Detecting copy number changes in genomic DNA - MAPH and MLPA

White, S.J.

Detecting copy number changes in genomic DNA – MAPH and MLPA

Stefan J. White

Detecting copy number changes
in genomic DNA – MAPH and MLPA

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 3 februari 2005
klokke 16.15 uur

door

Stefan John White

geboren te Lower Hutt, Nieuw Zeeland
in 1973

Promotiecommissie

Promotores:

Prof. Dr. M.H. Breuning
Prof. Dr. G.-J.B. van Ommen

Co-promotor:

Dr. J.T. den Dunnen

Referent:

Prof. Dr. H.G. Brunner (Universiteit Nijmegen)

Overige leden:

Prof. Dr. E. Bakker
Prof. Dr. H.J. Tanke

Printed by: Febodruk BV, Enschede, The Netherlands

Cover image of Mitre Peak and Milford Sound, from DavidWallPhoto.

The studies in this thesis were performed in the Center for Human and Clinical Genetics, Faculty of Medicine, Leiden University Medical Center, Leiden University, The Netherlands and were supported by ZonMw grant 912-04-047.

The publication of this thesis was financially supported by a grant from Agilent Technologies.

Is life just a game where we make up the rules,
While we're searching for something to say,
Or are we just simply spiralling coils,
Of self-replicating DNA?

Monty Python – The Meaning of Life

For my parents

Contents

Chapter 1

General Introduction 11

A. Duchenne Muscular Dystrophy 13

DMD gene 13

Dystrophin 13

Deletion and duplication detection in patients 14

Other mutations 15

Carrier detection 16

How do deletions and duplications occur? 17

Homologous recombination 17

LINEs 18

SINEs 18

Segmental duplications 19

Interchromosomal duplications 19

Intrachromosomal duplications 20

Non-homologous end joining 21

Rearrangements in the *DMD* gene 22

B. MAPH and MLPA 27

MAPH technique 27

MLPA technique 30

Analysis 31

Applications 35

C. Discussion 38

References 41

Chapter 2. Deletions and duplications in the *DMD* gene

Chapter 2.1

Comprehensive detection of genomic duplications and deletions in the *DMD* gene, by use of multiplex amplifiable probe hybridization (2002). *Am.J.Hum. Genet.* 71 (2):365-374. 53

Chapter 2.2

Duplications in the *DMD* gene are nonrandomly distributed and can be complex rearrangements. Manuscript in preparation. 65

Chapter 2.3

An alternative to FISH: detecting deletion and duplication carriers within 24 hours (2003). *J.Med.Genet.* 40 (10):e13. 79

Chapter 3. MAPH and LGMD

Sarcoglycanopathies and the risk of undetected deletion alleles in diagnosis. Submitted for publication. 85

Chapter 4. MAPH and mental retardation

Genomic imbalances in mental retardation (2004). *J.Med.Genet.* 41 (4):249-255. 97

Chapter 5. MLPA and intragenic deletions

Chapter 5.1

Two-color multiplex ligation-dependent probe amplification: Detecting genomic rearrangements in hereditary multiple exostoses (2004). *Hum. Mutat.* 24 (1):86-92. 107

Chapter 5.2

Genetic Heterogeneity in Rubinstein-Taybi syndrome: mutations in the *CBP* and *EP300* gene are both disease causing. Submitted for publication. 117

Chapter 6. MLPA and duplicon analysis

Complex SNP-related sequence variation in segmental genome duplications (2004). *Nat. Genet.* 36 (8):861-866. 133

Summary 143

Samenvatting 145

Publications 147

Curriculum vitae 149

Appendix

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).

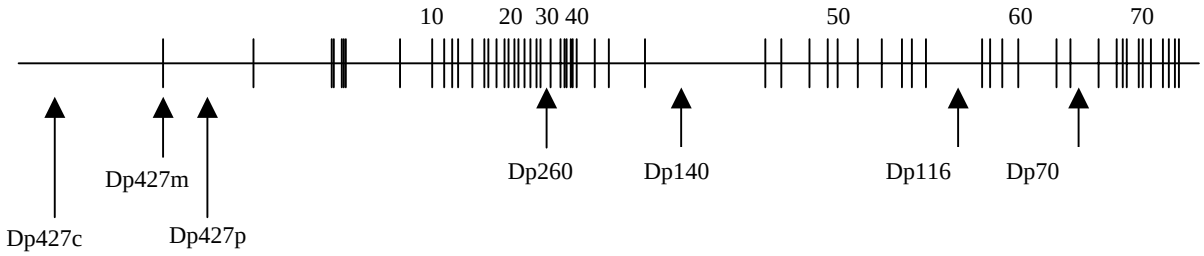

Figure 1. The *DMD* gene. In this diagram the promoter for the muscle isoform (Dp427m) is shown as exon 1, and the locations of the other promoters are indicated with arrows. Every 10th exon is numbered.

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).

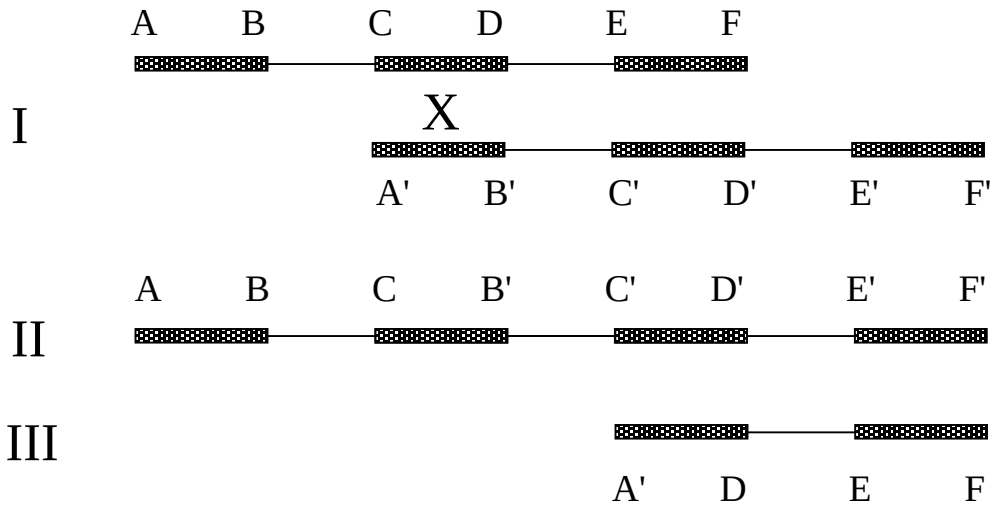

Figure 2. Non-allelic Homologous Recombination. Misalignment of highly similar sequences (I) can lead to non-equal crossing over, resulting in (II) a duplication and (III) the reciprocal deletion.

