Detecting copy number changes in genomic DNA - MAPH and MLPA

White, S.J.

Chapter 1

Introduction

General Introduction.

(Different sections of the introduction and discussion are based on the chapter "Detecting copy number changes in genomic DNA - MAPH and MLPA", White S.J., Breuning M.H., and den Dunnen J.T. In Cytometry 4ed, Volume 75 of the *Methods in Cell Biology*, in press).

It has long been recognized that some diseases run in the family. A classic example of this is hemophilia, which affected several European royal families descended from the English monarch Queen Victoria. Following the discovery of DNA as the hereditary material it became clear that it was changes in DNA that were responsible for many diseases, some inherited and some that appeared to have arisen spontaneously.

In principle there are 3 ways that the DNA within a genome can be changed. Qualitative alterations involve changes of one or more nucleotides in a given stretch of DNA, either by substitution or chemical modification. Positional rearrangements occur when a sequence moves from one specific location to another, for example translocations and inversions. The last class of changes are the quantitative rearrangements, involving the gain or loss of segments of DNA.

The purpose of the research described in this thesis is the development and optimization of techniques for detecting such duplications and deletions in genomic DNA. Many different methods for detecting copy number changes have been described (reviewed in (1)). The most commonly applied techniques to date have been FISH (Fluorescent In Situ Hybridization), Southern blotting (sometimes combined with Pulse-Field Gel Electrophoresis (PFGE)), and quantitative PCR. FISH combines hybridization of a fluorescently labeled probe with microscopic analysis of genomic DNA (2-4). Because of its relative reliability and ease of analysis, the technique has been widely used in diagnostic laboratories. Its main disadvantages are the workload involved, i.e. culturing cells, making chromosome preparations and performing hybridizations, and the fact that deletions smaller than ~40 kb cannot be routinely detected.

As an alternative, Southern blotting has also been broadly applied for the detection of deletions and duplications (5,6). In addition one has the possibility to scan for the unique junction fragments of the rearrangements and use these for diagnosis (7). Although this latter possibility requires some effort to identify a junction fragment, when found it does provide a powerful diagnostic tool. In this respect, the combination of PFGE and Southern blotting can be very informative, facilitating the use of a nearby probe to detect all rearrangements from a distance

(7,8). The problem with Southern blotting is that it is time consuming, and producing blots of the necessary quality is technically demanding.

Quantitative PCR has been tried by many (9,10) but has also turned out to be challenging and difficult to implement. Recently, the development of real-time PCR technology revived the diagnostic application of quantitative PCR (11-13) but it failed to take away another drawback, the difficulty of multiplexing, i.e. measuring copy number changes of several DNA sequences in parallel.

The development of new diagnostic methods is made easier by having a series of samples with known mutations to test. Within the department of Human and Clinical Genetics in Leiden several hundred DNA samples from patients suffering from Duchenne Muscular Dystrophy have been collected over the years. As is made clear in the next section, there are several features of the mutation spectrum of this disease that make it ideal for optimizing new quantitative methods.

A. Duchenne Muscular Dystrophy

In 1852 the English physician Meryon described nine boys affected by muscle wasting and weakness (14). He noted that it appeared to be of muscular rather than nervous origin, and described the degeneration of the voluntary muscular tissue. The same form of disease was discussed further by Duchenne in 1868 (described in (15)). He wrote about a disorder affecting primarily males, often familial and characterized by weakness of the lower musculature from an early age. This would spread to the upper body, and death from cardiac or respiratory failure usually occurred before the age of 20. This disease is now known as Duchenne Muscular Dystrophy (DMD).

The fact that it affected almost exclusively males meant that the causative gene was probably on the X chromosome. Ironically enough it was the extremely rare cases affecting females that first provided the clue as to where the gene was on the X chromosome. If a female had a translocation disrupting the gene affected in muscular dystrophy, the unaffected X-chromosome becomes inactivated and there will be no product from the DMD gene. This individual would then show all the manifestations of DMD. Combined analysis of such translocation cases (16-18), along with a cytogenetically visible deletion in a male patient (19), and restriction fragment length polymorphisms (RFLPs) (20,21) defined the breakpoint to a region on the short arm of the X chromosome, within the band Xp21. Further mapping defined the location of the gene more precisely (22), and the cDNA was eventually cloned in 1987 (5), and sequenced in 1988 (23).

DMD Gene

The gene remains the largest known, covering ~2.4 Mbases (24,25). It is composed of 79 exons (26), and there are at least 7 alternative promoters, leading to a number of different isoforms (Figure 1). The exons make up <1% of the total gene size, and the introns vary greatly in size, from <1 kb to ~250 kb.

Dystrophin

The protein product of the *DMD* gene is known as dystrophin. The full-length product from the muscle promoter has a molecular weight of 427 kDa (27), and is found primarily in skeletal and cardiac muscle. Other isoforms also show specific patterns of expression, and can be found in several other tissues (28-31).

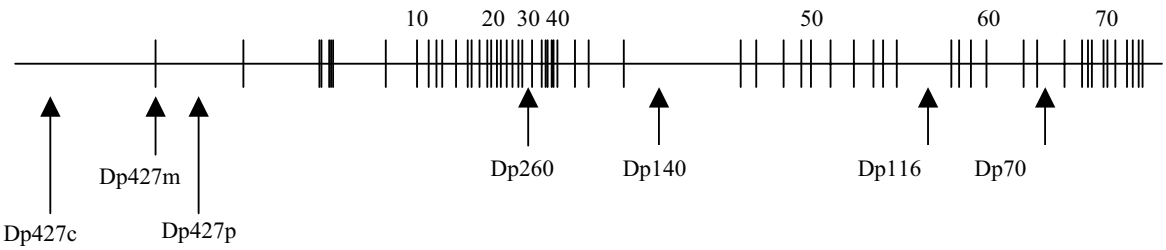

Figure 1. The *DMD* gene. In this diagram the promoter for the muscle isoform (Dp427m) is shown as exon 1, and the locations of the other promoters are indicated with arrows. Every 10th exon is numbered.

The amino terminal domain shows homology to the amino terminus of alpha-actin, and binds to the actin cytoskeleton (32). The central rod domain is the largest part of the protein, and is composed of a series of coiled-coil repeats. This is followed by a cystein-rich domain, and finally a carboxy terminal domain that interacts with several different proteins in the dystrophin-glycoprotein complex (33). The primary function of the protein appears to be structural, linking the cytoskeleton with the extracellular matrix. There is recent evidence that dystrophin may also be involved in different cell signalling pathways (34,35).

Deletion/duplication detection in patients

Cytogenetically invisible deletions were first detected with probes derived either randomly (36), from a region known to be missing in a patient (37), or a translocation breakpoint (38,39). Prior to the cloning of the cDNA it was shown by Pulse-Field Gel Electrophoresis (PFGE) that gross deletions and duplications could be found in >50% of DMD cases (7). By using cDNA fragments as a probe it was possible to examine the entire gene for both deletions and duplications using Southern blotting (40). This and other work showed that a deletion of one or more exons occurred in ~65% of cases, with the duplication frequency being estimated at 6% (41,42). The deletions clustered in 2 regions; a major hotspot between exons 45-52 and a minor hotspot between exons 2-19 (41,43,44). This clustering allowed for ~95% of all deletions in patients to be detected using 2 multiplex PCR kits, each amplifying 9 exons (45,46). These primer sets were later expanded and modified using fluorescently-labelled primers and dosage analysis, allowing duplications to be detected as well (9,10).

More recently, MAPH probe sets have been developed covering each of the 79 exons (47) (Chapter 2.1). Analysis of >100 samples showed a wide range of mutations, including several

that had not been previously detected using other techniques. Noteworthy here is that a duplication of exon 2 was the most common duplication found, yet had never been reported using alternative methods e.g. Southern blotting.

The ability to precisely determine the breakpoints led to an explanation for the difference in severity seen between DMD and Becker Muscular Dystrophy (BMD). BMD is milder, with the patients generally showing a later age of onset, slower disease progression and increased life expectancy. There are many cases where the deletion of a single exon is sufficient to cause DMD, whereas deletions of nearly half the gene have been found in BMD patients (48,49). Analysis of the breakpoints of the deletions lead to the development of the reading frame hypothesis (50). The theory was that mutations that disrupted the reading frame lead to a truncated protein, which could not carry out its physiological role of connecting the intracellular skeleton with the extracellular matrix. This destabilized the muscle cell, leading to all the symptoms of the disease. In contrast, an in-frame deletion would shorten the protein, but the presence of both the N-terminus and C-terminus would ensure that the protein would be at least partly functional, connecting the cytoskeleton and extra-cellular matrix.

Based on DNA analysis only, this theory is valid in about 90% of cases. Exceptions have been described, both for DMD patients with in-frame deletions and BMD with out-of-frame deletions. These are most commonly found at the 5' end of the gene (51). In-frame mutations causing DMD are either large or remove a domain essential for dystrophin function, whereas out-of-frame mutations causing BMD are presumably compensated for, either by the use of an alternative start codon (52), or exon skipping leading to an in-frame product (53,54).

Other mutations

In ~30% of cases no exonic deletion or duplication can be detected. The most common mutation in this subset of patients are changes of one or a few nucleotides leading to a truncated transcript, either via a premature stop codon or the disruption of the reading frame. Missense mutations have been rarely described (55-57). Given the large size of the gene and the fact that there are no apparent hot spots for point mutations, finding them is a time-consuming process. Screening for these changes has been performed by a number of different techniques, including heteroduplex analysis (58,59), denaturing high-performance liquid chromatography (DHPLC) (60), single strand conformation polymorphism (SSCP) (61), denaturing gradient gel electrophoresis (DGGE) (62) and direct sequencing (63).

With the advances in screening technology it is now possible to detect the causative mutation in ~96% of patients by screening at the DNA level (64). The use of RNA-based techniques, such as RT-PCR and the protein truncation test (PTT) (65,66), afford the possibility of detecting changes located deep within introns that activate cryptic splice sites (67,68), further increasing the detection rate. The disadvantages of using RNA relate primarily to the source. The amount of the dystrophin transcript within lymphocytes is very low, while the best mRNA source, a muscle biopsy, is not always easy to obtain.

There are several possible reasons as to why no mutations are found in the remainder of the patients. It may be that some patients do not have DMD, but instead suffer from another neuromuscular disorder. If the absence of dystrophin has not been confirmed it may be difficult to distinguish DMD from a severe form of limb-girdle muscular dystrophy purely on the clinical symptoms. Another, less likely, possibility is that there is a mutation in another gene that influences the regulation of dystrophin expression.

Knowing the mutation is not only important for confirming the diagnosis, but may also be a critical first step towards therapeutic intervention. Several different gene therapy strategies have been described for DMD (reviewed in (69)). For the mutation-specific approaches e.g. exon skipping (70,71), it is necessary to know which exon/exons need to be targeted.

Carrier detection

Finding the mutation is not only of importance for the patient. It also facilitates the genetic counseling of relatives who may be carriers. Prior to genetic testing this was primarily based on biochemical assays, of which elevated serum creatine kinase was the most common (72). The discovery of the *DMD* gene meant that a DNA-based test was possible, and most of the techniques applied for mutation screening in males are also applicable for testing females (73). In the early years of genetic testing it was not always possible to find the causative mutation in the patient. This made the determination of carrier status particularly difficult, often relying on haplotype analysis (74-77).

Identifying the mutation in the index patient simplifies carrier detection, as the entire gene no longer needs to be screened. Heterozygosity for a point mutation can be determined by sequencing. Different approaches have been described for determining the carrier status in cases involving a deletion or duplication (78,79). They are based around the analysis in the maternal DNA of an exon or exons known to be affected in the patient, and comparing these with a region known to be unaffected.

An important point to be considered in carrier testing is the possibility of somatic or germ line mosaicism (80-83). In a somatic mosaic case the mutation may be present in only a percentage of maternal cells. A negative result may be concluded, as the normal cells have masked the mutation present in the affected cells. Germ-line mosaicism will not be detectable in DNA isolated from blood cells. The frequency of mosaic cases has been estimated to be up to 14% (84). The high recurrence risk for apparently de novo mutations means that prenatal screening is still offered for females that have tested negative for carrier status.

How do deletions and duplications occur?

DMD is an unusual disease in that the majority of mutations found are deletions or duplications of one or more exons. Such rearrangements are also found in other diseases, but at a lower frequency. A great deal of effort has therefore gone into studying the mechanisms involved in these aberrant recombination events. It is known that the precursor for recombination is the double stranded break, or DSB. This occurs during normal cellular processes, for example recombination in meiosis (85,86) and V(D)J recombination in antibody processing (87). Less beneficially, it can also be caused by damaging environmental factors, such as ionizing radiation or free radicals.

When a DSB occurs, an attempt will be made by the cell to rejoin the free ends of the DNA molecule. Two main pathways are known to be involved (88,89), namely homologous recombination and non-homologous end joining. Both pathways are highly conserved from bacteria to humans, indicating the fundamental importance of the process (90,91).

Homologous recombination

Homologous recombination (HR) is a process that uses an intact template for the repair of a DSB. Once a DSB is recognized, a protein complex at the site is formed, and there is a partial 5' to 3' resection at the ends. Following strand invasion into an intact, homologous DNA molecule, an extension reaction by a DNA Polymerase fills in the gap. Precisely how the different strands are resolved will determine whether or not crossing over and/or gene conversion will occur (88).

Non-allelic homologous recombination (NAHR) can occur when two highly similar sequences misalign. Following a DSB the two homologous sequences are aligned, and crossing over occurs. This can lead to deletions and duplications (Figure 2).

