Alec J. Jeffreys
Maxine J. Allen
John A. L. Armour
Andrew Collick
Yuri Dubrova
Neale Fretwell
Tara Guram
Mark Jobling
Celia A. May
David L. Neil
Rita Neumann

Department of Genetics, University of Leicester

Mutation processes at human minisatellites

Minisatellites provide one of the most experimentally tractable systems for studying tandem repeat instability in man. Analysis of mutation processes has been greatly aided by the development of single molecule methods for recovering *de novo* mutants, and of techniques for exploring allele structure in detail. Application of these approaches to man has shown that minisatellites do not primarily mutate by processes such as replication slippage and unequal crossover intrinsic to the tandem repeat array. Instead, germline repeat instability is largely regulated by *cis*-acting elements near the array and involves unexpectedly complex processes of gene conversion, of potential relevance to the biology of meiosis. These processes can be explored both in humans and, in principle, in transgenic mouse models of human repeat instability.

1 Introduction

Minisatellites are tandem repeat loci typically 0.5–30 kbp long with repeat units in the range 6–100 bp. Thousands of minisatellites exist in the human genome, preferentially located near the ends of chromosomes, and frequently show variability in repeat copy number and therefore allele length. The extreme informativeness of the most variable minisatellites has led to their widespread use for identification and parentage analysis in forensic and legal medicine, using both multi-locus DNA fingerprinting and single-locus DNA profiling approaches. While much is known about allele length variation at human minisatellites, and how this variation can be exploited in DNA typing (see [1]), we are still largely ignorant of the mutation processes that generate this variability and maintain it in human populations. To explore these mechanisms, we and others are using minisatellites as a test-bed for developing new approaches for the detection and characterization of *de novo* mutation events in the human germline. One particular interest is to determine which of the "classic" models for tandem repeat instability, such as unequal exchange [2] and replication slippage [3], are involved in minisatellite mutation, and whether direct evidence can be adduced to support earlier speculations that minisatellites may play a role in meiotic recombination [4]. Tandem repeat instability is also of profound interest in relation to the phenomenon of triplet repeat expansion now known to be involved in seven different neurological disorders including Fragile-X mental retardation, Huntington's disease and myotonic dystrophy (see [5, 6]); again, our understanding of triplet expansion processes is still far from complete.

Knowledge of mutation rates and processes is also of relevance to forensic DNA typing, not only in paternity testing where paternal mutation could lead to a false exclusion, but also in fully understanding the population genetics of loci where genetic drift is substantially counteracted by recurrent mutation, and in assessing the potential for somatic mutation to lead to divergent genotypes in different tissues of the same individual.

Correspondence: Professor Sir A. J. Jeffreys, Department of Genetics, University of Leicester, Leicester, LE1 7RH, UK

Nonstandard abbreviations: MVR, minisatellite variant repeat; **SP-PCR**, small pool-polymerase chain reaction

Keywords: Minisatellite / Mutation / Recombination / Conversion / Sperm

2 Minisatellite mutation rates in pedigrees

The first indication that some minisatellites can show remarkably high germline mutation rates to new length alleles, arising from spontaneous changes in repeat copy number, came from multi-locus DNA fingerprint analysis of human families [4, 7] and recombinant inbred mouse strains [8] which revealed the frequent appearance of new mutant DNA fragments in offspring. Analysis of single human minisatellite loci showed mutation rates as high as 5% per gamete for the most variable minisatellite MS1, with mutation rates at different loci increasing with allele length heterozygosity in accordance with the predictions of the neutral mutation/random drift model [9]. At some loci such as MS1, paternal and maternal mutations arise with similar frequency, despite differences in numbers of mitotic cell divisions in the germline leading to sperm or oocytes. At other loci, evidence is accumulating that mutations preferentially arise in the male germline; the most extreme example to date is locus CEB1 with a mutation rate of 15% per sperm but only 0.3% per oocyte [10]. There is thus no clear correlation of mutation rate with germinal cell turnover, as might be expected for a mitotic mutation process. It is also not clear why mutation rates can vary by orders of magnitude between different loci; suggestions that instability may be promoted by factors such as short repeat units, high copy number arrays and repeat unit sequence homogeneity along the array [11] are contradicted by unstable loci such as CEB1, which have relatively long repeat units, low copy number arrays and extensive sequence diversity between repeat units in a single array [10, 12].