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

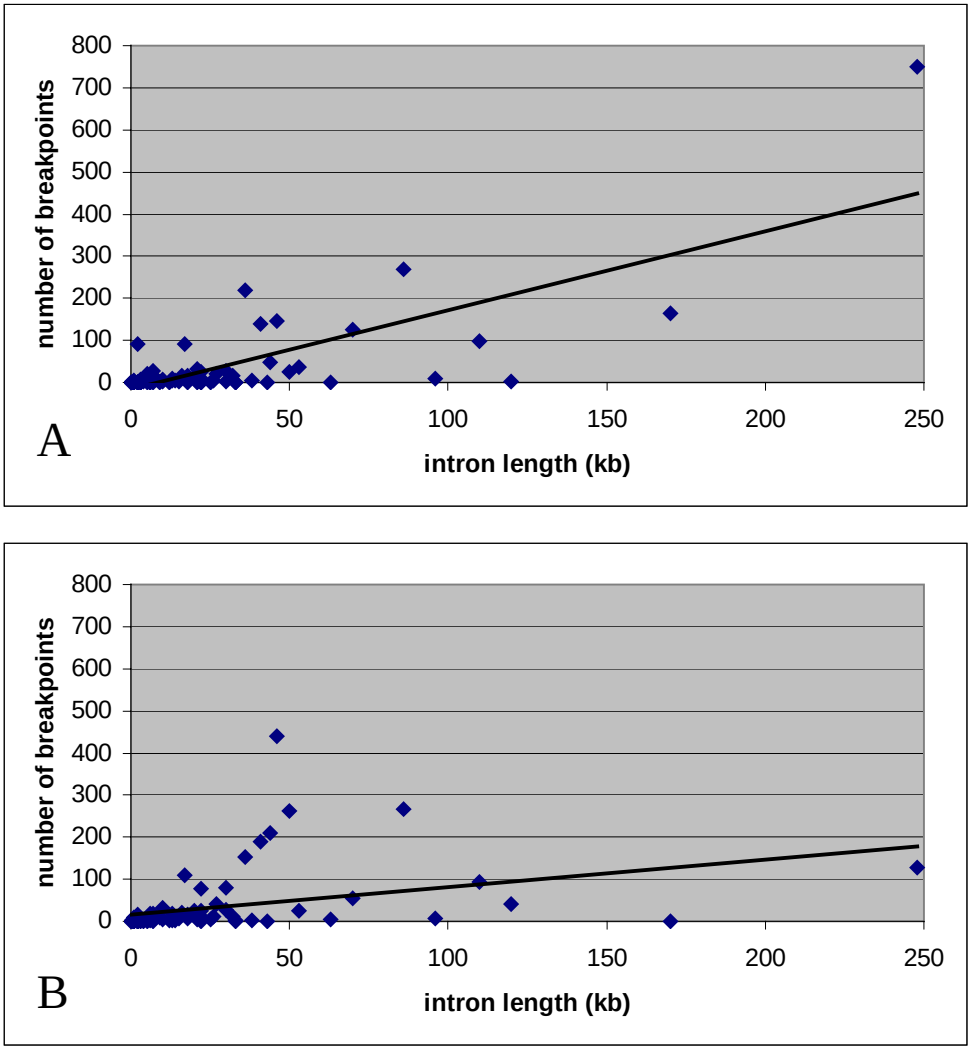

Figure 3. The relationship between intron size and breakpoint frequency for deletions within the *DMD* gene. A. 5' breakpoint frequency. B. 3' breakpoint frequency. The correlations are A; r=0.75, B; r=0.29.

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).

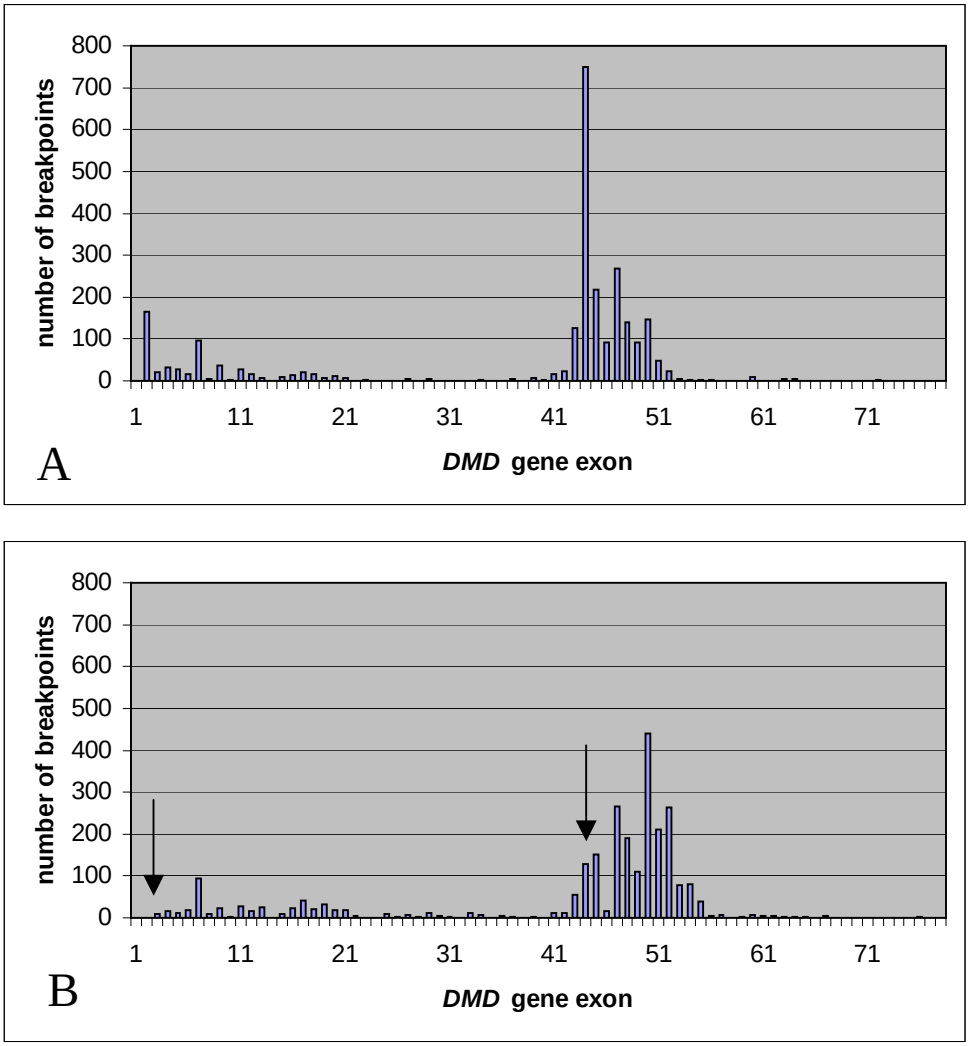

Figure 4. The distinction between (A) 5' and (B) 3' deletion breakpoints. The reduction in 3' breakpoints of the two largest introns (intron 2 and intron 44) is indicated by the arrows.