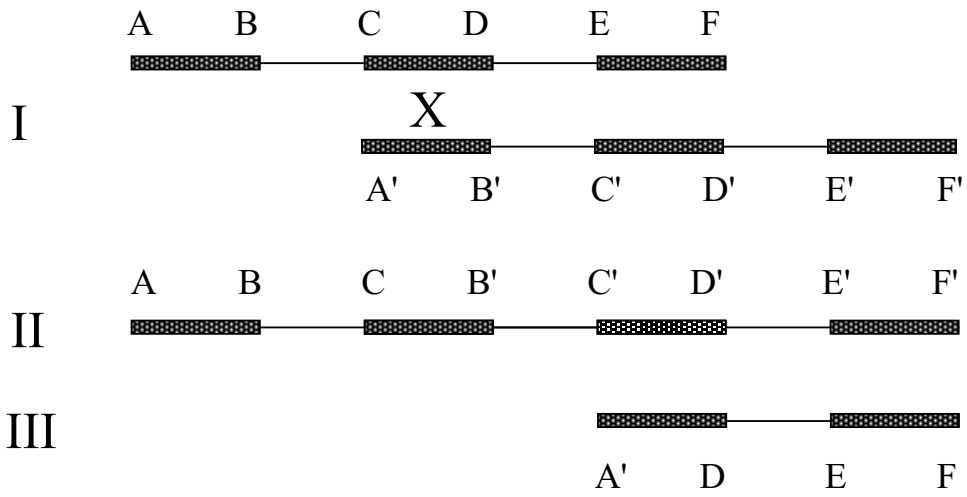

Figure 2. Non-allelic Homologous Recombination. Misalignment of highly similar sequences (I) can lead to non-equal crossing over, resulting in (II) a duplication and (III) the reciprocal deletion.

It is known that repetitive elements play an important role in NAHR. When the human genome sequence was published in 2001, it was stated that >50% of the human genome was composed of such sequences (92,93). These can be separated into different classes, with the majority being originally derived from transposable elements (94). The most common are described below.

Long Interspersed Nuclear Elements (LINEs)

LINEs make up about 20% of the human genome (93). They are capable of autonomous mobilization, and transpose via an RNA intermediate (95). Although the transposon is about 6 kb long, incomplete reverse transcription means that the majority of LINE elements are shortened or rearranged (96). They are found predominantly in AT rich DNA (97), thereby decreasing but not eliminating the chance of intragenic insertion (94). NAHR between such elements has been rarely described (98,99).

Short Interspersed Nuclear Elements (SINEs)

SINEs are sequences of 100-400 bp that are non-autonomous, requiring mobilization by other elements (95). In contrast to LINEs, SINEs show an increased density in GC rich regions of the genome, which also correspond with gene-rich areas. In total they make up ~13% of the genome, with the Alu element being the most prominent member of the group. Alu sequences are primate

specific (100), and can be divided into different groups on the basis of evolutionary age. These are present in ~1 million copies in each genome (92), or on average once every 3 kb. In contrast to LINEs, NAHR between Alu repeats has been seen much more frequently (95,101,102).

Segmental Duplications

Segmental duplications, also known as low copy repeats (LCRs), are defined as stretches of DNA, usually between 1-500 kb and with a high degree of homology (>90%), which are present in at least two copies within the genome (103). Such duplicated blocks have not been found in invertebrates, and occur more frequently in humans than lower mammals, for example the rat (104) or mouse (105). In total they comprise ~5 % of the human genome (103,106,107). A notable feature of LCR sequences in humans is the high frequency of Alu sequences found at the boundaries of segmental duplications (108). As it was predominantly the younger Alu sequences that were detected it was suggested that a large increase of Alu retroposition approximately 35 million years ago was the initial trigger for widespread, primate specific, duplication events (109). LCRs are found both interchromosomally and intrachromosomally.

Interchromosomal LCRs

Although these repeats are distributed throughout the genome, they predominate in the subtelomeric and pericentromeric regions. The reason for this is not known, but it has been suggested that such regions show an increased propensity for repeat accumulation (110), or that they are simply more tolerant for rearrangements.

These areas show high rates of recombination, with blocks of repeat units being duplicated and exchanged between multiple chromosomes. An example of such plasticity is found in the subtelomeric regions of the long arms of chromosomes 4 and 10, where swapping of repeat arrays between 4q and 10q has been reported to occur in 20% of the Dutch population (111,112). In addition, subtelomeric screening using FISH probes has shown a high degree of cross-hybridization across different chromosomes (113). The degree of recombination and sequence similarity of such regions has significantly hampered the sequencing of these areas, and it is noticeable that it is often these areas that remain as gaps in the “completed” human genome draft sequence (114).

Repeats on different chromosomes can lead to translocations, and there have been several common breakpoints described and characterized (115,116). It has been suggested that many of

the chromosomal rearrangements that have occurred during speciation are also mediated by LCRs, as studies have shown a correlation between such sequences and syntenic breakpoints between species (117-119).

Intrachromosomal LCRs

Intrachromosomal LCRs, also known as chromosome-specific duplications, are usually found on the same chromosome arm (106,120). Although there is a pericentric bias in localization for several chromosomes, in general such duplicons appear to be more evenly distributed throughout the chromosome than the interchromosomal repeats (103,106).

Several regions between the repeats show a recombination rate higher than would be expected due to random events, and many different disorders are known to be caused by NAHR between flanking LCRs (Table 1). Bailey et al. found 169 regions between 50 kb and 10 Mb flanked by low copy repeats (106), with rearrangements of 24 of these found in locations associated with specific disorders. It remains to be seen whether any of the remaining regions are also rearranged in other diseases.

Despite the fact that the LCR duplicons are often several hundreds of kilobases in length, the exact position of the breakpoints in different patients often shows a preference for a smaller, defined region (121). An explanation for this could be that the increased length of the duplicons allows (mis)alignment to take place, yet more specific sequences within the LCRs are actually involved in the recombination. It has been proposed that at least 200-500 bp of almost complete homology (minimal efficient processing segment; MEPS) is required for recombination (122,123). Additional factors are also involved in influencing the probability of recombination. For example, inversion polymorphisms can increase the chance of a rearrangement occurring between 2 repeats. This has been shown to be the case in the region affected in Williams-Beuren Syndrome (124,125), as well as between olfactory repeat genes in 8p23.1 (126,127).

To date all of the LCRs associated with human disease have been shown to be primate specific (110), and it is possible to give an approximate date to the origin of different LCRs with interspecies comparison. It has been suggested that some primate-specific characteristics have arisen from novel genes created by the fusion of gene fragments (110,114,120).

Disorder	Locus	Size of LCR (kb)	Typical Size of Rearrangement (kb)	Type of Rearrangement	References
Sotos	5q35	140	1300-2700	deletion	(128-130)
Williams-Beuren	7q11.23	320	1600	deletion	(125,131-134)
Angelman/Praeder-Willi	15q12	500	3500	deletion	(135,136)
Smith-Magenis	17p11.2	200	3700	deletion	(121,137-139)
HNPP	17p12	24	1400	deletion	(122,140,141)
Charcot Marie Tooth	17p12	24	1400	duplication	(141-144)
Neurofibromatosis Type 1	17q11	85	1500	deletion	(145-148)
Di George	22q11.2	200-400	1500-3000	deletion	(149-151)
Emery Muscular Dystrophy	Xq28	11	48	deletion / duplication / Inversion	(152-154)
Haemophilia A	Xq28	10	300-500	inversion	(155-157)

Table 1. A selection of regions known to be altered via recombination between specific Low Copy Repeats (LCRs). Described are the rearrangements that occur, the size of such rearrangements and the associated LCRs, and the corresponding disorder.

Non-homologous end joining

Non-homologous end joining (NHEJ) is a poorly understood process, where two DNA ends without significant sequence homology are joined. As the repair is not template based errors frequently occur, and deletions and insertions are often found at the junction. Breakpoints within a given region show less clustering compared to HR, although several sequences have been implicated in increasing the probability of a DSB (158). An example of such is the Chi sequence, which is a nonamer involved with recombination in prokaryotes (159) and has been linked with deletions in several different human disease genes (160-163). Alu sequences also have a 26 bp internal sequence homologous to the Chi sequence (164), suggesting that they may a play a role

in both HR and NHEJ. Detailed analysis of a series of deletion breakpoints showed that the presence of inverted repeats was a common motif in deletions, especially those suggestive of NHEJ i.e. deletions plus small insertions (165). It was proposed that stem-loop formation was responsible for the DSB.

NHEJ is a faster process than HR, which may explain why the majority of DSBs in mammalian cells are repaired by NHEJ (166). Which of the two repair pathways is used is presumably at least partially dependent on the position within the cell cycle, with HR being preferred when a sister chromatid is present (167).

Rearrangements within the *DMD* gene.

Given the above information, what can be said about the cause of the rearrangements found within the *DMD* gene? Although deletions can be found throughout the entire gene, they cluster within 2 areas. These regions also correspond to the location of the largest introns, and the larger introns do tend to show an increased number of breakpoints. Intron 44 is the largest intron within the gene, at ~248 kb, and 35% of all deletions have one of the breakpoints within this intron.

Analysis of the position of the breakpoints, however, reveals that any correlation between length and breakpoint frequency only holds for the 5' end of the breakpoint (figure 3). This is primarily due to a reduction in the number of 3' breaks in the two largest introns (only 15% of breakpoints in intron 44 are at the 3' end of a deletion, and in the case of intron 2 there are no 3' deletion breakpoints described) (Figure 4). The bias between 5' and 3' breakpoints suggests a directional influence in deletion formation, although it is unclear what causes the initial DSB. A direct association between meiotic DNA replication and DSB formation has been described in yeast (86), and the two recombination hotspots within the DMD gene colocalize with two of the most common deletion breakpoints, namely intron 7 and intron 44 (43).

Pozzoli and colleagues have performed a detailed analysis of the repeat content of the introns, and found that 37% of the intronic sequences are repetitive elements (168,169). A significant correlation between intron length and repeat content was found, suggesting that repeat expansion was at least part of the reason for the large introns found in the *DMD* gene. As repetitive sequences have been associated with recombination events, studies have been carried out to determine the exact sequences involved at the breakpoints (170-173). In the majority of cases no sequence similarities were seen at the breakpoints, suggesting that homologous recombination

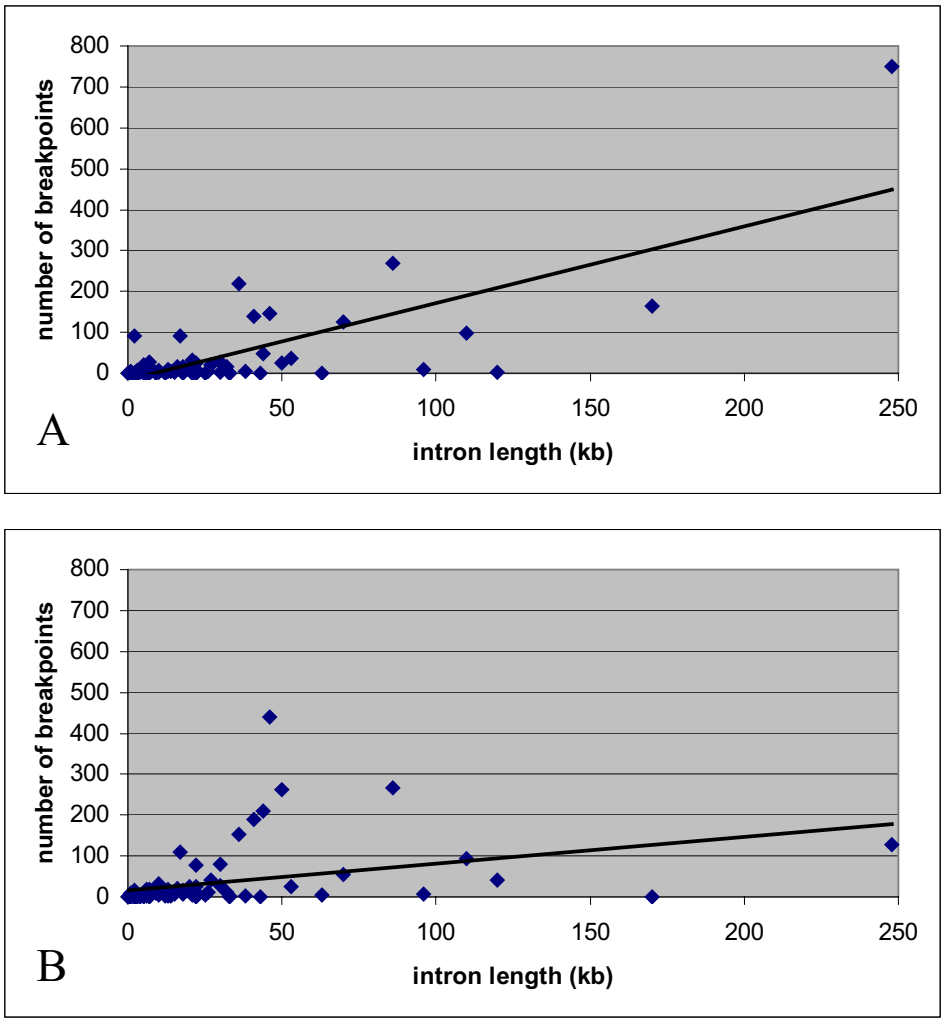

Figure 3. The relationship between intron size and breakpoint frequency for deletions within the *DMD* gene. A. 5' breakpoint frequency. B. 3' breakpoint frequency. The correlations are A; r=0.75, B; r=0.29.

between repetitive elements is not the predominant cause of rearrangements within the *DMD* gene.

Other factors may be involved that influence DNA conformation, increasing the probability of a DSB. Matrix attachment regions (MARs) are sequences of DNA that are involved in DNA looping and attachment to the protein scaffold (174). It has been proposed that a relative paucity of MAR sites might lead to greater torsional stress within these regions (171). This would then lead to an increased probability of DNA breakage, which in turn would lead to an increased chance of a rearrangement. Correlations between Loop Anchorage Regions (LARs) and recombination have also been described (175).