3 Somatic mutation

Minisatellite mutation events are not restricted to the germline but also occur in somatic tissues. However, standard hybridization analysis will only reveal such mutants if the tissue is partially or completely clonal for the new mutant allele, as can occur in cell lines and tumours [13]. Mutational clonality in lymphoblastoid cell lines derived from human pedigrees can lead to

0173-0835/95/0909-1577 $5.00+.25/0

somatic mutation being mis-scored as germinal in origin, as has also been noted for microsatellite mutation [14]. Mutational mosaicism can also arise in normal tissue as a result of early embryonic mutation, to generate somatic DNA with three, not two, alleles detectable by hybridization. Such events at minisatellites are extremely rare in humans (only one has been seen to date) but common at two mouse minisatellites [15–18]. Detailed analysis of mutant allele dosage and the segregation of mutant cells in embryonic and extra-embryonic tissues has suggested that many of these somatic mutation events in mice are confined to a very early developmental window of instability, during the first two cell divisions following fertilization [19]. However, these unstable mouse loci both consist of long arrays of very short repeat units (GGCA or GGGCA), unlike the human loci studied, and it is possible that these mouse loci mutate via pathways distinct from those operating at human minisatellites but more analogous to postzygotic triplet repeat instability, seen for example in Fragile-X mental retardation [20] and myotonic dystrophy [21].

4 No unequal exchange

If minisatellites mutate to new length alleles by unequal exchange between alleles, then new mutant alleles should be recombinant for flanking DNA markers. Analysis of limited numbers of germline mutants at three different loci has failed to detect an exchange of flanking markers [10, 22, 23], suggesting that inter-allelic unequal exchange is not the dominant mode of germline mutation at human minisatellites. Similarly, analysis of tumour DNAs and cell lines carrying somatic mutations has again excluded unequal mitotic exchange between alleles as a major process [13]. These studies unfortunately do not rule out unequal exchange as a contributory mechanism, nor do they clarify which of a host of alternative processes might be involved in minisatellite instability. Further investigation of minisatellite mutation required the development of methods for investigating allelic structure in detail before and after mutation, and of techniques capable of recovering unlimited numbers of new mutant alleles from any source of human DNA, rather than the limited numbers identifiable by pedigree, cell line and tumour DNA analysis.

5 Internal allelic structure

All human hypervariable minisatellites characterized to date vary not only in repeat copy number (allele length) but also in the interspersion pattern of variant repeat units within the array (Fig. 1A) [24]. This internal variation provides a powerful approach to the study of allelic variation and processes of mutation. Interspersion patterns can be determined by minisatellite variant repeat (MVR) mapping by PCR [25]. MVR-PCR uses a PCR primer at a fixed site in the DNA flanking the repeat array, together with primers specific for different repeat variants, to produce ladders of PCR products extending from the flanking DNA to each repeat unit of a given type. MVR distribution patterns can be determined by subsequent electrophoretic analysis of these MVR-PCR products. If only one site of repeat unit sequence variation is targeted, then MVR-PCR can be used to distinguish just two types of repeat, to generate binary codes of the two repeat types interspersed along an allele ("two-state" MVR-PCR). More internal structural information can be recovered by analysing additional sites of variability, if they exist, with appropriate MVR-PCR primers. For example, the 29 bp repeat unit of minisatellite MS32 contains two base substitutional polymorphic sites separated by 1 bp; simultaneous analysis of both sites by "four-state" MVR-PCR generates a quaternary code from an allele and doubles the information recoverable by analysis of either site alone [26].