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.

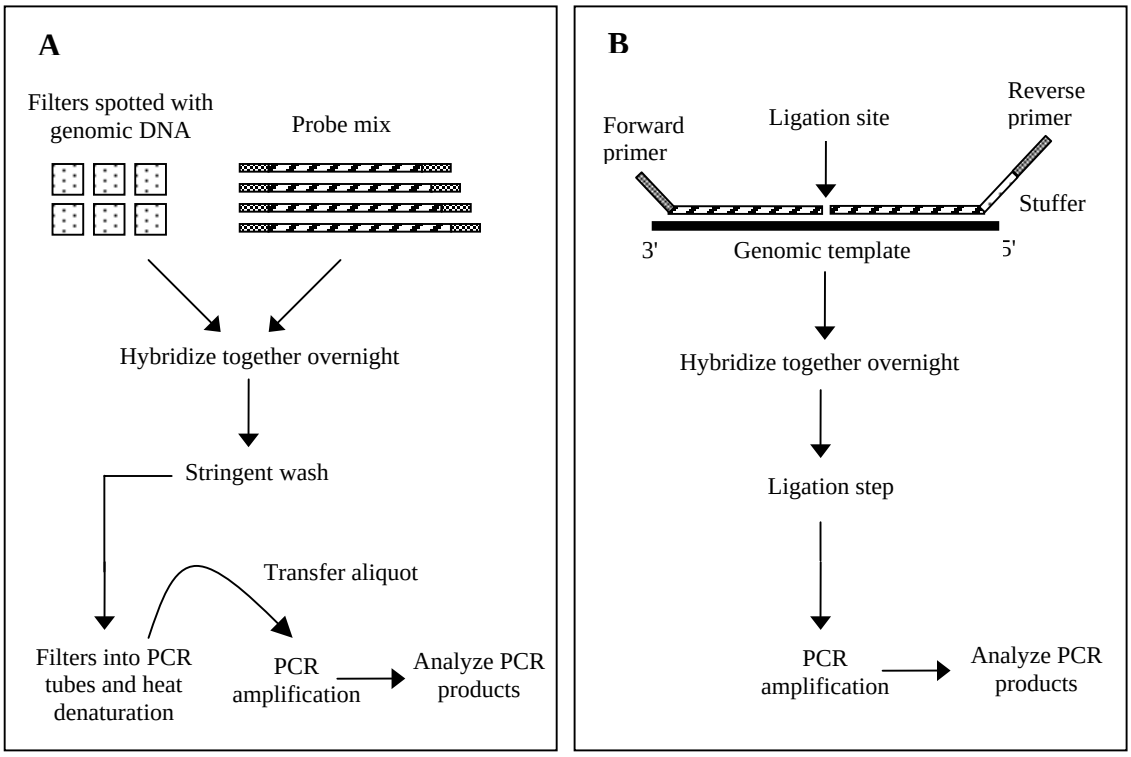

Figure 5. An outline of the two procedures. A. The MAPH technique. A series of probes with identical ends are hybridized to genomic DNA immobilized on a filter. Following stringent washing the specifically bound probes are released into solution, and an aliquot of this is used to seed a PCR reaction. 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. $C_{o}t1$, 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.

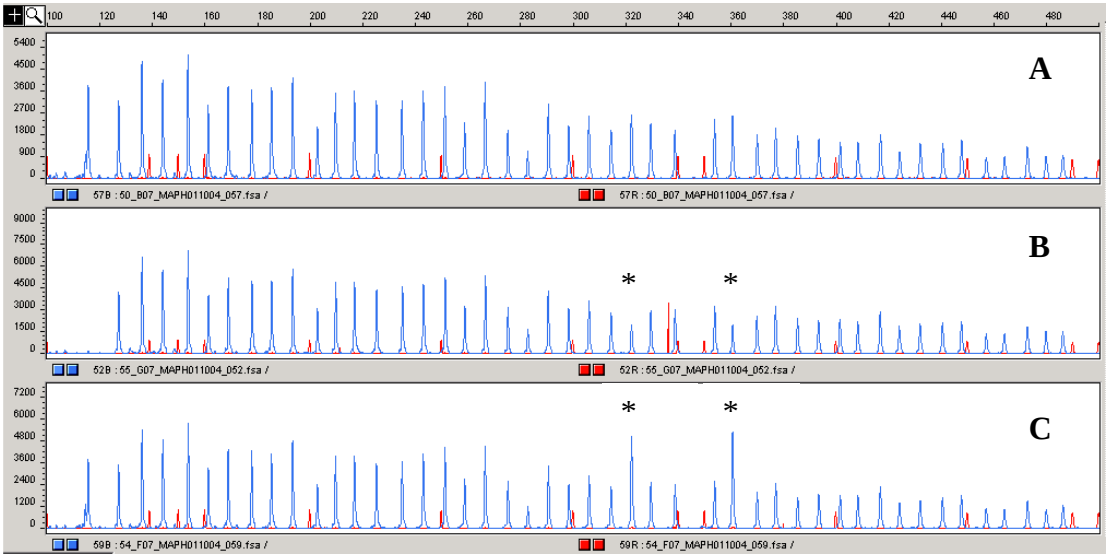

Figure 7. Peaks obtained following separation of one of the MLPA probe sets for the DMD gene on the ABI 3700. Panel A shows a typical pattern from an unaffected male. Panel B is from a female carrier with a deletion of two exons, and panel C is from a male with a duplication of the same two exons. Affected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