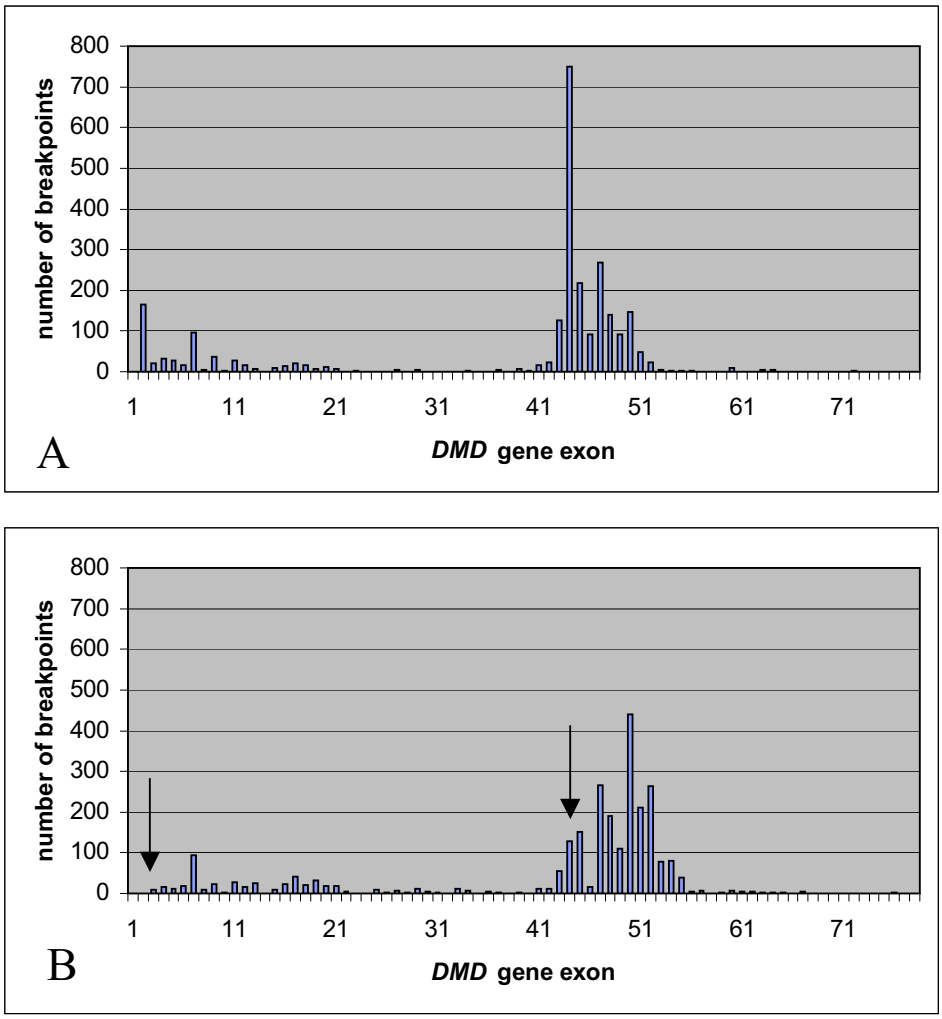

Figure 4. The distinction between (A) 5' and (B) 3' deletion breakpoints. The reduction in 3' breakpoints of the two largest introns (intron 2 and intron 44) is indicated by the arrows.

A study of DNA replication showed 6 origin of replication sites (*ori*s) in the DMD gene (176), and *ori* sites have been associated with chromosomal fragility (177). One of the replication termination sites was within the major deletion hotspot, in intron 44. As this intron makes up ~10% of the gene this could be coincidental, but it is known from studies in prokaryotes that such termination sites are also deletion hot spots (178). It has been shown in *E. coli* that in at least some cases this activity is also dependent on the presence of the Chi sequence (179) . A recent report (180) also showed a co-localization of MARs and *ori*s within the *DMD* gene, with one MAR falling within the major deletion hotspot.

The argument for the importance of chromatin structure is strengthened by the findings of Passos-Bueno and colleagues (181). In their study on deletion distribution in mosaic cases they found

that distal deletions were found in 72% of isolated cases, with 28% of the deletions occurring in the proximal region. Conversely the distribution between the proximal and distal regions amongst inherited cases was virtually identical (47% and 53% respectively). This suggested that proximal deletions were more likely to occur early in development during mitosis, with a corresponding increased risk of recurrence. It is known that transcription is coupled with changes in chromatin structure, which in turn may alter the chances of recombination (85). This may lead to certain regions becoming more accessible/vulnerable during different stages of development. It is probable therefore that the combination of specific sequence interactions and chromatin structure also underlie the disparity seen between the 2 hotspots.

Fewer studies have been carried out looking at the mechanisms involved in duplications. Analysis of 8 families showed that in each case the duplication originated from a single X chromosome, and unequal sister chromatid exchange was thought to be the mechanism responsible. In 7 of the 8 cases the results suggested that the mutation had occurred in the germline of the maternal grandfather (182,183). It was proposed that, as spermatogenesis needs several hundred cell divisions and oogenesis as few as 23, mutations will occur more frequently in the male germline. This theory, however, is not supported by analysis of deletions, where the orginal mutational event has been traced to a female in the majority of cases (184,185). This bias in parental origin of different types of mutations has also been observed for other genes on the X chromosome (186,187).

A more detailed study of three duplications showed that they were all duplicated in tandem (188). One case involved NAHR between two Alu sequences, whereas the other cases appeared to be derived from NHEJ, perhaps involving cleavage sites for DNA Topoisomerase enzymes. These proteins are known to be involved in DNA replication and transcription, breaking DNA molecules and altering chromatin conformation to allow these biological processes to take place (189,190).

It was noted by den Dunnen et al (41) that duplications were more commonly found in the proximal region than in the distal region. A survey of the Leiden DMD database www.dmd.nl (26 October 2004) shows that 59% of all duplications have at least one breakpoint within the proximal region (delineated by exons 2-19), compared with only 20% of the deletions. The locations of the most frequently occurring duplications are significantly different to those of the most common deletions (table 2), again suggesting a fundamental difference in mutation mechanism and/or timing of the different rearrangements.

Along with the differences in distribution, the relative frequency of specific regions being deleted or duplicated can also show marked differences. For example, a duplication of exon 2 is the

single most common duplication seen, yet there has not been a single report of a deletion of exon2 only. Precisely why this is remains unclear. It may be that such a deletion is embryonically lethal, that a deletion of exon 2 is not disease causing, or that structural restraints greatly reduce the probability of such a rearrangement. Alternatively, the mechanism responsible for the duplications may not involve unequal crossing over. Our analysis (chapter 2.2) suggests that the latter explanation is the more likely, although the different explanations are not mutually exclusive.

ranking	deleted exon(s)	%	ranking	duplicated exon(s)	%
1	45-47	7.0	1	2*	8.5
2	45	5.3	2=	3-7	3.5
3	48-50	5.1	2=	8-9	3.5
4	45-48	4.3	4=	2-7	3.0
5	45-50	4.0	4=	51	3.0
6	51	3.9	6=	3-11	2.0
7	44	3.5	6=	8-13	2.0
8	49-50	2.8	8=	3-4	1.5
9=	3-7	2.3	8=	9-14*	1.5
9=	45-52	2.3	8=	17	1.5
11	46-47	2.3	8=	22-25*	1.5
12	50	2.2	8=	43	1.5
13	45-49	2.0	8=	44	1.5

Table 2. The most frequently occurring deletions and duplications. It is clear that the region most commonly deleted is between exons 45-52; in contrast the duplications occur more frequently in the 5' end of the gene. A * indicates a duplication that has not been described as a deletion.

In conclusion therefore, although repetitive sequences are undoubtedly involved in some recombination events, these cannot be the predominant reason for the rearrangements seen in the *DMD* gene. The evidence suggests that other features, such as structural and functional domains, as well as changes in chromatin and DNA topology during replication and transcription, also influence the probability of DNA breakage and subsequent recombination. These factors, along with the repair mechanisms involved, are likely to be cell type dependent, which may underly the differences in distribution and mode of inheritance seen for the different types of mutations.

B. MAPH and MLPA

"The art of research [is] the art of making difficult problems soluble by devising means of getting at them"
-Sir Peter Medawar

The MAPH and MLPA protocols are outlined in figure 5, and discussed in more detail below.

MAPH

The MAPH assay is based upon the quantitative recovery of probes following hybridization to immobilized genomic DNA. Probes for MAPH analysis were initially created by cloning small DNA fragments into a vector (191). These fragments were generated either by PCR or restriction digestion of larger DNA constructs (e.g. PACs or BACs). The specific probes were then prepared by amplification from the vector using primers specific for the vector sequence. The drawback of the cloning approach is that the preparation of a large number of probes is arduous and time consuming.
An alternative method of probe preparation is to design primers to consist of two parts (192). A unique section, which is used for the amplification from genomic DNA, and an identical 5' priming sequence, which is the same for each probe. Once each probe is amplified from genomic DNA the products can then be combined into a probe mix. Because they all have the same ends the subsequent amplification can take place in exactly the same manner as for cloned probes.
A potential disadvantage of this approach is that during the original PCR amplification from genomic DNA, more than one PCR product could be amplified. These extra products would then be present in the probe mix, and the extra peaks would appear on any subsequent trace. In practice, however, we have not seen this as a common problem, and any problematical probes could be cloned and/or gel purified if necessary.
There are several criteria that the probes need to conform to. First, each sequence must be unique. This can be tested using the BLAST (193) or BLAT (194) program to check the sequence against the genomic database. Second, the hybridizing segment of each probe should ideally have a GC-content of between 40% and 60%. Some regions of the genome do not fall within these limits, and the probe length should be adjusted accordingly. Probes for regions with a GC-content as low as 30% have been successfully made, but these were all longer than 500 bp.

For difficult regions of the genome it may be necessary to try several different sequences before a good probe is found. Third, it is essential that each probe is of a different length, allowing the probes to be distinguished after electrophoretic separation.

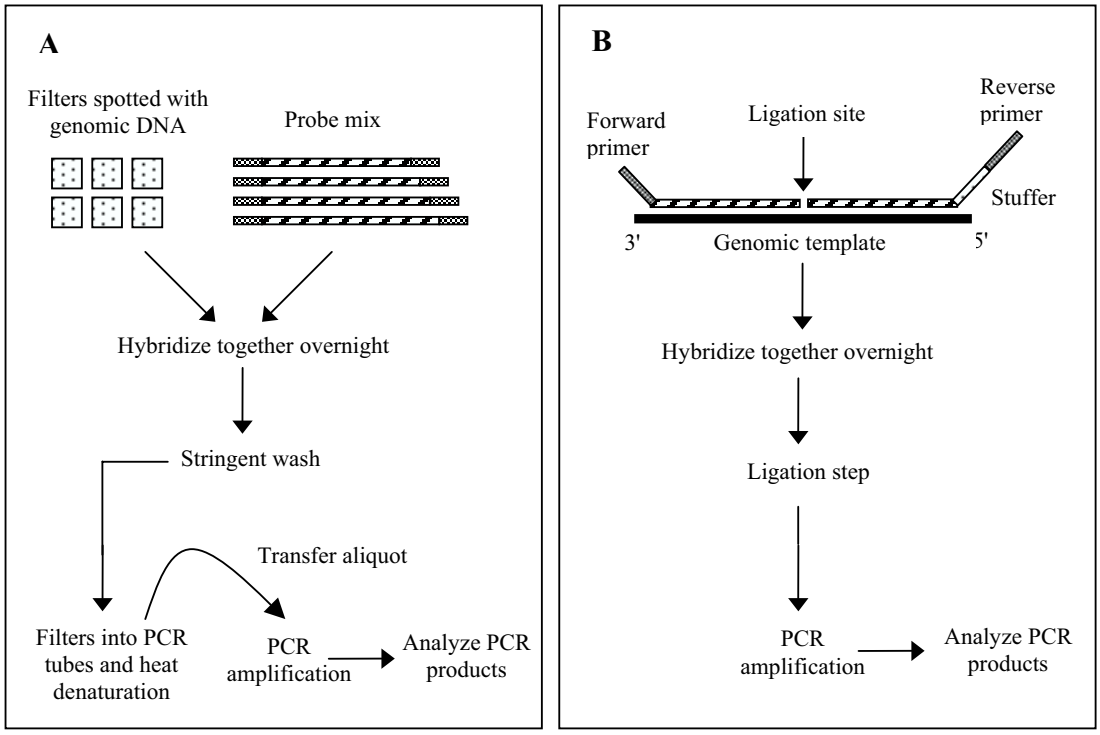

Figure 5. An outline of the two procedures. A. The MAPH technique. A series of probes with identical ends are hybridized to genomic DNA immobilized on a filter. Following stringent washing the specifically bound probes are released into solution, and an aliquot of this is used to seed a PCR reaction. B. The MLPA technique. Only if the 2 half-probes hybridize adjacently on a target sequence can they be ligated and amplified using the universal primers in a subsequent PCR reaction. Probes can be distinguished by using stuffer sequences of different lengths.

Mixes are prepared by combining all probes so that the final concentration of each is ~0.5 ng/µl to 1.0 ng/µl. We routinely store all probe mixes at –20 ^{o}C, and these have been successfully used up to 18 months after preparation with repeated freeze/thaw cycles.

Protocols have been outlined in previous articles (47,191), and complete descriptions are available at the following websites (http://www.nottingham.ac.uk/~pdzjala/maph/ and http://www.dmd.nl/DMD_MAPH.html).

As it is necessary to remove the unbound probes after the hybridization step the genomic DNA has to be immobilized. This is done by denaturing the DNA in NaOH and applying it to a small nylon filter. Because the filters are combined in a single tube during the hybridization step they

need to be marked in such a way that they can be easily distinguished from each other. It is possible to cut the filters into different shapes, but the most convenient method is to number each filter with a sharp pencil. This mark will remain visible after all the washing steps.

We have found that the easiest way of applying the DNA to the filter is to pipette each DNA sample into a well of micotiter plate, already containing the NaOH. Each filter can then be put into the appropriate well, and the DNA allowed to adsorb onto the filter (figure 6).