MVR-PCR analysis of total genomic DNA generates PCR products from both alleles simultaneously, to produce extremely variable diploid digital codes of considerable potential use in forensic identification [24, 25]. For allelic diversity studies, however, it is necessary to determine codes from individual alleles. This can be achieved in three ways: first, by physical separation or differential PCR amplification of alleles of different length prior to MVR-PCR [25]; second, by deducing parental single-allele codes from the diploid codes of parents and their offspring [25]; and third and most simply, by using allele-specific flanking PCR primers targeted to known sites of base substitutional polymorphism in the DNA flanking the minisatellite so as to produce MVR-PCR products from only one allele in an appropriate flanking heterozygote [27].

Single-allele coding has been most extensively carried out on minisatellite MS32 [25]. Two-state MVR codes have been established for more than 1100 alleles to date and have allowed a detailed analysis of allelic variability at this locus. Allelic diversity so revealed is far greater than can be distinguished by standard allele length analysis; approximately 50 different length alleles can be resolved at MS32, compared with an estimated 100 000 000 or more different alleles distinguishable by MVR-PCR in the global human population. While almost all alleles so far typed at MS32 are different, it is possible to compare allele codes to identify groups of alleles that are closely related and which have therefore diverged from a recent common ancestral allele (Fig. 1B). Such allele groups show a curious phenomenon, namely that most variability between alleles in repeat copy number and MVR code is restricted to the beginning of the repeat array [25, 28]. This indicates a polarity of variability along the array and implies the existence of a terminal mutation hotspot. Similar polarity has also been seen at three other minisatellites and may be a general feature of VNTR loci [29–31]. Such polarity would not be predicted by classic processes of slippage or unequal exchange. Analogous polarity has also been observed at an unstable tetranucleotide repeat on the X chromosome [32], and also in the Fragile-X $(CGG)_n$ repeat array such that most variation is confined to one end where all repeats are homogeneous and not disrupted by AGG variants which can occur in the less variable segment of the array [33, 34]. For Fragile-X, it is suggested that repeat homogeneity promotes instability, for example by slippage, and thus creates polar variation. This does not apply to minisatellite polarity, where the