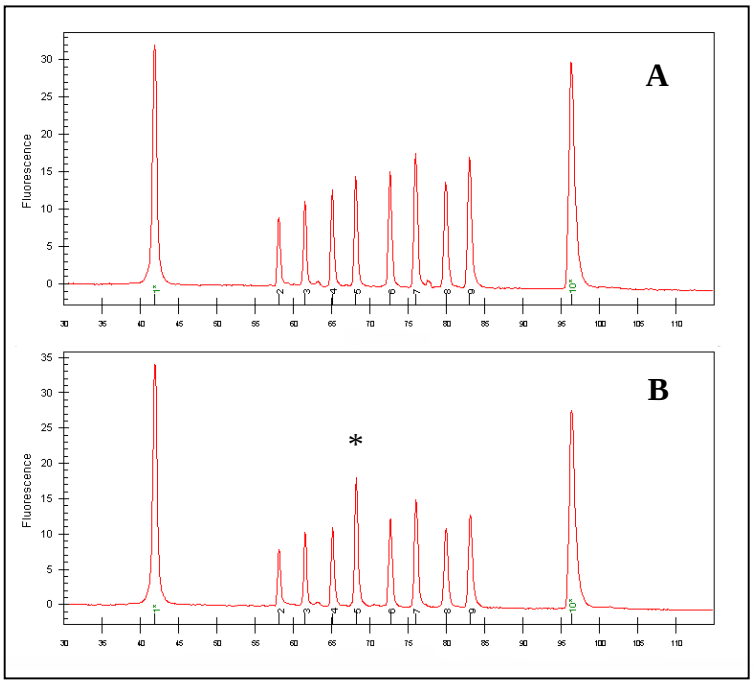

Figure 8. An example of traces obtained by analyzing MAPH products on the Lab-on-a-chip from Agilent. In this example the probes have a spacing of 30-40 base pairs. As can be seen, this spacing could be reduced if necessary to allow more probes to be used. A typical pattern from an unaffected individual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

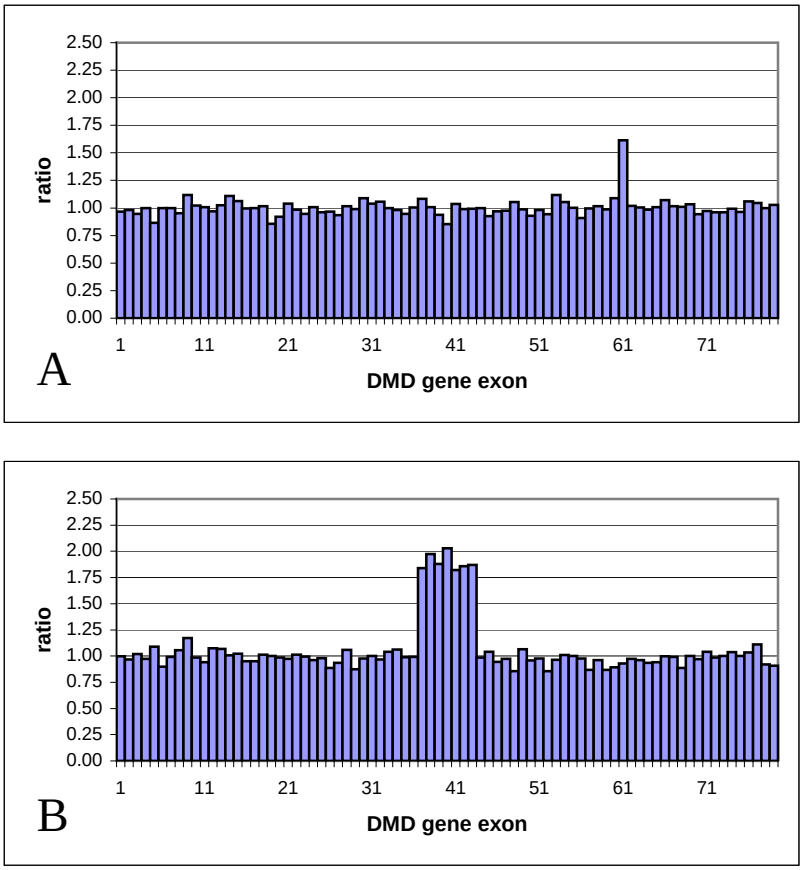

Figure 9. Analysis of MLPA reactions with 2 DMD probe sets following normalization. Under normal circumstances the ratios of all probes clustered around 1.0, indicating an unaffected individual. Graph A shows a single probe with a ratio around 1.5, this is from a female carrier of an exon 61 duplication. 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.

207. White,S.J., Vink,G.R., Kriek,M. et al (2004) Two-color multiplex ligation-dependent probe amplification: Detecting genomic rearrangements in hereditary multiple exostoses. *Hum.Mutat.*, **24**, 86-92.

208. Albertson,D.G., Pinkel,D. (2003) Genomic microarrays in human genetic disease and cancer. *Hum.Mol.Genet.*, **12 Spec No 2**, R145-R152.

209. Snijders,A.M., Nowak,N., Segraves,R. et al (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat.Genet.*, **29**, 263-264.

210. Vissers,L.E., De Vries,B.B., Osoegawa,K. et al (2003) Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am.J.Hum.Genet.*, **73**, 1261-1270.

211. Ishkanian,A.S., Malloff,C.A., Watson,S.K. et al (2004) A tiling resolution DNA microarray with complete coverage of the human genome. *Nat.Genet.*, **36**, 299-303.

212. Locke,D.P., Segraves,R., Nicholls,R.D. et al (2004) BAC microarray analysis of 15q11-q13 rearrangements and the impact of segmental duplications. *J.Med.Genet.*, **41**, 175-182.

213. Yeakley,J.M., Fan,J.B., Doucet,D. et al (2002) Profiling alternative splicing on fiber-optic arrays. *Nat.Biotechnol.*, **20**, 353-358.

214. Oliphant,A., Barker,D.L., Stuelpnagel,J.R. et al (2002) BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *BioTechniques*, **Suppl**, 56-1.

215. Kennedy,G.C., Matsuzaki,H., Dong,S. et al (2003) Large-scale genotyping of complex DNA. *Nat.Biotechnol.*, **21**, 1233-1237.

216. Bignell,G.R., Huang,J., Greshock,J. et al (2004) High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res.*, **14**, 287-295.

217. Lucito,R., Healy,J., Alexander,J. et al (2003) Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res.*, **13**, 2291-2305.

218. Iafrate,A.J., Feuk,L., Rivera,M.N. et al (2004) Detection of large-scale variation in the human genome. *Nat.Genet.*, **36**, 949-951.

219. Sebat,J., Lakshmi,B., Troge,J. et al (2004) Large-scale copy number polymorphism in the human genome. *Science*, **305**, 525-528.

220. Fredman,D., White,S.J., Potter,S. et al (2004) Complex SNP-related sequence variation in segmental genome duplications. *Nat.Genet.*, **36**, 861-866.