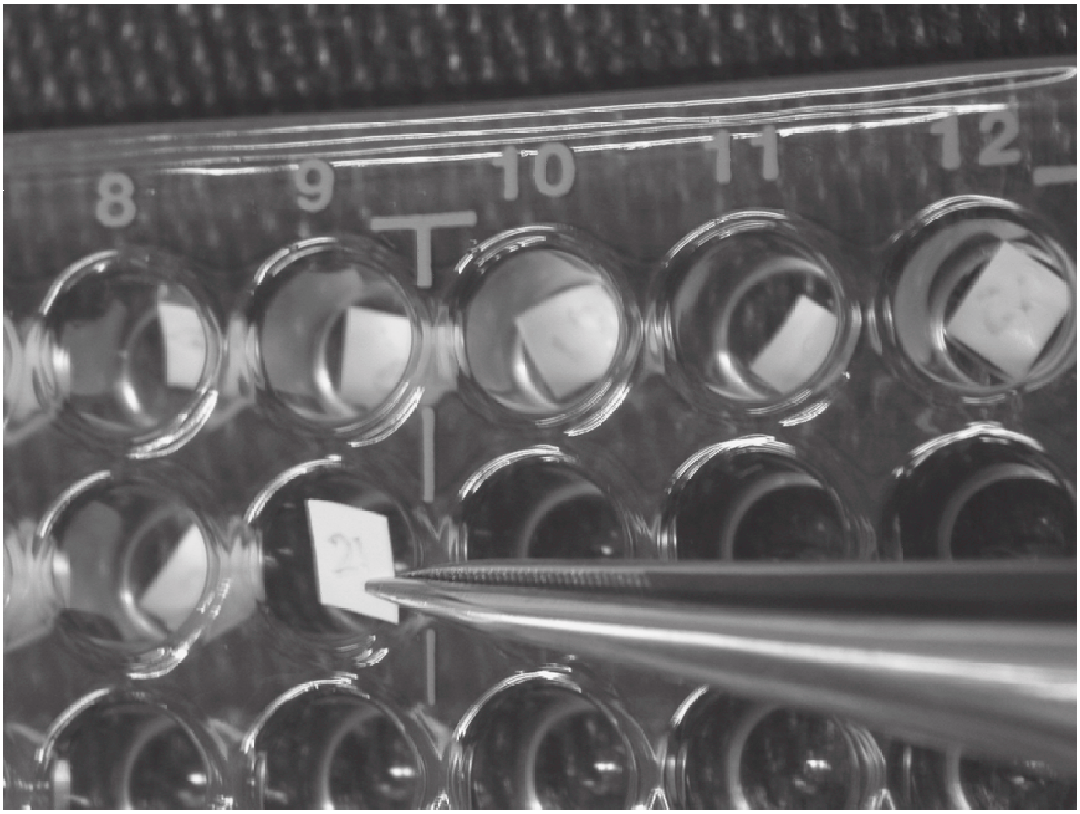

Figure 6. The use of a microtiter plate for applying denatured genomic DNA to the filters. The number on the filter held in the forceps can be clearly seen; this will remain easily visible after all the washing steps.

After the filters are dry the DNA is fixed to the filter, either by UV cross-linking, or by heating the filters at 80 ^{o}C for 1 hour. The filters are then combined in a prehybridization mixture for 2 hours to overnight, followed by addition of the hybridization mix. This mixture contains the probes of interest along with competitive DNA (e.g. C_ot1, fragmented *E. coli* or herring sperm DNA as well as sequences that block any interactions between the common ends of the probes). After overnight incubation the hybridization mixture is removed, followed by stringent washing. To avoid accidental loss of the filters during this step we first pour off washing solution into a

beaker, as it is easier to recover filters from this than the U-bend of a sink. Alternatively, a tea strainer or something similar can be used.

Following the washing steps each filter is placed individually into a PCR tube containing 1x PCR buffer. After heating to 95 ^{o}C for 5 minutes the previously bound probes are released into solution, and an aliquot of this is added to a PCR mix. Optimization of the PCR reaction regarding such factors as cycle number and denaturation time is necessary to obtain optimal results.

MLPA

MLPA (195) is based upon the ligation of two adjacently-annealing oligonucleotides, followed by a quantitative PCR amplification of the ligated products. The left hand half-probe is usually the shorter of the two probes (typically 45-70 nt), and is chemically synthesized. This half-probe is composed of two sections, a unique, annealing sequence and terminal priming sequences common to all probes. The right hand half-probe can be up to 440 nt, which cannot be routinely synthesized. Instead, a series of M13 vectors has been created, each with a "spacer" sequence of a different length (195). Two complementary oligonucleotides are annealed to each other, followed by ligation into the modified M13 vector. It is the spacer which determines the final length of the half probe. The resulting M13 construct containing the annealed oligonucleotides is isolated in single-stranded form, and the cloned sequence, including the spacer and priming sequence can be isolated from the vector by restriction digestion.

It is also possible to use chemically synthesized oligos for the right hand half-probe as well, but to facilitate ligation it is necessary that this sequence is 5' phosphorylated. Although chemically synthesizing both probes greatly accelerates the probe production process, due to length constraints during synthesis the number of probes that can be subsequently combined within a single set is limited.

The MLPA protocol is described in full in Schouten et al., 2002 (193), and is also available at the following web site (http://www.mrc-holland.com/). As outlined in the notes of the protocol, several modifications are possible. To save on reagents the PCR volume can be reduced from 50 µl to 25 µl without influencing the results. Combining the two pre-PCR mixes together and adding 5 µl of the ligated products on ice makes the overall procedure easier, particularly when working with many samples. As with MAPH, it may be necessary to adjust the number of PCR cycles to obtain sufficient product for analysis.

Analysis

For both MAPH and MLPA the methods of data analysis are effectively identical. The PCR products are usually separated by electrophoresis and each probe is quantified, with the relative amount of each product being proportional to the copy number of the locus being tested. In the original MAPH protocol (191) the probes were labeled with a ^{33}P-5'-end-labelled primer and separated on a polyacrylamide gel, with the bands being quantified using a phosphoimager screen. A faster and easier method is to use a fluorescently labeled primer during the PCR amplification (47) and separate the products on a sequencing system, either a polyacrylamide gel or through polymer-filled capillaries (Figure 7). This approach also simplifies the analysis of the products, as software (e.g. GeneScan and Genotyper from Applied Biosystems) has been integrated into these systems. To obtain sufficient signal it may be necessary to concentrate the PCR reaction, as the amount of product to be added depends on the sensitivity of the system.

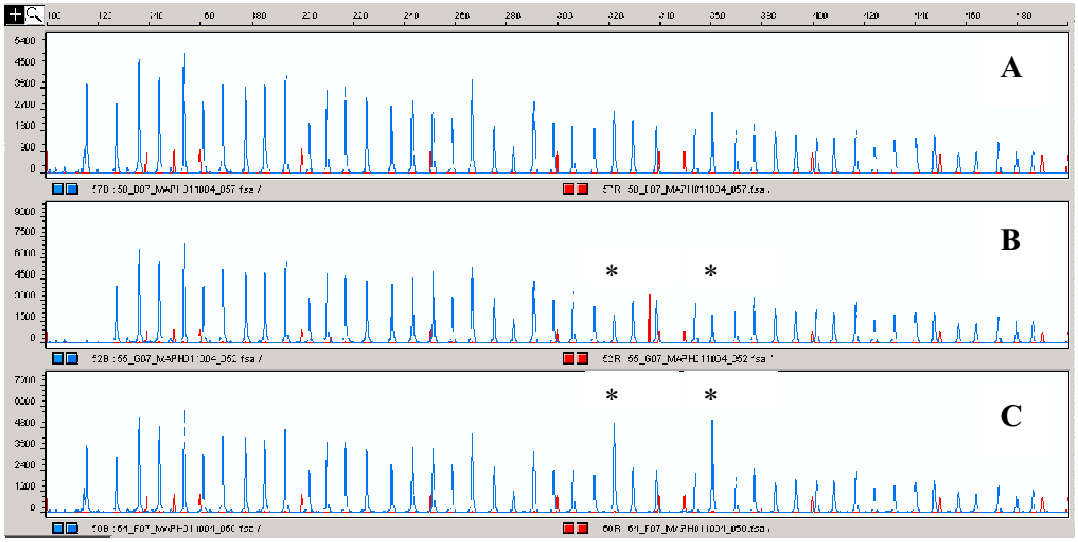

Figure 7. Peaks obtained following separation of one of the MLPA probe sets for the DMD gene on the ABI 3700. Panel A shows a typical pattern from an unaffected male. Panel B is from a female carrier with a deletion of two exons, and panel C is from a male with a duplication of the same two exons. Affected probes are indicated with asterisks.

Another possibility is to use the Lab-on-a-chip from Agilent (Figure 8). This DNA chip can separate a maximum of 12 samples in 45 minutes. The resolution is ~20 bp, meaning that on the DNA500 chip up to 15 different probes can be analyzed, and it is not necessary to use a fluorescently labeled primer. This approach might be attractive when relatively few loci and/or samples need to be tested (78).

Several different methods have been described for data analysis. At the most basic level peaks can be visually compared (195), e.g. by overlaying control and test traces. In theory any deletion or duplication should be sufficiently obvious without any calculations being performed. Whilst this is usually true for deletions, especially single exon duplications will not always be obvious. For this reason, and to automate the analysis it is recommended to perform a computer-based analysis.

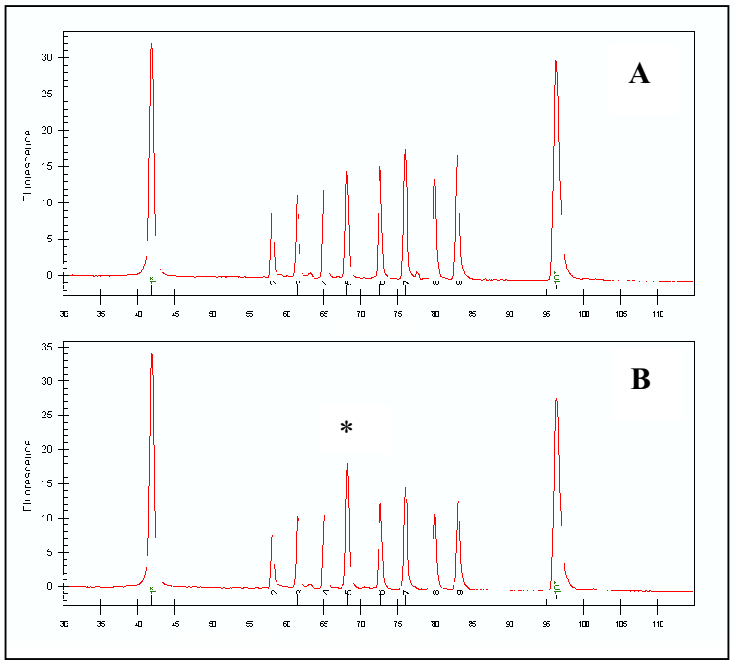

Figure 8. An example of traces obtained by analyzing MAPH products on the Lab-on-a-chip from Agilent. In this example the probes have a spacing of 30-40 base pairs. As can be seen, this spacing could be reduced if necessary to allow more probes to be used. A typical pattern from an unaffected individual is shown in trace A, a single locus duplication is indicated by an asterisk in trace B.

Software such as GeneScan provides data about the peak height and peak area, both of which have been used in different reports (47,196-198) with equivalent results. Most methods of calculation described have been based on dividing the value of a given probe by the sum of 2 or more other probes to obtain a ratio. These are preferably control probes that are unlinked to the loci in question. The "nearest neighbor" method can be used, where the values of the four nearest peaks are added together. This means that each probe will be normalized against a different group of probes. It is also possible to add up the values of all probes in a trace (a global approach), and divide each probe by that value. Exactly which method is used is dependent on several factors, primarily how many probes and samples you expect to be affected. For example,

subtelomeric rearrangements are found in about 5% of mentally retarded patients, and it is unlikely that more than two probes would be affected. It should therefore be no problem to use the nearest neighbor or global approach, as the vast majority of probes and samples will be unaffected. In contrast, sets composed of probes for all exons within one gene should be analyzed by another method. It is possible that most if not all of the exons within a gene are deleted, which mean that none of these probes should be used for normalization purposes. It is therefore essential to add control probes for loci elsewhere in the genome.

The ratio obtained for each probe is averaged across a series of samples to obtain a normalized value, usually corresponding to a copy number of 2 (1 for X-linked probes in a male). The original ratio for each sample is then divided by the normalized value, and in an unaffected situation should give a number distributed around 1. Under ideal circumstances a deletion (1:2) would give a value 50% lower, i.e. 0.5, and a duplication (3:2) a value 50% higher i.e. 1.5 (Figure 9). Of course each probe will show a certain level of variation, and it is this variation that determines the degree of certainty of each measurement. The variation is dependent on several factors. Amplification by PCR will introduce a certain level of variability; this can be determined by performing several amplifications from a single pre-MAPH/MLPA. Each probe itself will have a certain amount of variance, depending on several factors such as GC-content, degree of homology with other regions in the genome, possible interactions with other probes, amplification efficiency, etc. These influences can be estimated by calculating the standard deviation for each probe over a series of control samples. The accuracy of each probe in detecting a true copy number change is also important. For this reason it is desirable to be able to test each probe on samples with known mutations.

Different criteria can be applied for deciding whether a given result is significant or not. At the most basic level thresholds can be set, usually at 0.75 and 1.25. A more statistical approach can be used by calculating the standard deviation either for each probe across all samples, or for all probes within a sample. As the variation of each probe also depends on the quality of the DNA being tested, it does not always follow that the standard deviation seen on control samples will be the same as that seen on test samples. This usually, however, provides a good estimate as to the overall reliability of any given probe. The use of bivariate analysis has also been described (196). This relies on duplicate testing of each sample, and allows the user to decide beforehand what the false negative rate will be. A full explanation is provided at the following web site (http://www.nottingham.ac.uk/~pdzjala/maph/ststats.pdf).