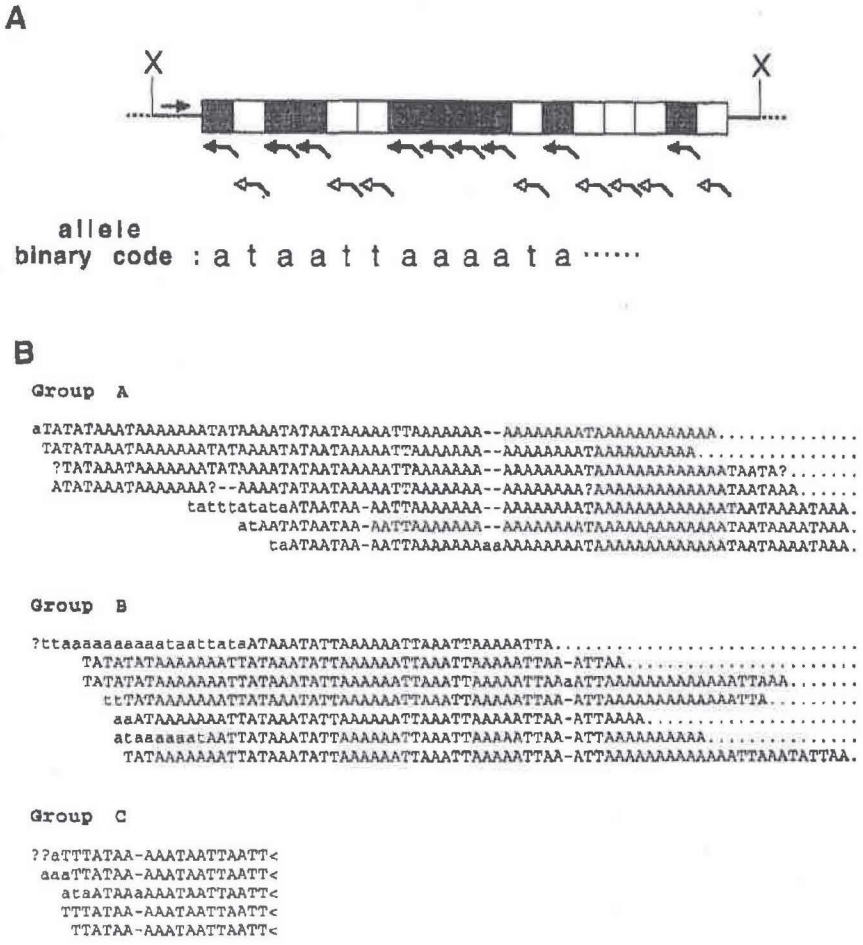

Figure 1. Minisatellite allele analysis by MVR-PCR. (A) The principles of MVR-PCR. DNA profiling assays minisatellite allele length variability by digestion with a restriction enzyme X which cleaves in the genomic DNA flanking the tandem repeat array, followed by gel electrophoresis and Southern blot hybridisation. In contrast, MVR-PCR assays allelic variability in the interspersion pattern of variant repeats (white, shaded boxes) within an allele. Two-state MVR-PCR uses a primer at a fixed site in the DNA flanking the minisatellite, plus primers specific for one or other repeat type, to generate PCR products extending from the flanking DNA to each repeat of a given type, from which the allelic binary code of repeat types (a-, t-type) can be read. (B) Examples of groups of closely related MS32 alleles discovered in a sample of 1100 alleles mapped by two-state MVR-PCR. Related alleles are aligned with gaps introduced to improve alignments. Repeats shared amongst related alleles are shown in uppercase. Alignable groups are usually population-specific; thus Group A alleles are all Caucasian in origin, Group B all African and Group C all Japanese. Note that most variability is concentrated at the beginning of the alleles, even within the very short (18–22 repeat) alleles of Group C.

most variable part of the array contains normal levels of repeat unit sequence heterogeneity (Fig. 1B).

Groups of closely related or identical minisatellite alleles usually share a common haplotype of flanking DNA variants, suggesting that whatever mutation processes operate, they do not normally involve recombinational exchanges of flanking DNA. This confirms that inter-allelic unequal exchange is not the major mutational process. However, related groups do frequently show evidence of flanking haplotype switching [30] although it is difficult to deduce reliable rates of exchange from these population data, and impossible to determine whether these exchanges occur as a result of tandem repeat mutation or, instead, represent recombination or conversion events operating strictly in the flanking DNA.

6 New approaches to mutation detection

Mutation processes cannot be reliably deduced from population data, and can only be investigated by direct structural analysis of new mutant alleles. The recovery of mutants from pedigrees is too inefficient for detailed mutation analysis, and further can give no information about individual mutation rates. Pedigree analysis, of course, uses offspring to determine whether either parental gamete carried a mutation. Given the potential for PCR to analyse DNA at the single molecule level, an alternative is to detect *de novo* mutations directly in single gametes. Since 10^8 or more sperm can be recovered from a man, equivalent to 10^8 offspring, unlimited numbers of mutants could be identified and characterized. This approach is not appropriate for studying mutation in the female germline, in view of the limited availability of human oocytes, but can be easily adapted to the detection of mutants in bulk somatic DNA.