Chapter 2.1

White S., Kalf M., Liu Q., Villerius M., Engelsma D., Kriek M., Vollebregt E., Bakker B., van Ommen G.J., Breuning M.H., den Dunnen J.T. (2002). Comprehensive detection of genomic duplications and deletions in the *DMD* gene, by use of multiplex amplifiable probe hybridization. *Am.J.Hum.Genet.* 71 (2):365-374.

Comprehensive Detection of Genomic Duplications and Deletions in the *DMD* Gene, by Use of Multiplex Amplifiable Probe Hybridization

Stefan White,[1] Margot Kalf,[1] Qiang Liu,[2] Michel Villerius,[1] Dieuwke Engelsma,[1] Marjolein Kriek,[1] Ellen Vollebregt,[1] Bert Bakker,[1] Gert-Jan B. van Ommen,[1] Martijn H. Breuning,[1] and Johan T. den Dunnen[1]

[1]Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; and [2]Departments of Molecular Genetics and Molecular Diagnosis, City of Hope National Medical Center, Duarte, CA

Duplications and deletions are known to cause a number of genetic disorders, yet technical difficulties and financial considerations mean that screening for these mutations, especially duplications, is often not performed. We have adapted multiplex amplifiable probe hybridization (MAPH) for the screening of the *DMD* gene, mutations in which cause Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy. MAPH involves the quantitative recovery of specifically designed probes following hybridization to immobilized genomic DNA. We have engineered probes for each of the 79 exons of the *DMD* gene, and we analyzed them by using a 96-capillary sequencer. We screened 24 control individuals, 102 patients, and 23 potential carriers and detected a large number of novel rearrangements, especially small, one- and two-exon duplications. A duplication of exon 2 alone was the most frequently occurring mutation identified. Our analysis indicates that duplications occur in 6% of patients with DMD. The MAPH technique as modified here is simple, quick, and accurate; furthermore, it is based on existing technology (i.e., hybridization, PCR, and electrophoresis) and should not require new equipment. Together, these features should allow easy implementation in routine diagnostic laboratories. Furthermore, the methodology should be applicable to any genetic disease, it should be easily expandable to cover >200 probes, and its characteristics should facilitate high-throughput screening.

Introduction

Most techniques currently applied to reveal disease-causing mutations are PCR based and do not readily produce quantitative data. Consequently, although copy-number changes (i.e., deletions and duplications) are frequently involved, they will go undetected unless specific techniques are applied (Petrij-Bosch et al. 1997; Wijnen et al. 1998; Morgan et al. 1999). The major reason behind this failure is economical: obtaining quantitative data is feasible but is technically demanding, labor intensive, and, thus, costly. When specific precautions are taken, Southern blotting and quantitative PCR are able to detect deletions and/or duplications, but they are both laborious and difficult to implement on a routine basis.

A technique that might fill this gap has recently been described—namely, multiplex amplifiable probe hybridization (MAPH) (fig. 1) (Armour et al. 2000). MAPH is based on the quantitative recovery of probes, after their hybridization to immobilized DNA. Each probe is engineered to allow simultaneous amplification with only one set of primers. This overcomes one of the most difficult elements of quantitative multiplex PCR—namely, differences, between primers, in annealing efficiency. In the original MAPH protocol, after the hybridization step, the PCR products were radioactively labeled during PCR amplification and were analyzed on a polyacrylamide gel. To facilitate a more-automated, higher-throughput application of MAPH, we have chosen to label the products fluorescently and to separate them on a 96-capillary sequencer. Each product is identifiable on the basis of length, with the size of the resultant peak being directly proportional to the copy number of the relevant probe. Changes in peak heights will therefore reflect deletions and duplications in genomic DNA.

We have applied the modified MAPH protocol to scan for copy-number changes in patients with Duchenne muscular dystrophy (DMD) (MIM #310200). DMD is the most commonly inherited neuromuscular disease, affecting 1 in 3,500 male individuals (Worton and Thompson 1988). It is an X-linked disorder that is caused by mutations in the *DMD* gene. This gene is the largest known, covering 2.4 Mb (den Dunnen et al. 1989; Boyce et al. 1991) and containing 79 exons that encode a 14-kb mRNA (Koenig et al. 1987). Translation-truncating

Received March 28, 2002; accepted for publication May 20, 2002; electronically published July 8, 2002.

Address for correspondence and reprints: Dr. J. T. den Dunnen, Human and Clinical Genetics, Leiden University Medical Center, Wassenaarseweg 72, 2333AL Leiden, The Netherlands. E-mail: ddunnen@lumc.nl

0002-9297/2002/7102-0015$15.00

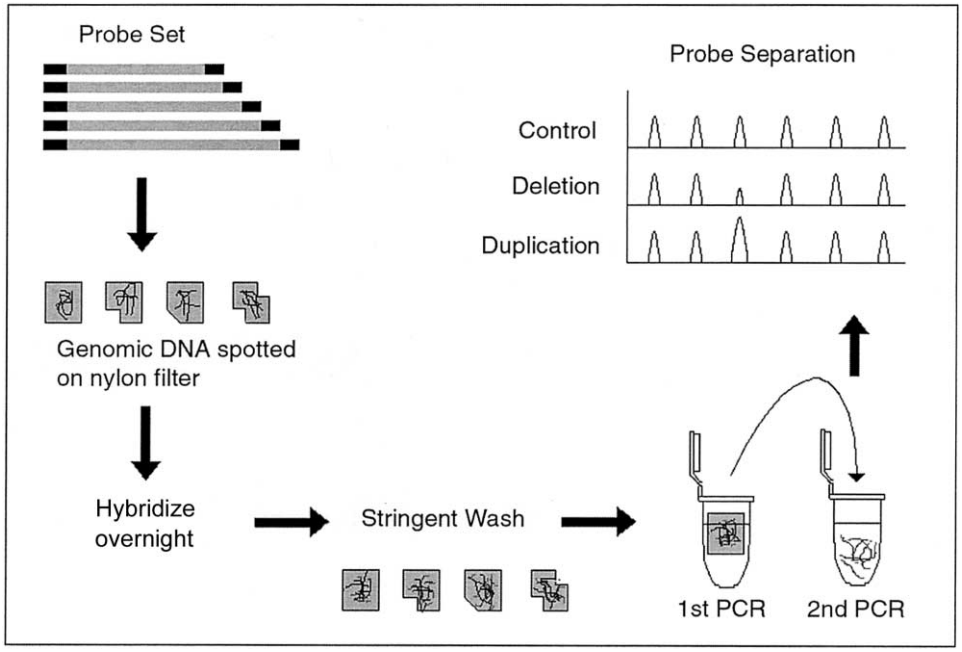

Figure 1 Outline of the MAPH technique. Probes are prepared such that all can be amplified with one primer pair. After overnight hybridization to immobilized genomic DNA, unbound probes are removed by stringent washing. Bound probes are then released and amplified in a quantitative manner. By fluorescent labeling and capillary electrophoresis, it is possible to both discriminate and quantify each probe. Changes in peak heights correspond to copy-number changes (i.e., deletions and duplications).

mutations in *DMD* lead to the lethal phenotype of DMD, whereas mutations that retain the reading frame generally cause the less severe phenotype of Becker muscular dystrophy (BMD) (Monaco et al. 1988; Koenig et al. 1989). An accurate molecular diagnosis is therefore essential both to confirm the clinical diagnosis and to distinguish the two allelic forms.