The exact criteria that are used for determining a significant result, be it a fixed figure or a certain number of standard deviations, will determine the false positive and false negative rate. Setting

the thresholds at a relatively low level will lead to a correspondingly low false negative rate. This will however lead to a relatively higher false positive rate. One of the strengths of both MAPH and MLPA is that processing many samples does not take much more time than processing a few

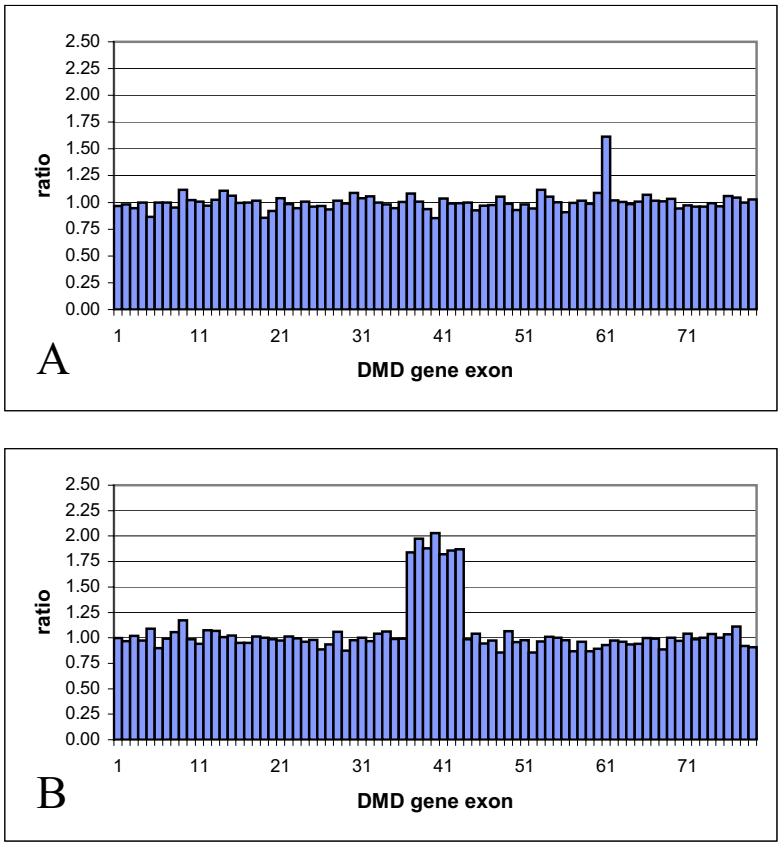

Figure 9. Analysis of MLPA reactions with 2 DMD probe sets following normalization. Under normal circumstances the ratios of all probes clustered around 1.0, indicating an unaffected individual. Graph A shows a single probe with a ratio around 1.5, this is from a female carrier of an exon 61 duplication. In graph B there are seven probes with ratios around 2.0, corresponding to a duplication of exons 37-43 in a male patient.

samples. Thus it should be easy to test all samples twice, and only probes that deviate in a significant manner in both duplicates should then be retested with another technique. If the majority of samples do not show any changes, however, then it can be argued that routine duplicate testing, at least on the same DNA isolate, is not necessary. If both tests need to show the same deviation, and the first is normal, then the second test is not going to change the conclusion. Therefore only samples that show significant changes in the first round need to be retested.

Whilst detecting deletions (1:2) and duplications (3:2) is relatively straightforward, distinguishing relatively smaller changes, e.g. in mixed or mosaic samples, requires testing each sample several

times to determine confidence limits. This is one of the strengths of both MAPH and MLPA, as multiple testing requires little extra time or expense. It is also advantageous to be able to use data from more than one probe, as this strengthens the statistical analysis.

In addition to false positives and false negatives because of variations affecting the calculations, incorrect conclusions can also be drawn due to inherent limitations of each technique. For example, if only part of the genomic region that binds a MAPH probe is deleted then the signal obtained from that probe will be proportionally lower, but perhaps not enough to be significant.

Due to the sensitivity of the MLPA reaction to mismatches at the ligation site of the two half-probes, even a single nucleotide change may be enough to prevent successful ligation. In such a case the result will be scored as a deletion, a conclusion that will not be changed by repeated analyses. For this reason it is critical to confirm all single probe deletions found with MLPA with another technique.

Applications

There have been several reports on the use of MAPH and MLPA for detecting copy number changes. These have focused primarily on screening either single genes for exonic deletions and duplications, or chromosomal regions for rearrangements.

Given the potential to screen up to 50 loci simultaneously an obvious target was to develop probes for each of the subtelomeric regions of the human chromosomes (43 in total). Such a probe set can be used to detect trisomies and unbalanced translocations. In addition, rearrangements in the subtelomeric regions have been shown to be involved in mental retardation (MR), with rearrangements found in ~5% of cases (199). Subtelomeric screening using MAPH has been described in three reports. The first used a combination of FISH and MAPH (200). A total of 70 samples from MR patients had been screened with FISH, and one deletion was detected. Analysis with MAPH gave the same result. The second report was of the MAPH screening of 37 DNA samples from patients referred for Fragile X screening (196). In this study 6 rearrangements were found (16%), one of the highest percentages found in subtelomeric screening of MR patients. No FISH screening was performed, and the rearrangements were confirmed by semi-quantitative PCR. A problem with FISH confirmation of MAPH results is that a negative result with FISH does not mean that the MAPH finding is not correct. The difference in resolution afforded by the techniques (a minimum probe size of ~40kb vs. ~100bp) means that MAPH may have detected a genuine but small alteration that simply cannot be resolved with FISH. For this reason confirmation with a high resolution technique such as

MAPH, MLPA, quantitative PCR or Southern blotting may be necessary. The third report (201) (Chapter 4) analysed 188 patients with developmental delay, and found subtelomeric rearrangements in 9 cases (5%). In this study a number of interstitial loci were also tested, resulting in a further 8 rearrangements being detected.

In addition to the screening of the *DMD* gene in DMD/BMD patients (Chapter 2.1, 2.2, 2.3), we have also looked at genes involved in other muscular disorders, namely the sarcoglycanopathies. Using this approach it was possible to identify the identical single exon deletion within the *SGCG* gene in three, unrelated limb-girdle muscular dystrophy (LGMD) patients (Chapter 3).

MAPH has also been applied to the analysis of deletions and amplifications in Chronic Myeloid Leukemia (192). This clonal malignancy is characterized by the generation of a *BCR-ABL1* fusion gene after a translocation involving chromosomes 9 and 22 (the Philadelphia (Ph) chromosome). The marked difference in disease progression seen in these patients was thought to be at least partly due to deletions at the breakpoints of the chromosome 9 derivative. Probes were made at the breakpoints (9q34 and 22q12), and were used to test DNA samples with known deletions and duplications. Using titration experiments the authors were able to show that they could still see deletions and amplifications when only 60% to 70% of the cells were affected. This is a demonstration of the ability of MAPH to detect mutations in mosaic cases.

Another application of MAPH, testing its limits of sensitivity, was the analysis of a previously identified polymorphic region on 8p23.1 (202). This region is flanked by olfactory repeats, and it had been shown that up to 25% of the normal population carries an inversion polymorphism between these repeats (126). Additionally, an apparently benign duplication of the region had been described, with no obvious effect on the carriers. By designing several MAPH probes within this region, along with the use of semi-quantitative FISH, it was possible to estimate the copy number of this region. It could be shown that the majority of individuals had between 2 and 7 copies of the allele, whereas carriers of the apparent duplication had in fact 9-12 copies.

MLPA probe sets are commercially available, and there have been several reports detailing their application. To date these have been primarily focused on screening for exonic deletions and duplications in specific disease genes. The first genes screened were those known to be commonly deleted in different cancers. These include *BRCA1* (197,203) involved in breast cancer, and *MSH2* and *MLH1* involved in colon cancer (204,205). More recently, two sets together covering all exons of the *DMD* gene have become available (Chapter 2.2).

Subtelomeric screening has also been performed using MLPA. A total of 4 rearrangements were found in a group of 75 DNA samples from mentally retarded patients (198), which is similar to the average percentage found in other studies using alternative methods. An additional

rearrangement was suspected, but further examination showed that there was a 3 bp deletion at the probe site. This case illustrates the importance of confirming apparent deletions found with a single probe, as even a single nucleotide change may disturb the ligation of the two half-probes.

An interesting application lies in the detection of aneuploidy directly from amniotic fluid lysates (206). In a study of 492 samples there were 18 aneuploidies identified, with no false positives or false negatives. The probe set used contained 8 probes for each of chromosomes 13, 18 and 21, 4 probes for chromosome X and Y and 8 probes for other chromosomes. To determine if an aneuploidy was present the mean normalized ratio of all probes for a given chromosome was used, as it was stated that not all relevant probes showed the aneuploidy. By combining the results of multiple probes the sensitivity of the assay is enhanced, and in this case the vast majority of samples were tested only once.

To circumvent the time-consuming and expensive cloning required in MLPA probe preparation, we have looked at the possibility of using synthetic oligonucleotides for both half-probes (207) . Due to length limitations during the chemical synthesis the maximum size of ligated probes is ~130 bp, considerably less than the 490 bp obtained from cloned products. To partially circumvent these restrictions we combined the single base pair resolution of capillary electrophoresis with the ability to analyze multiple fluorescent labels simultaneously. We made two sets of probes, each being amplified with a defined primer pair, and each set being labeled with a different fluorophore. As the primers function under the same PCR amplification conditions it is possible to combine all probes in the same reaction. This approach has been applied to the analysis of the *EXT1* and *EXT2* genes in patients with multiple osteochondromas (Chapter 5.1), as well as the *CBP* and *EP300* genes in Rubinstein-Taybi patients (Chapter 5.2).

C. Discussion

Although MLPA and MAPH share many similarities, there are significant differences that will influence which of the two will be considered for a given purpose. MLPA requires ~100 ng of genomic DNA, considerably less than the 1µg required for MAPH analysis. While 1 µg is not an excessive amount, it may be a problem when dealing with older archived samples, or when analysing amniotic fluid.

A genuine drawback of the MAPH technique is the necessity to immobilize the genomic DNA, usually performed by spotting onto a small nylon filter. This is inconvenient, and the subsequent washing steps are time consuming. In contrast, MLPA is essentially a one-tube assay, particularly important for diagnostic purposes as it minimizes the chance of sample mix-up during the assay. It also makes automation of the entire procedure easier. For these reasons MLPA is becoming the technique of choice, especially in diagnostic settings.

As a MAPH probe is a PCR product, probe preparation requires little more than 2 primers and a PCR reaction. This is in contrast to the standard manner of MLPA probe production, which requires a cloning step followed by single-stranded DNA isolation. The time and expense required can be justified for probe sets that are commercially desirable, yet there will often be cases when probes will be needed for extremely rare disorders, or for determining the size of a rearrangement. In these cases cloning is not an attractive option, and MLPA with synthetic oligos only should be considered (207).

It is easy to foresee that new MLPA/MAPH assays will be developed that will focus not only on specific genes but also on high-resolution analysis of chromosomal regions associated with a range of diseases. Several large deletions and duplications are known to be associated with a specific spectrum of disorders, e.g. microdeletion syndromes and contiguous gene disorders. Such rearrangements can be of the order of several megabases, making them cytogenetically visible. Often many genes are found within these areas and the extent of the rearrangement is directly correlated with disease severity, i.e. the sum of the genes affected determines the overall phenotype. The use of MAPH/MLPA assays should make it straightforward to determine the extent of the deletion/duplication.

MAPH and MLPA facilitate the versatile detection of copy number changes in a selected set of up to 50 target sequences. What has been lacking until recently is a technology that can be used in a scanning mode and genome-wide. In many cases it is not yet known where genes involved in specific disorders reside, e.g. mental retardation. As a first step to identify the genes responsible one would want to scan the entire genome for cases where a (large) deletion/duplication is

involved. These rare cases might then point to the genes involved, which could then be analyzed in more detail and in a much larger set of patients. Similarly, it would be desirable to have such a scanning tool to analyze genomic DNA of newborns presenting with initially unclear health problems. In a significant fraction of these newborns, de novo genomic copy number changes might be involved which, when detected, could provide a much earlier diagnosis and thereby valuable information for health care.

To fill in this gap, array CGH (array-based Comparative Genomic Hybridization) has been developed. It has been successfully applied by spotting BACs/PACs (or PCR products thereof) onto a glass slide, and hybridizing labeled genomic DNA to these (208-210) to pick up genomic copy number changes. The method could be successfully applied to detect large rearrangements, especially in cancer-related samples, but recently also to reveal rearrangements in patients with mental retardation and dysmorphisms. However, these arrays have a limited resolution. Even with the 30000 BAC/PAC array recently described (211) only relatively large (>50-100 kb) deletions / duplications will be detected. Furthermore, due to the presence of low copy repeats, some regions of the genome are not amenable to analysis in this way (212).

An obvious approach to resolve the issue of resolution would be to design an array-based MLPA/MAPH assay. In such an assay it is no longer necessary for each probe to be of a different length for subsequent electrophoretic separation. This feature might help to resolve the main technical hurdle of this approach, namely the maximum degree of multiplicity of the assay, i.e. the number of loci that can be simultaneously amplified and resolved when combined in a single reaction. We have successfully developed an array analyzing 26 MLPA probes simultaneously, and preliminary results from our laboratory indicate that a 200-plex, array-based MLPA assay is possible (Kalf et al., unpublished).

For high resolution, genome-wide analysis of copy number changes the most promising developments may be the application of whole-genome SNP-typing assays. An approach very similar to MLPA, also based on the ligation and quantitative amplification of oligonucleotides has been developed (213,214). The primary difference lies in the analysis method. Instead of separating products of a different length by electrophoresis, the PCR products are captured by complementary sequences attached to microspheres. As up to 1500 different microspheres can be distinguished using a specific color-coding system, it should be possible to analyze 1500 loci in genomic DNA simultaneously. In addition, the system has been set up for micro-titer plate analysis, up to 96 samples can be processed in parallel. This then tackles another important issue of these technologies, which is the number of different patient samples that can be processed in a specific time. In that respect, array-based technologies currently have significant limitations.

The development of a 10,000 human SNP-typing microarray was initially reported (215) with a 120,000 array recently becoming available (188). Instead of hybridizing the entire genome, a reduction in complexity is achieved by digesting the genomic DNA with a restriction enzyme, followed by attachment of a linker. This sequence allows for the subsequent amplification of the intervening sequences. The amplified fragments are then hybridized to short complementary sequences on an array. Deletions can be detected with such tools by first using the SNP itself i.e. inheritance of 'null alleles', or indirectly derived from the presence of exceptionally large homozygous regions which might point to a hidden deletion. Rearrangements should also be detectable using the amount of signal generated per SNP, with lower signal (and homozygosity) pointing to deleted regions and increased signals to duplicated regions (216).

