., Theilmann, J., Telenius, H., Zeisler, J., Squitieri, F., Lin, B. Y., Bassett, A., *et al.* (1994). A worldwide study of the Huntington's disease mutation: the sensitivity and specificity of measuring Cag repeats. *N. Engl. J. Med.* **330,** 1401–1406.
45. McNeil, S. M., Novelletto, A., Srinidhi, J., Barnes, G., Kornbluth, I., Altherr, M. R., Wasmuth, J. J., Gusella, J. F., Macdonald, m. E., *et al.* (1997). Reduced penetrance of the Huntington's disease mutation. *Hum. Mol. Genet.* **6,** 775–779.
46. Zuhlke, C., Riess, O., Bockel, B., Lange, H., and Thies, U. (1993). Mitotic stability and meiotic variability of the (CAG)n repeat in the Huntington disease gene. *Hum. Mol. Genet.* **2,** 2063–2067.
47. Legius, E., Cuppens, H., Dierick, H., Van, Z.-K., Dom, R., Fryns, J. P., Evers, K.-G., Decruyenaere, M., Demyttenaere, K., Marynen, P., *et al.* (1994). Limited expansion of the (CAG)n repeat of the Huntington gene: a premutation (?). *Eur. J. Hum. Genet.* **2,** 44–50.
48. Goldberg, Y. P., Kremer, B., Andrew, S. E., Theilmann, J., Graham, R. K., Squitieri, F., Telenius, H., Adam, S., Sajoo, A., *et al.* (1993). Molecular analysis of new mutations for Huntington's disease—intermediate alleles and sex of origin effects. *Nature Genet.* **5,** 174–179.
49. Novelletto, A., Persichetti, F., Sabbadini, G., Mandich, P., Bellone, E., Ajmar, F., Pergola, M., Del, S.-L., MacDonald, M. E., Gusella, J. F., *et al.* (1994). Analysis of the trinucleotide repeat expansion in Italian families affected with Huntington disease. *Hum. Mol. Genet.* **3,** 93–98.
50. Leeflang, E. P., Fan, F., Losekoot, M., and Van Ommen, G. J. B. (1995). Single sperm analysis of the transition from high normal to diseased triplet repeat length in the Huntington's disease locus. *Am. J. Hum. Genet.* **57,** 140.
51. MacDonald, M. E., Novelletto, A., Lin, C., Tagle, D., Barnes, G., Bates, G., Taylor, S., Allitto, B., and Altherr, M. (1992). The Huntington's disease candidate region exhibits many different haplotypes. *Nature Genet.* **1,** 99–103.
52. Rubinsztein, D. C., Leggo, J., Barton, D. E., and Ferguson-Smith, M. A. (1993). Site of (CCG) polymorphism in the HD gene. *Nature Genet.* **5,** 214–215.
53. Andrew, S. E., Goldberg, Y. P., Theilmann, J., Zeisler, J., and Hayden, M. R. (1994). A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene, implications for diagnostic accuracy and predictive testing. *Hum. Mol. Genet.* **3,** 65–67.
54. Leeflang, E. P., Tavare, S., Marjoram, P., Neal, C. O. S., Srinidhi, J., MacFarlane, H., MacDonald, M. E., Gusella, J. F., de Young, M., Wexler, N. S., and Arnheim, N. (1997). Dynamic mutation rates at the Huntington's disease locus. [Unpublished data]
55. Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., Folstein, S., Ross, C., Franz, M., Abbott, M., Gray, J., Conneally, P., Young, A., Penney, J., Hollingsworth, Z., Shoulson, I., Lazzarini, A., Falek, A., Koroshetz, W., Sax, D., Bird, E., Vonsattel, J., Bonilla, E., Alvir, J., Bickham Conde, J., Cha, J.-H., Dure, L., Gomez, F., Ramos, M., Sanchez-Ramos, J., Snodgrass, S., de Young, M., Wexler, N., Moscowitz, C., Penchaszadeh, G., MacFarlane, H., Anderson, M., Jenkins, B., Srinidhi, H., Barnes, G., Gusella, J., and MacDonald, M. (1993). Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genet.* **4,** 387–392.
56. Trottier, Y., Biancalana, V., and Mandel, J. L. (1994). Instability of CAG repeats in Huntington's Disease: relation to parental transmission and age of onset. *J. Med. Genet.* **31,** 377–382.
57. Wexler, N. S., Rose, E. A., and Housman, D. E. (1991). Molecular approaches to hereditary diseases of the nervous system: Huntington's disease as a paradigm. *Annu. Rev. Neurol.* **14,** 503–529.
58. Jansen, G., Willems, P., Coerwinkel, M., Nillesen, W., Smeets, H., Vits, L., Howeler, C., Brunner, H., and Wieringa, B. (1994). Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in (CTG)n repeat variation and selection against extreme expansion in sperm. *Am. J. Hum. Genet.* **54,** 575–585.
59. Wierdl, M., Dominska, M., and Petes, T. D. (1997). Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* **146,** 769–779.
60. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., and Inouye, M. (1966). Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31,** 77–84.
61. Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., *et al.* (1993). Clues to the pathogenesis of familial colorectal cancer. *Science* **260,** 812–816.
62. Peltomaki, P., Aaltonen, L. A., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Green, J. S., Jass, J. R., Weber, J. L., Leach, F. S., *et al.* (1993). Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **260,** 810–812.
63. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363,** 558–561.
64. Strand, M., Prolla, T. A., Liskay, R. M., and Petes, T. D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365,** 274–276. [Published erratum appears in *Nature* **368**(6471), 569, 1994]
65. Modrich, P., and Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination and cancer biology. *Annu. Rev. Biochem.* **65,** 101–133.

66. Mitas, M. (1997). Trinucleotide repeats associated with human disease. *Nucleic Acids Res.* **25,** 2245–2253.
67. Richards, R. I., and Sutherland, G. R. (1994). Simple repeat DNA is not replicated simply. *Nature Genet.* **6,** 114–116.
68. De Wind, N., Dekker, M., Berns, A., Radman, M., and te Riele, H. (1995). Inactivation of the mouse MSH2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell* **82,** 321–330.
69. Drost, J. B., and Lee, W. R. (1995). Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. *Environ. Mol. Mutagen.* **25,** 48–64.