In approximately two-thirds of cases, the mutation is a deletion or duplication of one or more of the exons, clustered in two hotspot regions (Forrest et al. 1987; Koenig et al. 1987; Darras et al. 1988; den Dunnen et al. 1989; Gillard et al. 1989). In affected male patients, deletion detection is relatively simple. For multiplex PCR, two nine-exon sets (the Chamberlain et al. [1988] and Beggs et al. [1990] sets) have been designed around these hotspots, which together detect 90%–95% of the deletions in male patients. Alternative methods must be applied to determine the exact boundaries of the deletion, as well as to detect duplications and carrier status in female individuals. The size (2.4 Mb) and complexity (79 exons) of *DMD,* however, make this a daunting task. Quantitative Southern blotting has been the most commonly used technique (den Dunnen et al. 1989; Hu et al. 1990; Yamagishi et al. 1996). By the comparison of band intensities between test samples and control samples, it is possible to detect copy-number changes in the individual exons. The preparation of high-quality blots is technically demanding, and six to eight hybridizations are required in order to scan all the exons. In addition, duplication detection is difficult, especially in female carriers and when the duplications are small (i.e., covering only one or two exons). Quantitative PCR is another, more recently applied technique (Ioannou et al. 1992; Mansfield et al. 1993; Yau et al. 1996), in which a multiplex PCR is performed that has a limited number of cycles, ensuring that quantitative products are yielded. Again, the technique is technically demanding, and the incomplete coverage of the exons means that mutations outside the hotspots will be missed. Not surprisingly, therefore, mutation-analysis reports differ considerably in the frequency of duplications detected, ranging between 0 and 6%, depending on the techniques applied for analysis (Koenig et al. 1987; den Dunnen et al. 1989; Hu et al. 1990; Mendell et al. 2001).

Here we describe a MAPH-based method that scans all 79 *DMD* exons for deletions and duplications. We have been able to detect and define a large series of new mutations—in particular, duplications—with several not detected by Southern blotting and/or quantitative PCR. The simplicity of this technique should allow its easy implementation in diagnostic laboratories, and its utility means that it can be readily adapted for the screening of duplications and deletions in any genetic disease.

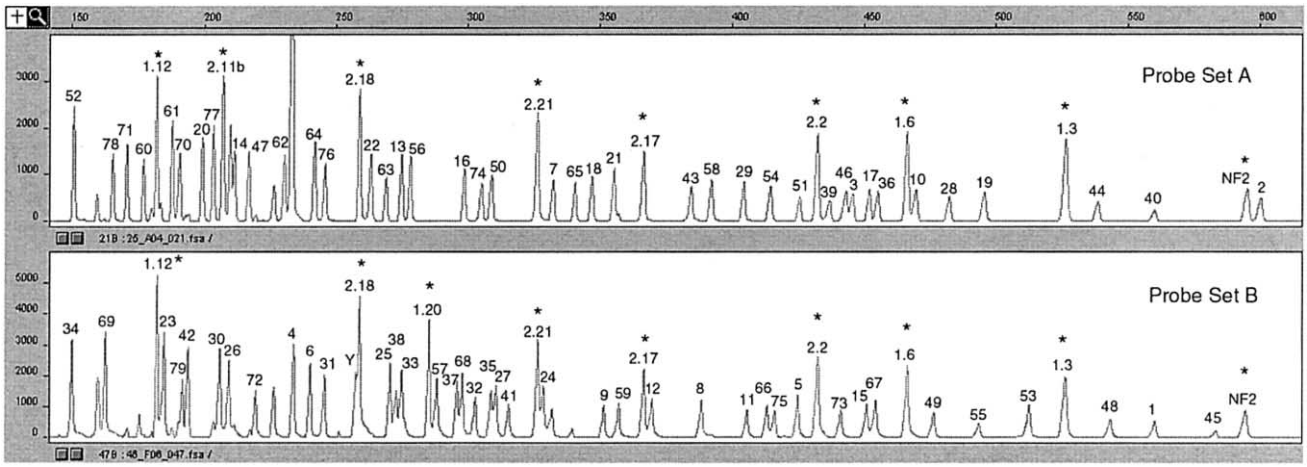

Figure 2 Example of trace patterns obtained from an unaffected male individual. The numbers refer to *DMD* exon numbers: "1.*x*" and "2.*x*" (where *x* is the exon number) refer to BRCA1 and BRCA2, respectively, and "NF2" denotes a probe homologous to the first exon of the *NF2* gene. Asterisks (*) indicate control peaks, and unlabeled peaks indicate noise. Probes range in size from 151 bp (*DMD* exon 34) to 602 bp (*DMD* exon 2).

Methods

Probe Generation

All probes were based on individual exons. Some *DMD* probes were created using primers from the Chamberlain et al. (1988) and Beggs et al. (1990) kits. The remainder were based on sequences obtained from the Leiden Muscular Dystrophy Pages. To facilitate analysis, we prepared control probes by using genomic sequence obtained from GenBank. For each product, the presence of duplicated and/or repetitive sequences was excluded using the BLAST program. The sequences were checked against the nr (nonredundant) and htgs (high-throughput genomic sequences) databases. No probe showed an intraspecies homology >90% for a stretch of ⩾30 nt (expected value $>e^{-11}$).

Products were amplified from genomic DNA by PCR and were cloned into the pGEM-T easy vector (Promega). The correct insert was confirmed by sequencing through use of the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems). This was performed at the Leiden Genome Technology Center, where reactions were analyzed on the ABI 3700 Sequencer (Applied Biosystems).

Each probe was amplified from the vector by use of the primers MAPH-F_1 (GGCCGCGGGAATTCGATT) and MAPH-R_1 (GCCGCGAATTCACTAGTG). Products were purified with the Qiagen PCR cleanup kit (Qiagen) and were then added to the appropriate probe mixture, which had a final concentration of 100–500 pg/μl. Probe sets A and B were prepared containing 40 and 39 *DMD* exons, respectively. Nine-probe control mixtures were made specifically for use with each probe set.