A very similar, linker-based approach has also been described, namely Representational Oligonucleotide Microarray Analysis (ROMA) (217). This used oligonucleotide microarrays specifically designed to detect the PCR fragments, so was not intended for SNP analysis. This does not allow the array to be used for LOH analysis, but does give more flexibility with regard to probe localization.

Copy number analysis is not only of importance in studying disease. As this thesis was being finalised three reports using either Array CGH (218), ROMA (219) and MLPA (Chapter 6) (220) to examine copy number polymorphisms in healthy individuals were published. Whatever methods are used, high resolution quantitative analysis will be an important step following the generation of the draft human genome sequence. The information obtained will not only be important for the identification of new disease genes, but will also give greater insight into the functioning and plasticity of the genome.

Reference List

1. Armour,J.A., Barton,D.E., Cockburn,D.J. et al (2002) The detection of large deletions or duplications in genomic DNA. *Hum.Mutat.*, **20**, 325-337.

2. Van Ommen,G.J.B., Breuning,M.H., and Raap,A.K. (1995) FISH in genome research and molecular diagnostics. *Curr.Opin.Genet.Dev.*, **5**, 304-308.

3. Ligon,A.H., Kashork,C.D., Richards,C.S. et al (2000) Identification of female carriers for Duchenne and Becker muscular dystrophies using a FISH-based approach. *Eur.J.Hum.Genet.*, **8**, 293-298.

4. Petrij,F., Dauwerse,H.G., Blough,R.I. et al (2000) Diagnostic analysis of the Rubinstein-Taybi syndrome: five cosmids should be used for microdeletion detection and low number of protein truncating mutations. *J.Med.Genet.*, **37**, 168-176.

5. Koenig,M., Hoffman,E.P., Bertelson,C.J. et al (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*, **50**, 509-517.

6. Petrij-Bosch,A., Peelen,T., Van Vliet,M. et al (1997) BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. *Nat.Genet.*, **17**, 341-345.

7. Den Dunnen,J.T., Bakker,E., Klein-Breteler,E.G. et al (1987) Direct detection of more than 50% Duchenne muscular dystrophy mutations by field-inversion gels. *Nature*, **329**, 640-642.

8. Van Der Maarel,S.M., Deidda,G., Lemmers,R.J. et al (1999) A new dosage test for subtelomeric 4;10 translocations improves conventional diagnosis of facioscapulohumeral muscular dystrophy (FSHD). *J.Med.Genet.*, **36**, 823-828.

9. Abbs,S.J., Bobrow,M. (1992) Analysis of quantitative PCR for the diagnosis of deletion and duplication carriers in the dystrophin gene. *J.Med.Genet.*, **29**, 191-196.

10. Ioannou,P., Christopoulos,G., Panayides,K. et al (1992) Detection of Duchenne and Becker muscular dystrophy carriers by quantitative multiplex polymerase chain reaction analysis. *Neurol.*, **42**, 1783-1790.

11. Wilke,K., Duman,B., and Horst,J. (2000) Diagnosis of haploidy and triploidy based on measurement of gene copy number by real-time PCR. *Hum.Mutat.*, **16**, 431-436.

12. Anhuf,D., Eggermann,T., Rudnik-Schoneborn,S. et al (2003) Determination of SMN1 and SMN2 copy number using TaqMan technology. *Hum.Mutat.*, **22**, 74-78.

13. Boehm,D., Herold,S., Kuechler,A. et al (2004) Rapid detection of subtelomeric deletion/duplication by novel real-time quantitative PCR using SYBR-green dye. *Hum.Mutat.*, **23**, 368-378.

14. Meryon,E. (1852) On granular and fatty degeneration of the voluntary muscles. *Medico-Chirurgical Transactions*, **35**, 73-84.

15. Emery,A.E., Emery,M.L.H. (1995) *The History of a Genetic Disease - Duchenne Muscular Dystrophy or Meryon's Disease*. Royal Society of Medicine Press Limited.

16. Lindenbaum,R.H., Clarke,G., Patel,M. et al (1979) Muscular Dystrophy in an X:1 translocation female suggests that Duchenne locus is on X-chromosome short arm. *J.Med.Genet.*, **16**, 389-392.

17. Jacobs,P.A., Hunt,P.A., Mayer,M. et al (1981) Duchenne muscular dystrophy (DMD) in a female with an X/autosome translocation: further evidence that the DMD locus is at Xp21. *Am.J.Hum.Genet.*, **33**, 513-518.

18. Boyd,Y., Buckle,V.J., Holt,S. et al (1986) Muscular dystrophy in girls with X;autosome translocations. *J.Med.Genet.*, **23**, 484-490.

19. Francke,U., Ochs,H.D., De Martinville,B. et al (1985) Minor Xp21 chromosome deletion in a male associated with expression of Duchenne muscular dystrophy,chronic granulomatous disease, retinitis pigmentosa and McLeod syndrome. *Am.J.Hum.Genet.*, **37**, 250-267.

20. Murray,J.M., Davies,K.E., Harper,P.S. et al (1982) Linkage relationship of a cloned DNA sequence on the short arm of the X chromosome to Duchenne muscular dystrophy. *Nature*, **300**, 69-71.

21. Davies,K.E., Pearson,P.L., Harper,P.S. et al (1983) Linkage analysis of two cloned DNA sequences flanking the Duchenne Muscular Dystrophy locus on the short arm of the human X chromosome. *Nucl.Acids Res.*, **11**, 2303-2312.

22. Van Ommen,G.J.B., Verkerk,J.M.H., Hofker,M.H. et al (1986) A physical map of 4 million base pairs around the Duchenne muscular dystrophy gene on the human X-chromosome. *Cell*, **47**, 499-504.

23. Koenig,M., Monaco,A.P., and Kunkel,L.M. (1988) The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell*, **53**, 219-228.

24. Mandel,J.L. (1989) Dystrophin. The gene and its product. *Nature*, **339**, 584-586.

25. Den Dunnen,J.T., Grootscholten,P.M., Dauwerse,J.D. et al (1992) Reconstruction of the 2.4 Mb human DMD-gene by homologous YAC recombination. *Hum.Mol.Genet.*, **1**, 19-28.

26. Tennyson,C.N., Klamut,H.J., and Worton,R.G. (1995) The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat.Genet.*, **9**, 184-190.

27. Hoffman,E.P., Brown,R.H., and Kunkel,L.M. (1987) Dystrophin, the protein product of the Duchenne muscular dystrophy locus. *Cell*, **51**, 919-928.

28. Bar,S., Barnea,E., Levy,Z. et al (1990) A novel product of the Duchenne muscular dystrophy gene which greatly differs from the known isoforms in structure and tissue distribution. *Biochem.J.*, **272**, 557-560.

29. Gorecki,D.C., Monaco,A.P., Derry,J.M.J. et al (1992) Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. *Hum.Mol.Genet.*, **1**, 505-510.

30. Byers,T.J., Lidov,H.G.W., and Kunkel,L.M. (1993) An alternative dystrophin transcript specific to peripheral nerve. *Nat.Genet.*, **4**, 77-81.

31. Lidov,H.G., Selig,S., and Kunkel,L.M. (1995) Dp140: a novel 140 kDa CNS transcript from the dystrophin locus. *Hum.Mol.Genet.*, **4**, 329-335.

32. Ahn,A.H., Kunkel,L.M. (1993) The structural and functional diversity of dystrophin. *Nat.Genet.*, **3**, 283-291.

33. Campbell,K.P., Kahl,S.D. (1989) Association of dystrophin and an integral membrane glycoprotein. *Nature*, **338**, 259-262.

34. Rando,T.A. (2001) The dystrophin-glycoprotein complex, cellular signaling, and the regulation of cell survival in the muscular dystrophies. *Muscle Nerve*, **24**, 1575-1594.

35. Ilsley,J.L., Sudol,M., and Winder,S.J. (2001) The interaction of dystrophin with beta-dystroglycan is regulated by tyrosine phosphorylation. *Cell Signal.*, **13**, 625-632.

36. Hofker,M.H., Wapenaar,M.C., Goor,N. et al (1985) Isolation of probes detecting restriction fragment length polymorphisms from X-chromosome specific libraries: potential use for diagnosis of Duchenne muscular dystrophy. *Hum.Genet.*, **70**, 148-156.

37. Kunkel,L.M., Monaco,A.P., Middlesworth,W. et al (1985) Specific cloning of DNA fragments absent from the DNA of a male patient with an X-chromosome deletion. *Proc.natl.Acad.Sci.USA*, **82**, 4778-4782.

38. Ray,P.N., Belfall,B., Duff,C. et al (1985) Cloning of the breakpoint of an X:21 translocation associated with Duchenne muscular dystrophy. *Nature*, **318**, 672-675.

39. Monaco,A.P., Bertelson,C.J., Middlesworth,W. et al (1985) Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. *Nature*, **316**, 842-845.

40. Darras,B.T., Francke,U. (1988) Normal human genomic restriction-fragment patterns and polymorphisms revealed by hybridization with the entire dystrophin cDNA. *Am.J.Hum.Genet.*, **43**, 612-619.

41. Den Dunnen,J.T., Grootscholten,P.M., Bakker,E. et al (1989) Topography of the DMD gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. *Am.J.Hum.Genet.*, **45**, 835-847.

42. Hu,X., Burghes,A.H.M., Ray,P.N. et al (1988) Partial gene duplications in Duchenne and Becker muscular dystrophies. *J.Med.Genet.*, **25**, 369-376.

43. Oudet,C., Hanauer,A., Clemens,P. et al (1992) Two hot spots of recombination in the DMD-gene correlate with the deletion prone regions. *Hum.Mol.Genet.*, **1**, 599-603.

44. Forrest,S.M., Cross,G.S., Speer,A. et al (1987) Preferential deletion of exons in Duchenne and Becker Muscular Dystrophies. *Nature*, **329**, 638-640.

45. Beggs,A.H., Koenig,M., Boyce,F.M. et al (1990) Detection of 98-percent DMD/BMD gene deletions by polymerase chain reaction. *Hum.Genet.*, **86**, 45-48.

46. Chamberlain,J.S., Gibbs,R.A., Ranier,J.E. et al (1988) Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucl.Acids Res.*, **23**, 11141-11156.

47. White,S., Kalf,M., Liu,Q. et al (2002) Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. *Am.J.Hum.Genet.*, **71**, 365-374.

48. England,S.B., Nicholson,L.V.B., Johnson,M.A. et al (1990) Very mild muscular dystrophy associated with the deletion of 46% of dystrophin. *Nature*, **343**, 180-182.

49. Love,D.R., Flint,T.J., Genet,S.A. et al (1991) Becker muscular dystrophy patient with a large intragenic dystrophin deletion: implications for functional minigenes and gene therapy. *J.Med.Genet.*, **28**, 860-864.

50. Monaco,A.P., Bertelson,C.J., Liechti-Gallati,S. et al (1988) An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics*, **2**, 90-95.

51. Muntoni,F., Gobbi,P., Sewry,C. et al (1994) Deletions in the 5' region of dystrophin and resulting phenotypes. *J.Med.Genet.*, **31**, 843-847.

52. Winnard,A.V., Mendell,J.R., Prior,T.W. et al (1995) Frameshift deletions of exons 3-7 and revertant fibers in Duchenne muscular dystrophy: mechanisms of dystrophin production. *Am.J.Hum.Genet.*, **56**, 158-166.

53. Sherratt,T.G., Vulliamy,T., Dubowitz,V. et al (1993) Exon skipping and translation in patients with frameshift deletions in the dystrophin gene. *Am.J.Hum.Genet.*, **53**, 1007-1015.

54. Prior,T.W., Bartolo,C., Papp,A.C. et al (1997) Dystrophin expression in a Duchenne muscular dystrophy patient with a frame shift deletion. *Neurol.*, **48**, 486-488.

55. Prior,T.W., Bartolo,C., Papp,A.C. et al (1994) Identification of a missense mutation, single base deletion and a polymorphism in the dystrophin exon 16. *Hum.Mol.Genet.*, **3**, 1173-1174.

56. Prior,T.W., Papp,A.C., Snyder,P.J. et al (1993) A missense mutation in the dystrophin gene in a Duchenne muscular dystrophy patient. *Nat.Genet.*, **4**, 357-360.

57. Ishikawa-Sakurai,M., Yoshida,M., Imamura,M. et al (2004) ZZ domain is essentially required for the physiological binding of dystrophin and utrophin to beta-dystroglycan. *Hum.Mol.Genet.*, **13**, 693-702.

58. Prior,T.W., Wenger,G.D., Papp,A.C. et al (1995) Rapid DNA haplotyping using a multiplex heteroduplex approach: application to Duchenne muscular dystrophy carrier testing. *Hum.Mutat.*, **5**, 263-268.

59. Prior,T.W., Papp,A.C., Snyder,P.J. et al (1994) Heteroduplex analysis of the dystrophin gene: application to point mutation and carrier detection. *Am.J.Med.Genet.*, **50**, 68-73.

60. Bennett,R.R., Dunnen,J., O'Brien,K.F. et al (2001) Detection of mutations in the dystrophin gene via automated DHPLC screening and direct sequencing. *BMC.Genet.*, **2**, 17.

61. Mendell,J.R., Buzin,C.H., Feng,J. et al (2001) Diagnosis of Duchenne dystrophy by enhanced detection of small mutations. *Neurol.*, **57**, 645-650.

62. Hofstra,R.M., Mulder,I.M., Vossen,R. et al (2004) DGGE-based whole-gene mutation scanning of the dystrophin gene in Duchenne and Becker muscular dystrophy patients. *Hum.Mutat.*, **23**, 57-66.

63. Flanigan,K.M., von Niederhausern,A., Dunn,D.M. et al (2003) Rapid direct sequence analysis of the dystrophin gene. *Am.J.Hum.Genet.*, **72**, 931-939.

64. Yan,J., Feng,J., Buzin,C.H. et al (2004) Three-tiered noninvasive diagnosis in 96% of patients with Duchenne muscular dystrophy (DMD). *Hum.Mutat.*, **23**, 203-204.