We have developed two single-molecule PCR approaches for detecting new mutations in sperm and somatic DNA. The first approach uses gel electrophoretic enrichment of minisatellite mutants of abnormal length, compared with progenitor alleles, from bulk sperm or somatic DNA prior to recovery of individual mutant molecules by PCR [28]. For technical reasons, this approach is most effective for recovering large deletion mutants and can be used to detect mutants at a frequency as low as 10^{-6}/cell. However, the sperm deletion mutants recovered by this method from minisatellite MS32, for which the tech-

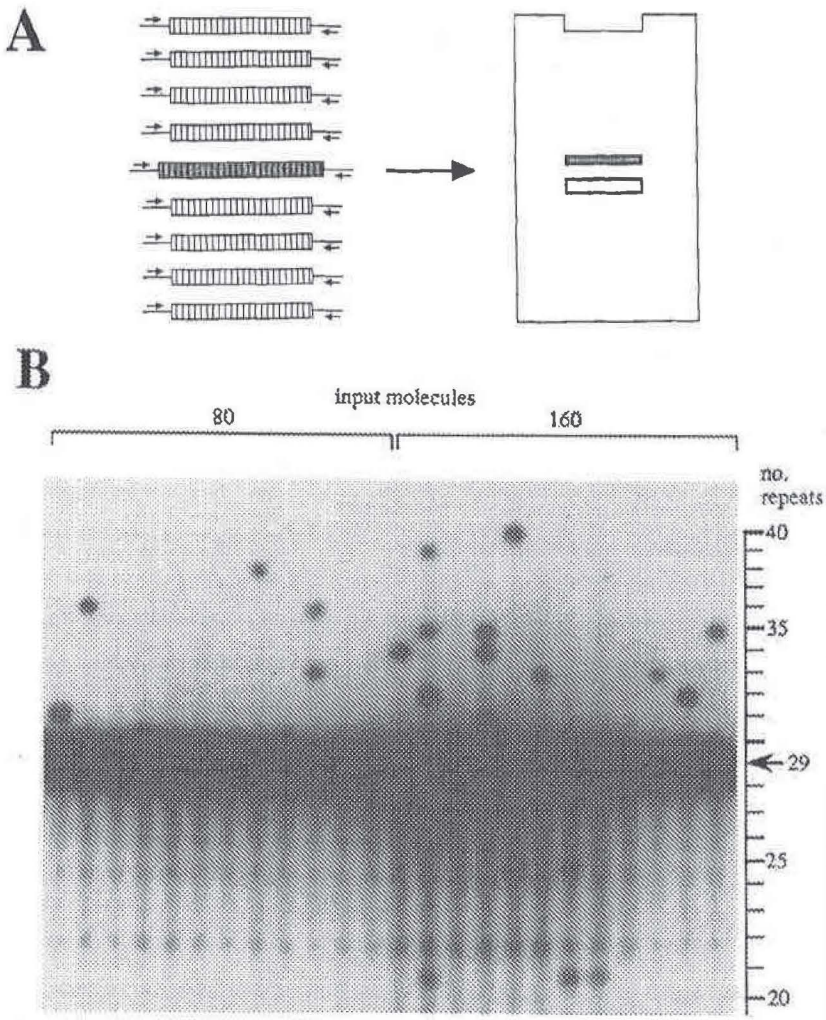

Figure 2. Minisatellite mutation detection by small pool PCR. (A) The principles of SP-PCR. Genomic DNA is diluted until an aliquot contains a limiting number of minisatellite molecules. If a mutant molecule (shaded) is present, it will now make up a significant proportion of all molecules. Each molecule is PCR amplified using primers in the flanking DNA (arrows). Products of the mutant molecule can now be resolved from the progenitor allele (white) by agarose gel electrophoresis and detected by Southern blot hybridisation (right). (B) Example of sperm mutation detection at MS32 by SP-PCR on multiple aliquots of dilute sperm DNA. The individual tested was heterozygous for a 29-repeat allele plus a 200-repeat allele too large to be amplified in SP-PCR. Mutant alleles are detected as PCR products different in size from the 29-repeat progenitor, which shows an intense signal.

nique was developed, are now known to be rare and highly atypical of the bulk of *de novo* mutations arising in the male germline.