MAPH

MAPH was performed using a protocol adapted from the original MAPH protocol (Armour et al. 2000), as follows (for detailed protocol, see the Leiden Muscular Dystrophy Pages). At least 1 μg genomic DNA was denatured in 1 μl 1N NaOH and spotted on a small nylon filter, followed by UV cross-linking. Up to 16 filters were hybridized together in one tube; the filters were prehybridized in 1 ml prehybridization solution (0.5 M sodium phosphate [pH 7.2], 1 mM EDTA, 7% SDS, and 100 ng/μl Herring Sperm DNA [Gibco BRL]) for ~2 h at 60°C; this solution was replaced with 200 μl prehybridization solution that contained 2 μg denatured Cot_1 DNA (Gibco BRL) and was incubated at 60°C for 30 min. Probe mixture (1 μl combined probes, 1 μl Cot_1 DNA [1 μg/μl], 1 μl Herring Sperm DNA [10 μg/μl; Gibco BRL], 1 μl blocking mixture [blocking primers that each had a final concentration of 20 μM], and 3 μl H_2O) was denatured by the addition of 2 μl 1N NaOH and incubated at 37°C for 1 min. After cooling on ice, 3 μl 1M NaH_2PO_4 was added, and the mixture was added to the tube that contained the filters. Hybridization was performed overnight at 60°C. Washing was performed the next day with five times in 25 ml salt-sodium citrate (SSC) and 1% SDS, followed by five washes in 25 ml 0.1 × SSC and 0.1% SDS, all at 60°C. Each filter was transferred to a PCR tube, and a five-cycle PCR amplification was performed under the following conditions: 94°C for 5 min; five cycles of 94°C for 45 s, 57°C for 1 min, and 68°C for 1 min; and 68°C for 10 min.

Two and one-half microliters of this mixture was transferred to a second PCR, which was performed under the

Table 1

Samples Screened

Sample (Sex)	Method(s)[a]	Mutation Found[b] by: Other Methods	Mutation Found[b] by: MAPH
D1 (M)	1,3	Del 5-7	Del 5-7
D2 (F)			Dup 52-55
D3 (M)	3	Del 50	Del 50
D4 (F)			Dup 50-55
D5 (M)	1,3	nm	Dup 50-55
D6 (M)	3	Del 8-44	Del 8-44
D7 (M)	1,3	nm	nm
D8 (M)	1,3	Dup 43	Dup 43
D9 (F)			Dup 58-63
D10 (M)	1,3	?[c]	Dup 58-63
D11 (M)	3	Del 49-52	Del 49-52
D12 (M)	3	Del 48-50	Del 48-50
D13 (F)			Del 45
D14 (M)	3	Del 1m	Del 1m
D15 (M)	3	nm	Del 53
D16 (M)	1,3	nm	nm
D17 (M)	3	Del (3)-(10)	Del 4-12
D18 (M)	3	Del 14-60	Del 14-60
D19 (M)	1,3	Del 69	Del 64-67
D20 (F)	3	Del (45)-(50)	Del 49-54
D21 (M)	3	Del 2-(33)	Del 2-30
D22 (M)	3	nm	nm
D23 (M)	3	Dup 12	Dup 12
D24 (F)			Dup 2
D25 (M)	2,3	nm	Dup 2
D26 (M)	1,3	nm	nm
D27 (F)			Dup 12-13
D28 (F)			Dup 44
D29 (M)	3	Dup 2-9	Dup 2-9
D30 (F)			nm
D31 (M)	3	Del 1-79	Del 1-79
D32 (F)	3	Dup 44-57	Dup 44-57
D33 (F)			Dup 2-7
D34 (M)	1,3	Dup 2-7	Dup 2-7
D35 (M)	1,3	Dup 2-7	Dup 2-7
D36 (F)	3	Del 10-(?)	Del 10-46
D37 (M)	3	Del 3-(?)	Del 3-19
D38 (F)	3	Del (50)	Del 48-50
D39 (M)			nm
D40 (M)	1,3	Del 46-51	Del 46-51
D41 (M)	3	Del XJ10[d]	Del 4-13
D42 (M)	2,3	Del 3-16	Del 3-16
D43 (M)	1,3	nm	nm
D44 (F)	2,3	Dup 2[e]	Dup 2
D45 (M)			Dup 2
D46 (F)	3	nm	nm
D47 (M)	1,3	nm	nm
D48 (M)	1,3	nm	nm
D49 (M)	3	Del 8-(16)	Del 8-39
D50 (M)	1,3	nm	nm
D51 (F)	2,3	nm	nm
D52 (M)	1,2,3	Dup 51	Dup 51
D53 (M)	1,3	nm	nm
D54 (M)	1,3	nm	nm
D55 (M)	1,3	nm	nm
D56 (M)	1,3	nm	nm
D57 (M)			Dup 6
D58 (M)	1,3	nm	nm
D59 (M)	1,2,3	nm	nm
D60 (M)	1,3	nm	nm
D61 (M)	1,3	Del 2-7	Del 3-6
D62 (M)	1,2,3	Del 20-29	Del 20-29
D63 (F)		Dup 2-(7)	Dup 3-7
D64 (M)		Dup 17	Dup 17
D65 (M)	1,2,3	nm	nm
D66 (F)			nm
D67 (M)	1,2,3	Del 19-43	Del 21-43
D68 (M)	1,2,3	Del 19-43	Del 21-43
D69 (M)	1,3	nm	nm
D70 (F)		Dup 3-7	Dup 3-7
D71 (M)		Dup 3	Dup 3
D72 (F)			Dup 3
D73 (M)	1,2,3	nm	nm
D74 (M)		Dup 51-55	Dup 51-55
D75 (F)	2,3	Dup 51-55	Dup 51-55
G1 (M)	1,4	nm	Del 21
G2 (M)	1,4	nm	Dup 10-11
G3 (M)	1,4	nm	Dup 18-23
G4 (M)	1,4	nm	nm
G5 (M)	1,4	nm	Del 48
G6 (M)	1,4	nm	Dup 6-7
G7 (M)	1,4	nm	Del 66
G8 (M)	1,4	nm	nm
G9 (M)	1,4	nm	nm
G10 (F)	1,4	nm	Del 48-50
G11 (M)	1,4	nm	nm
H1 (M)	1	nm	nm
H2 (M)	1	nm	nm
H3 (M)	1	nm	nm
H4 (F)	1	nm	nm
H5 (M)	1	nm	Del 45-50
H6 (M)	1	nm	Dup 44
H7 (M)	1	nm	nm
H8 (M)	1	nm	nm
H9 (M)	1	nm	nm
H10 (M)	1	nm	Dup 3-4
L1 (M)	1,4	nm	Dup 8-13
L2 (M)	1,4	nm	Del 18
L3 (M)	1,4	nm	nm
L4 (M)	1,4	nm	nm
L5 (M)	1,4	nm	Dup 5-6
L6 (M)	1,4	nm	Dup 54
L7 (M)	1,4	nm	Dup 8-9
L8 (M)	1,4	nm	nm
L9 (M)	1,4	nm	Dup 45-52
L10 (M)	1,4	nm	Dup 2
L11 (M)	1,4	nm	Dup 8-13
L12 (M)	1,4	nm	Dup 53-55
L13 (M)	1,4	nm	Dup 61-64
L14 (M)	1,4	nm	nm
L15 (M)	1,4	nm	Dup 51-57
L16 (M)	1,4	nm	Dup 8-9
L17 (M)	1,4	nm	Dup 2
L18 (M)	1,4	nm	Dup 3-30
L19 (M)	1,4	nm	Dup 20
L20 (M)	1,4	nm	Dup 14-21
L21 (M)	1,4	nm	nm
L22 (M)	1,4	nm	Dup 2
L23 (M)	1,4	nm	Dup 8-9
L24 (M)	1,4	nm	Dup 42-43
L25 (M)	1,4	nm	Dup 6-7
L26 (M)	1,4	nm	Del 56
L27 (M)	1,4	nm	nm
M1 (M)	1,4	nm	Dup 18-32
M2 (M)	1,4	nm	Dup 20-27