65. Roest,P.A.M., Roberts,R.G., Van Der Tuijn,A.C. et al (1993) Protein truncation test (PTT) to rapidly screen the DMD-gene for translation-terminating mutations. *Neuromusc.Disord.*, **3**, 391-394.

66. Roest,P.A.M., Bout,M., Van Der Tuijn,A.C. et al (1996) Splicing mutations in DMD/BMD detected by RT-PCR/PTT: detection of a 19AA insertion in the cysteine-rich domain of dystrophin compatible with BMD. *J.Med.Genet.*, **33**, 935-939.

67. Beroud,C., Carrie,A., Beldjord,C. et al (2004) Dystrophinopathy caused by mid-intronic substitutions activating cryptic exons in the DMD gene. *Neuromuscul.Disord.*, **14**, 10-18.

68. Tuffery-Giraud,S., Chambert,S., Demaille,J. et al (1999) Point mutations in the dystrophin gene: evidence for frequent use of cryptic splice sites as a result of splicing defects. *Hum.Mutat.*, **14**, 359-368.

69. van Deutekom,J.C., van Ommen,G.J. (2003) Advances in Duchenne muscular dystrophy gene therapy. *Nat.Rev.Genet.*, **4**, 774-783.

70. van Deutekom,J.C., Bremmer-Bout,M., Janson,A.A.M. et al (2001) Antisense-induced exon skipping restores dystrophin expression in DMD patient-derived muscle cells. *Hum.Mol.Genet.*, **10**, 1547-1554.

71. Aartsma-Rus,A., Janson,A.A., Kaman,W.E. et al (2004) Antisense-induced multiexon skipping for Duchenne muscular dystrophy makes more sense. *Am.J.Hum.Genet.*, **74**, 83-92.

72. Bullock,D.G., McSweeny,F.M., Whitehead,T.P. et al (1979) Serum creatine kinase activity and carrier status for Duchenne muscular dystrophy. *Lancet*, **2**, 1151-1152.

73. Yau,S.C., Bobrow,M., Mathew,C.G. et al (1996) Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis. *J Med Genet*, **33**, 550-558.

74. Bakker,E., Hofker,M.H., Goor,N. et al (1985) Prenatal diagnosis and carrier detection of Duchenne Muscular Dystrophy with closely linked RFLPs. *Lancet*, **i**, 655-658.

75. Bakker,E., Bonten,E.J., de Lange,L.F. et al (1986) DNA probe analysis for carrier detection and prenatal diagnosis of Duchenne muscular dystrophy: A standard diagnostic procedure. *J.Med.Genet.*, **23**, 573-580.

76. Clemens,P.R., Fenwick,R.G., Chamberlain,J.S. et al (1991) Carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophy families, using dinucleotide repeat polymorphisms. *Am.J.Hum.Genet.*, **49**, 951-960.

77. Schwartz,L.S., Tarleton,J., Popovich,B. et al (1992) Fluorescent multiplex analysis and carrier detection for Duchenne/Becker muscular dystrophy. *Am.J.Hum.Genet.*, **51**, 721-729.

78. White,S.J., Sterrenburg,E., van Ommen,G.J. et al (2003) An alternative to FISH: detecting deletion and duplication carriers within 24 hours. *J.Med.Genet.*, **40**, e113.

79. Joncourt,F., Neuhaus,B., Jostarndt-Foegen,K. et al (2004) Rapid identification of female carriers of DMD/BMD by quantitative real-time PCR. *Hum.Mutat.*, **23**, 385-391.

80. Bakker,E., Van Broeckhoven,C., Bonten,E.J. et al (1987) Germline Mosaicism and Duchenne Muscular Dystrophy mutations. *Nature*, **329**, 554-556.

81. Bakker,E., Bonten,E.J., Veenema,H. et al (1989) Prenatal diagnosis of Duchenne muscular dystrophy: a three-year experience in a rapidly evolving field. *J.Inherited Metab.Dis.*, **12**, 174-190.

82. Van Essen,A.J., Abbs,S.J., Baiget,M. et al (1992) Parental origin and germline mosaicism of deletions and duplications of the dystrophin gene: a European study. *Hum.Genet.*, **88**, 249-257.

83. Van Essen,A.J., Kneppers,A.L.J., van der Hout,A.H. et al (1997) The clinical and molecular genetic approach to Duchenne and Becker muscular dystrophy: an updated protocol. *J.Med.Genet.*, **34**, 805-812.

84. Bakker,E., Veenema,H., Den Dunnen,J.T. et al (1989) Germinal mosaicism increases the recurrence risk for "new" Duchenne muscular dystrophy mutations. *J.Med.Genet.*, **26**, 553-559.

85. Wu,T.C., Lichten,M. (1994) Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science*, **263**, 515-518.

86. Borde,V., Goldman,A.S., and Lichten,M. (2000) Direct coupling between meiotic DNA replication and recombination initiation. *Science*, **290**, 806-809.

87. Fugmann,S.D., Lee,A.I., Shockett,P.E. et al (2000) The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu.Rev.Immunol.*, **18**, 495-527.

88. Haber,J.E. (2000) Partners and pathwaysrepairing a double-strand break. *Trends Genet.*, **16**, 259-264.

89. Jackson,S.P. (2002) Sensing and repairing DNA double-strand breaks. *Carcinogenesis*, **23**, 687-696.

90. Cromie,G.A., Connelly,J.C., and Leach,D.R. (2001) Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol.Cell*, **8**, 1163-1174.

91. Sonoda,E., Takata,M., Yamashita,Y.M. et al (2001) Homologous DNA recombination in vertebrate cells. *Proc.Natl.Acad.Sci.U.S.A*, **98**, 8388-8394.

92. Li,W.H., Gu,Z., Wang,H. et al (2001) Evolutionary analyses of the human genome. *Nature*, **409**, 847-849.

93. Lander,E.S., Linton,L.M., Birren,B. et al (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

94. Prak,E.T., Kazazian,H.H.jr. (2000) Mobile elements and the human genome. *Nat.Rev.Genet.*, **1**, 134-144.

95. Deininger,P.L., Batzer,M.A. (1999) Alu repeats and human disease. *Mol.Genet.Metab*, **67**, 183-193.

96. Smit,A.F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr.Opin.Genet.Dev.*, **9**, 657-663.

97. Jurka,J., Klonowski,P. (1996) Integration of retroposable elements in mammals: selection of target sites. *J.Mol.Evol.*, **43**, 685-689.

98. Burwinkel,B., Kilimann,M.W. (1998) Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *J.Mol.Biol.*, **277**, 513-517.

99. Segal,Y., Peissel,B., Renieri,A. et al (1999) LINE-1 elements at the sites of molecular rearrangements in Alport syndrome-diffuse leiomyomatosis. *Am.J.Hum.Genet.*, **64**, 62-69.

100. Houck,C.M., Rinehart,F.P., and Schmid,C.W. (1979) A ubiquitous family of repeated DNA sequences in the human genome. *J.Mol.Biol.*, **132**, 289-306.

101. Harteveld,K.L., Losekoot,M., Fodde,R. et al (1997) The involvement of Alu repeats in recombination events at the alpha-globin gene cluster: characterization of two alphazero-thalassaemia deletion breakpoints. *Hum.Genet.*, **99**, 528-534.

102. Kolomietz,E., Meyn,M.S., Pandita,A. et al (2002) The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. *Genes Chromosomes.Cancer*, **35**, 97-112.

103. Bailey,J.A., Yavor,A.M., Massa,H.F. et al (2001) Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.*, **11**, 1005-1017.

104. Tuzun,E., Bailey,J.A., and Eichler,E.E. (2004) Recent segmental duplications in the working draft assembly of the brown norway rat. *Genome Res.*, **14**, 493-506.

105. Cheung,J., Wilson,M.D., Zhang,J. et al (2003) Recent segmental and gene duplications in the mouse genome. *Genome Biol.*, **4**, R47.

106. Bailey,J.A., Gu,Z., Clark,R.A. et al (2002) Recent segmental duplications in the human genome. *Science*, **297**, 1003-1007.

107. Eichler,E.E. (2001) Segmental duplications: what's missing, misassigned, and misassembled--and should we care? *Genome Res.*, **11**, 653-656.

108. Bailey,J.A., Liu,G., and Eichler,E.E. (2003) An Alu transposition model for the origin and expansion of human segmental duplications. *Am.J.Hum.Genet.*, **73**, 823-834.

109. Liu,G., Zhao,S., Bailey,J.A. et al (2003) Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.*, **13**, 358-368.

110. Samonte,R.V., Eichler,E.E. (2002) Segmental duplications and the evolution of the primate genome. *Nat.Rev.Genet.*, **3**, 65-72.

111. van Deutekom,J.C., Bakker,E., Lemmers,R.J. et al (1996) Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. *Hum.Mol.Genet.*, **5**, 1997-2003.

112. Lemmers,R.J., Van Der Maarel,S.M., van der Wielen,M.J. et al (1998) Inter- and intrachromosomal sub-telomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum.Mol.Genet.*, **7**, 1207-1214.

113. Knight,S.J., Lese,C.M., Precht,K.S. et al (2000) An optimized set of human telomere clones for studying telomere integrity and architecture. *Am.J.Hum.Genet.*, **67**, 320-332.

114. Eichler,E.E., Clark,R.A., and She,X. (2004) An assessment of the sequence gaps: Unfinished business in a finished human genome. *Nat.Rev.Genet.*, **5**, 345-354.

115. Kurahashi,H., Shaikh,T.H., Hu,P. et al (2000) Regions of genomic instability on 22q11 and 11q23 as the etiology for the recurrent constitutional t(11;22). *Hum.Mol.Genet.*, **9**, 1665-1670.

116. Kurahashi,H., Shaikh,T., Takata,M. et al (2003) The constitutional t(17;22): another translocation mediated by palindromic AT-rich repeats. *Am.J.Hum.Genet.*, **72**, 733-738.

117. Dehal,P., Predki,P., Olsen,A.S. et al (2001) Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution. *Science*, **293**, 104-111.

118. Armengol,L., Pujana,M.A., Cheung,J. et al (2003) Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. *Hum.Mol.Genet.*, **12**, 2201-2208.

119. Bailey,J.A., Baertsch,R., Kent,W.J. et al (2004) Hotspots of mammalian chromosomal evolution. *Genome Biol.*, **5**, R23.

120. Eichler,E.E. (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.*, **17**, 661-669.

121. Bi,W., Park,S.S., Shaw,C.J. et al (2003) Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2. *Am.J.Hum.Genet.*, **73**, 1302-1315.

122. Reiter,L.T., Hastings,P.J., Nelis,E. et al (1998) Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. *Am.J.Hum.Genet.*, **62**, 1023-1033.

123. Waldman,A.S., Liskay,R.M. (1988) Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol.Cell Biol.*, **8**, 5350-5357.

124. Osborne,L.R., Li,M., Pober,B. et al (2001) A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. *Nat.Genet.*, **29**, 321-325.

125. Bayes,M., Magano,L.F., Rivera,N. et al (2003) Mutational mechanisms of Williams-Beuren syndrome deletions. *Am.J.Hum.Genet.*, **73**, 131-151.

126. Giglio,S., Calvari,V., Gregato,G. et al (2002) Heterozygous submicroscopic inversions involving olfactory receptor-gene clusters mediate the recurrent t(4;8)(p16;p23) translocation. *Am.J.Hum.Genet.*, **71**, 276-285.

127. Giglio,S., Broman,K.W., Matsumoto,N. et al (2001) Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. *Am.J.Hum.Genet.*, **68**, 874-883.

128. Kurotaki,N., Imaizumi,K., Harada,N. et al (2002) Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat.Genet.*, **30**, 365-366.

129. Kurotaki,N., Harada,N., Shimokawa,O. et al (2003) Fifty microdeletions among 112 cases of Sotos syndrome: low copy repeats possibly mediate the common deletion. *Hum.Mutat.*, **22**, 378-387.

130. Miyake,N., Kurotaki,N., Sugawara,H. et al (2003) Preferential paternal origin of microdeletions caused by prezygotic chromosome or chromatid rearrangements in sotos syndrome. *Am.J.Hum.Genet.*, **72**, 1331-1337.

131. Peoples,R., Franke,Y., Wang,Y.K. et al (2000) A physical map, including a BAC/PAC clone contig, of the Williams-Beuren syndrome--deletion region at 7q11.23. *Am.J.Hum.Genet.*, **66**, 47-68.

132. Urban,Z., Helms,C., Fekete,G. et al (1996) 7q11.23 deletions in Williams syndrome arise as a consequence of unequal meiotic crossover. *Am.J.Hum.Genet.*, **59**, 958-962.

133. Francke,U. (1999) Williams-Beuren syndrome: genes and mechanisms. *Hum.Mol.Genet.*, **8**, 1947-1954.

134. Robinson,W.P., Waslynka,J., Bernasconi,F. et al (1996) Delineation of 7q11.2 deletions associated with Williams-Beuren syndrome and mapping of a repetitive sequence to within and to either side of the common deletion. *Genomics*, **34**, 17-23.

135. Amos-Landgraf,J.M., Ji,Y., Gottlieb,W. et al (1999) Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am.J.Hum.Genet.*, **65**, 370-386.

136. Robinson,W.P., Dutly,F., Nicholls,R.D. et al (1998) The mechanisms involved in formation of deletions and duplications of 15q11-q13. *J.Med.Genet.*, **35**, 130-136.

137. Slager,R.E., Newton,T.L., Vlangos,C.N. et al (2003) Mutations in RAI1 associated with Smith-Magenis syndrome. *Nat.Genet.*, **33**, 466-468.

138. Potocki,L., Chen,K.S., Park,S.S. et al (2000) Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat.Genet.*, **24**, 84-87.

139. Shaw,C.J., Bi,W., and Lupski,J.R. (2002) Genetic proof of unequal meiotic crossovers in reciprocal deletion and duplication of 17p11.2. *Am.J.Hum.Genet.*, **71**, 1072-1081.

140. Inoue,K., Dewar,K., Katsanis,N. et al (2001) The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. *Genome Res.*, **11**, 1018-1033.