The second single-molecule approach is technically simpler and involves PCR amplification of the entire allele from multiple dilute aliquots of sperm DNA (small pool PCR, SP-PCR; Fig. 2; [35]). Since the haploid genome of man contains 3 pg DNA, a small pool of 300 pg sperm DNA will contain 100 molecules of a given locus (50 per allele). For a locus such as MS32 with a mutation rate of 1%/gamete, each small pool will contain on average one mutant molecule. The PCR products from the mutant will therefore make up 1% of the total products following SP-PCR, an amount which enables the mutant PCR products to be readily detected by gel electrophoresis and Southern blot hybridization. Analysis of multiple pools of DNA enables 10000s of sperm to be surveyed for mutation in a single experiment. SP-PCR can only be applied to alleles short enough to be amplified efficiently in their entirety (<5 kbp); recent developments in long-range PCR [36, 37] may make it possible to analyse longer alleles for mutation, though electrophoretic resolution of mutants from their progenitors will remain problematical.

SP-PCR has been extensively validated at MS32, in particular to determine whether abnormal length PCR products are authentic mutants rather than artefacts arising during PCR [35]. Various lines of evidence indicate authenticity, including the quantal nature and appropriate intensity of mutant sperm PCR products, the correct proportionality between input DNA and yield of mutants, the much lower frequency and different spectrum of mutants detected by corresponding analyses of somatic (blood) DNA, and the curious structures of new mutant alleles (see below) which are incompatible with PCR artefacts. Mutation rates as low as 10^{-4}/sperm can be reliably measured by SP-PCR; below this level, PCR artefact noise will progressively impede mutation detection. SP-PCR is also being adapted to other minisatellite loci, and has been applied to the $(CAG)_n$ triplet repeat array in myotonic dystrophy to enable the heterogeneous smear of new mutant alleles detected in the sperm and somatic DNA of affected individuals to be resolved into individual mutant molecules [38].

7 Sperm mutation rates at MS32

SP-PCR has now been used to measure sperm mutation rates directly in numerous men [35]. The mean mutation rate is 0.8% per sperm, as predicted from pedigree analysis and suggesting that SP-PCR is not being heavily biased by sampling mutation in abnormal sperm that cannot contribute to the next generation. Curiously, the sperm mutation rate per allele does not increase with allele length (over the sampled range of 22–164 repeats), a result which appears to be incompatible with mutation processes intrinsic to the repeat array, where the mutation rate should increase with target size (array length). Most mutation events involve the gain or loss of one or a few repeat units, with the size distribution of mutant alleles relative to the progenitor being apparently constant, and again independent of allele length (Fig. 3). Most of these mutants are specific to sperm in that they are not seen in blood DNA (Fig. 3). Sperm mutation at MS32 shows a remarkable 3:1 bias in favour of gains rather than losses of repeat units; both alleles in a man show the same bias, indicating an asymmetric mutation process rather than one allele gaining repeats at the expense of the other. Pedigree analysis at other minisatellites has also provided evidence for expansion bias in the male germline, suggesting that this may be a common feature of VNTR loci [10, 35]. The scale of the bias at MS32 has major population genetic implications; in particular, array lengths will not drift in a stochastic fashion but will increase deterministically at a mean rate of 1 repeat added per 1000 years. A major question therefore is what forces act to prevent such loci from expanding indefinitely. There are various possibilities, including counteracting deletions in the female germline (although such deletions have not yet been seen at MS32), increased rate of large deletions as alleles grow in length, truncating selection acting on chromosome dysfunction induced by long arrays, failure of very long alleles to engage in the mutation process, and the population spread of mutations that reduce or eliminate the mutational competence of minisatellites (see below). Which of these processes operate to prevent minisatellites from swamping the genome is as yet unknown.