141. Shaw,C.J., Shaw,C.A., Yu,W. et al (2004) Comparative genomic hybridisation using a proximal 17p BAC/PAC array detects rearrangements responsible for four genomic disorders. *J.Med.Genet.*, **41**, 113-119.

142. Valentijn,L.J., Bolhuis,P.A., Zorn,I. et al (1992) The peripheral myelin gene PMP22/GAS3 is duplicated in Charcot-Marie-Tooth disease type 1A. *Nat.Genet.*, **1**, 166-170.

143. Pentao,L., Wise,C.A., Chinault,A.C. et al (1992) Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nat.Genet.*, **2**, 292-300.

144. Lupski,J.R., Wise,C.A., Kuwano,A. et al (1992) Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A. *Nat.Genet.*, **1**, 29-33.

145. Riva,P., Castorina,P., Manoukian,S. et al (1996) Characterization of a cytogenetic 17q11.2 deletion in an NF1 patient with a contiguous gene syndrome. *Hum.Genet.*, **98**, 646-650.

146. Lopez-Correa,C., Dorschner,M., Brems,H. et al (2001) Recombination hotspot in NF1 microdeletion patients. *Hum.Mol.Genet.*, **10**, 1387-1392.

147. Dorschner,M.O., Sybert,V.P., Weaver,M. et al (2000) NF1 microdeletion breakpoints are clustered at flanking repetitive sequences. *Hum.Mol.Genet.*, **9**, 35-46.

148. Jenne,D.E., Tinschert,S., Reimann,H. et al (2001) Molecular characterization and gene content of breakpoint boundaries in patients with neurofibromatosis type 1 with 17q11.2 microdeletions. *Am.J.Hum.Genet.*, **69**, 516-527.

149. Shaikh,T.H., Kurahashi,H., Saitta,S.C. et al (2000) Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum.Mol.Genet.*, **9**, 489-501.

150. Edelmann,L., Pandita,R.K., Spiteri,E. et al (1999) A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum.Mol.Genet.*, **8**, 1157-1167.

151. Edelmann,L., Pandita,R.K., and Morrow,B.E. (1999) Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. *Am.J.Hum.Genet.*, **64**, 1076-1086.

152. Bione,S., Maestrini,E., Rivella,S. et al (1994) Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nat.Genet.*, **8**, 323-327.

153. Small,K., Iber,J., and Warren,S.T. (1997) Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. *Nat.Genet.*, **16**, 96-99.

154. Small,K., Warren,S.T. (1998) Emerin deletions occurring on both Xq28 inversion backgrounds. *Hum.Mol.Genet.*, **7**, 135-139.

155. Rossiter,J.P., Young,M., Kimberland,M.L. et al (1994) Factor VIII gene inversions causing severe hemophilia A originate almost exclusively in male germ cells. *Hum.Mol.Genet.*, **3**, 1035-1039.

156. Naylor,J.A., Buck,D., Green,P. et al (1995) Investigation of the factor VIII intron 22 repeated region (int22h) and the associated inversion junctions. *Hum.Mol.Genet.*, **4**, 1217-1224.

157. Naylor,J.A., Nicholson,P., Goodeve,A. et al (1996) A novel DNA inversion causing severe hemophilia A. *Blood*, **87**, 3255-3261.

158. Abeysinghe,S.S., Chuzhanova,N., Krawczak,M. et al (2003) Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. *Hum.Mutat.*, **22**, 229-244.

159. Stahl,M.M., Kobayashi,I., Stahl,F.W. et al (1983) Activation of Chi, a recombinator, by the action of an endonuclease at a distant site. *Proc.Natl.Acad.Sci.U.S.A*, **80**, 2310-2313.

160. Wyatt,R.T., Rudders,R.A., Zelenetz,A. et al (1992) BCL2 oncogene translocation is mediated by a chi-like consensus. *J.Exp.Med.*, **175**, 1575-1588.

161. Marshall,B., Isidro,G., and Boavida,M.G. (1996) Insertion of a short Alu sequence into the hMSH2 gene following a double cross over next to sequences with chi homology. *Gene*, **174**, 175-179.

162. Schmucker,B., Ballhausen,W.G., and Pfeiffer,R.A. (1996) Mosaicism of a microdeletion of 486 bp involving the CGG repeat of the FMR1 gene due to misalignment of GTT tandem repeats at chi-like elements flanking both breakpoints and a full mutation. *Hum.Genet.*, **98**, 409-414.

163. Laccone,F., Junemann,I., Whatley,S. et al (2004) Large deletions of the MECP2 gene detected by gene dosage analysis in patients with Rett syndrome. *Hum.Mutat.*, **23**, 234-244.

164. Rudiger,N.S., Gregersen,N., and Kielland-Brandt,M.C. (1995) One short well conserved region of Alu-sequences is involved in human gene rearrangements and has homology with prokaryotic chi. *Nucleic Acids Res.*, **23**, 256-260.

165. Chuzhanova,N., Abeysinghe,S.S., Krawczak,M. et al (2003) Translocation and gross deletion breakpoints in human inherited disease and cancer II: Potential involvement of repetitive sequence elements in secondary structure formation between DNA ends. *Hum.Mutat.*, **22**, 245-251.

166. Sargent,R.G., Brenneman,M.A., and Wilson,J.H. (1997) Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol.Cell Biol.*, **17**, 267-277.

167. Helleday,T. (2003) Pathways for mitotic homologous recombination in mammalian cells. *Mutat.Res.*, **532**, 103-115.

168. Pozzoli,U., Sironi,M., Cagliani,R. et al (2002) Comparative analysis of the human dystrophin and utrophin gene structures. *Genetics*, **160**, 793-798.

169. Pozzoli,U., Elgar,G., Cagliani,R. et al (2003) Comparative analysis of vertebrate dystrophin loci indicate intron gigantism as a common feature. *Genome Res.*, **13**, 764-772.

170. Blonden,L.A.J., Grootscholten,P.M., Den Dunnen,J.T. et al (1991) 242 breakpoints in the 200-kb deletion-prone P20 region of the DMD-gene are widely spread. *Genomics*, **10**, 631-639.

171. McNaughton,J.C., Cockburn,D.J., Hughes,G. et al (1998) Is gene deletion in eukaryotes sequence-dependent? A study of nine deletion junctions and nineteen other deletion breakpoints in intron 7 of the human dystrophin gene. *Gene*, **222**, 41-51.

172. Sironi,M., Pozzoli,U., Cagliani,R. et al (2003) Relevance of sequence and structure elements for deletion events in the dystrophin gene major hot-spot. *Hum.Genet.*, **112**, 272-288.

173. Nobile,C., Toffolatti,L., Rizzi,F. et al (2002) Analysis of 22 deletion breakpoints in dystrophin intron 49. *Hum.Genet.*, **110**, 418-421.

174. Boulikas,T. (1995) Chromatin domains and prediction of MAR sequences. *Int.Rev.Cytol.*, **162A**, 279-388.

175. Svetlova,E.Y., Razin,S.V., and Debatisse,M. (2001) Mammalian recombination hot spot in a DNA loop anchorage region: A model for the study of common fragile sites. *J.Cell Biochem.*, **81**, 170-178.

176. Verbovaia,L.V., Razin,S.V. (1997) Mapping of replication origins and termination sites in the Duchenne muscular dystrophy gene. *Genomics*, **45**, 24-30.

177. Toledo,F., Coquelle,A., Svetlova,E. et al (2000) Enhanced flexibility and aphidicolin-induced DNA breaks near mammalian replication origins: implications for replicon mapping and chromosome fragility. *Nucleic Acids Res.*, **28**, 4805-4813.

178. Bierne,H., Ehrlich,S.D., and Michel,B. (1991) The replication termination signal terB of the Escherichia coli chromosome is a deletion hot spot. *EMBO J.*, **10**, 2699-2705.

179. Horiuchi,T., Nishitani,H., and Kobayashi,T. (1995) A new type of E. coli recombinational hotspot which requires for activity both DNA replication termination events and the Chi sequence. *Adv.Biophys.*, **31**, 133-147.

180. Iarovaia,O.V., Bystritskiy,A., Ravcheev,D. et al (2004) Visualization of individual DNA loops and a map of loop domains in the human dystrophin gene. *Nucleic Acids Res.*, **32**, 2079-2086.

181. Passos-Bueno,M.R., Bakker,E., Kneppers,A.L.J. et al (1992) Different mosaicism frequencies for proximal and distal Duchenne muscular dystrophy (DMD) mutations indicate difference in etiology and recurrence risk. *Am.J.Hum.Genet.*, **51**, 1150-1155.

182. Hu,X., Burghes,A.H.M., Bulman,D.E. et al (1989) Evidence for mutation by unequal sister chromatid exchange in the Duchenne muscular dystrophy gene. *Am.J.Hum.Genet.*, **44**, 855-863.

183. Hu,X., Ray,P.N., Murphy,E. et al (1990) Duplicational mutation at the Duchenne muscular dystrophy locus: its frequency, distribution, origin and phenotype/genotype correlation. *Am.J.Hum.Genet.*, **46**, 682-695.

184. Darras,B.T., Blattner,P., Harper,J.F. et al (1988) Intragenic deletions in 21 Duchenne muscular dystrophy (DMD) / Becker muscular dystrophy families studied with the dystrophin cDNA: location of breakpoints on HindIII and BglII exon-containing fragment maps, meiotic and mitotic origin of the mutations. *Am.J.Hum.Genet.*, **43**, 620-629.

185. Grimm,T., Meng,G., Liechti-Gallati,S. et al (1994) On the origin of deletions and point mutations in Duchenne muscular dystrophy: most deletions arise in oogenesis and most point mutations result from events in spermatogenesis. *J.Med.Genet.*, **31**, 183-186.

186. Ketterling,R.P., Vielhaber,E., Bottema,C.D. et al (1993) Germ-line origins of mutation in families with hemophilia B: the sex ratio varies with the type of mutation. *Am.J.Hum.Genet.*, **52**, 152-166.

187. Becker,J., Schwaab,R., Moller-Taube,A. et al (1996) Characterization of the factor VIII defect in 147 patients with sporadic hemophilia A: family studies indicate a mutation type-dependent sex ratio of mutation frequencies. *Am.J.Hum.Genet.*, **58**, 657-670.

188. Hu,X., Ray,P.N., and Worton,R.G. (1991) Mechanisms of tandem duplication in the Duchenne muscular dystrophy gene include both homologous and nonhomologous intrachromosomal recombination. *EMBO J.*, **10**, 2471-2477.

189. Gale,K.C., Osheroff,N. (1992) Intrinsic intermolecular DNA ligation activity of eukaryotic topoisomerase II. Potential roles in recombination. *J.Biol.Chem.*, **267**, 12090-12097.

190. Wang,J.C., Caron,P.R., and Kim,R.A. (1990) The role of DNA topoisomerases in recombination and genome stability: a double-edged sword ? *Cell*, **62**, 403-406.

191. Armour,J.A., Sismani,C., Patsalis,P.C. et al (2000) Measurement of locus copy number by hybridisation with amplifiable probes. *Nucl.Acids Res.*, **28**, 605-609.

192. Reid,A.G., Tarpey,P.S., and Nacheva,E.P. (2003) High-resolution analysis of acquired genomic imbalances in bone marrow samples from chronic myeloid leukemia patients by use of multiple short DNA probes. *Genes Chromosomes.Cancer*, **37**, 282-290.

193. Altschul,S.F., Gish,W., Miller,W. et al (1990) Basic local alignment search tool. *J.Mol.Biol.*, **215**, 403-410.

194. Kent,W.J. (2002) BLAT--the BLAST-like alignment tool. *Genome Res.*, **12**, 656-664.

195. Schouten,J.P., McElgunn,C.J., Waaijer,R. et al (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.*, **30**, e57.

196. Hollox,E.J., Atia,T., Cross,G. et al (2002) High throughput screening of human subtelomeric DNA for copy number changes using multiplex amplifiable probe hybridisation (MAPH). *J.Med.Genet.*, **39**, 790-795.

197. Montagna,M., Dalla,P.M., Menin,C. et al (2003) Genomic rearrangements account for more than one-third of the BRCA1 mutations in northern Italian breast/ovarian cancer families. *Hum.Mol.Genet.*, **12**, 1055-1061.

198. Rooms,L., Reyniers,E., van Luijk,R. et al (2004) Subtelomeric deletions detected in patients with idiopathic mental retardation using multiplex ligation-dependent probe amplification (MLPA). *Hum.Mutat.*, **23**, 17-21.

199. Flint,J., Knight,S. (2003) The use of telomere probes to investigate submicroscopic rearrangements associated with mental retardation. *Curr.Opin.Genet.Dev.*, **13**, 310-316.

200. Sismani,C., Armour,J.A., Flint,J. et al (2001) Screening for subtelomeric chromosome abnormalities in children with idiopathic mental retardation using multiprobe telomeric FISH and the new MAPH telomeric assay. *Eur.J.Hum.Genet.*, **9**, 527-532.

201. Kriek,M., White,S.J., Bouma,M.C. et al (2004) Genomic imbalances in mental retardation. *J.Med.Genet.*, **41**, 249-255.

202. Hollox,E.J., Armour,J.A., and Barber,J.C. (2003) Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. *Am.J.Hum.Genet.*, **73**, 591-600.

203. Hogervorst,F.B., Nederlof,P.M., Gille,J.J. et al (2003) Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. *Cancer Res.*, **63**, 1449-1453.

204. Gille,J.J., Hogervorst,F.B., Pals,G. et al (2002) Genomic deletions of MSH2 and MLH1 in colorectal cancer families detected by a novel mutation detection approach. *Br.J.Cancer*, **87**, 892-897.

205. Taylor,C.F., Charlton,R.S., Burn,J. et al (2003) Genomic deletions in MSH2 or MLH1 are a frequent cause of hereditary non-polyposis colorectal cancer: identification of novel and recurrent deletions by MLPA. *Hum.Mutat.*, **22**, 428-433.

206. Slater,H.R., Bruno,D.L., Ren,H. et al (2003) Rapid, high throughput prenatal detection of aneuploidy using a novel quantitative method (MLPA). *J.Med.Genet.*, **40**, 907-912.