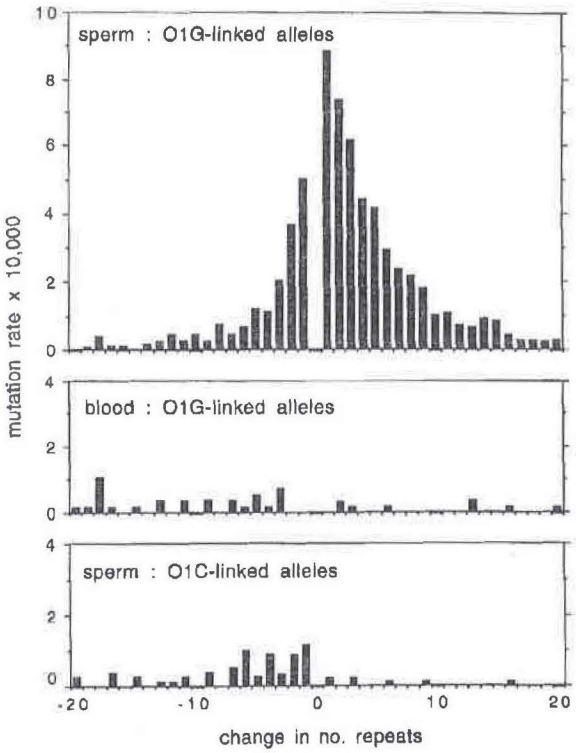

Figure 3. Germline and somatic mutation rates to new length alleles at MS32 as measured by SP-PCR. Rates are given as the frequency per progenitor molecule of mutant molecules of a given size class relative to the progenitor. Top, mean distribution for 25 different O1G-linked alleles in sperm, determined from 180 000 progenitor molecules. Middle, mean distribution for 4 different O1G-linked alleles in blood (56 000 progenitor molecules tested). Bottom, mean distribution for 7 different O1C-linked alleles in sperm (90 000 molecules tested).

8 Complex conversion events

MVR-PCR structural analysis of new mutant alleles identified in pedigrees or recovered by SP-PCR analysis of sperm DNA has revealed an unexpectedly complex and bizarre mutation process operating at MS32 in the male germline (Fig. 4) [35]. As predicted from allele diversity studies, most mutation events are extremely polar, involving repeat unit changes at the extreme end of the array over the region previously identified as a variability hotspot. In at least half of the sperm gain mutants, mutation involves the transfer of repeats or repeat unit blocks from one allele (the "donor") into the other allele (the "recipient"). These inter-allelic transfers are most obvious when the donor allele contains MVR types not present in the recipient (Fig. 4). Transfer is frequently accompanied by complex rearrangements in the recipient allele, including "target site" duplications in repeat blocks flanking, or adjacent to, the site of insertion, and occasional multiple rounds of imperfect amplification of repeat blocks at or near the site of insertion to create mutant alleles much longer than the recipient and in which the beginning of the mutating allele is profoundly remodelled. In some cases, scrambling of MVR types can be so extensive as to create repeat segments in the recipient allele which have no obvious origin in either parental allele. Inter-allelic transfer usually, but not always, involves the acquisition of repeat unit blocks from the corresponding region of the donor allele, implying that alleles of different lengths are paired at their 5′ end prior to transfer. Even closely-flanking DNA markers are seldom cotransferred during inter-allelic transfer, though one example has recently been discovered by pedigree analysis of minisatellite MS31 (D.L.Neil and A. J. Jeffreys, unpublished). This suggests that the mutation process is largely restricted to the repeat array itself. Some sperm mutants do not involve inter-allelic transfer but can nevertheless show polarity and complex rearrangements compatible with an analogous mutation process involving transfer between sister chromatids rather than alleles. There is evidence that the balance between inter- and intra-allelic events, and the degree of polarity of intra-allelic mutation, can vary between loci [12].

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.