(*continued*)

Table 1 (continued)

[a] 1 = PCR by use of the Chamberlain et al. (1988) and Beggs et al. (1990) sets; 2 = quantitative Southern blotting; 3 = quantitative multiplex PCR; 4 = point-mutation detection.

[b] Del = deletion; Dup = duplication; nm = no mutation found. Numbers denote exons; those in parentheses indicate an uncertain breakpoint.

[c] Duplication of 30–50 kb around exon 60 detected using pulsed-field gel electrophoresis.

[d] Probe located in intron 7.

[e] Detectable by quantitative PCR, not evident with quantitative Southern blotting.

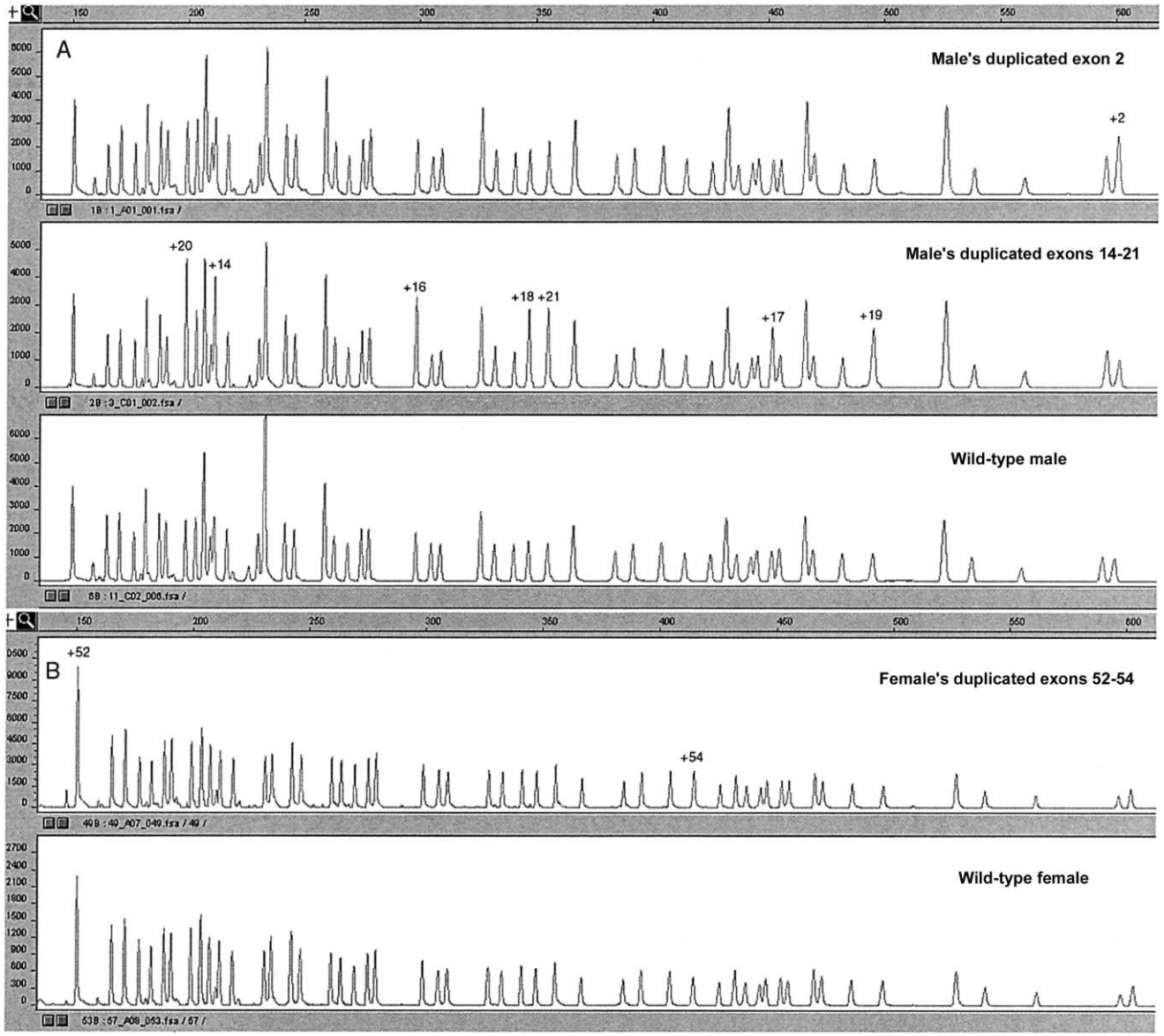

Figure 3 Patterns obtained from analysis of patients by use of probe set A. *A*, Male patient's duplicated exon 2, male patient's duplicated exons 14–21, and male control individual. *B*, Female carrier's duplicated exons 52–54 and female control individual.

same conditions as the first reaction except that one of the primers was fluorescently labeled and the reaction was for 23 cycles. Two microliters of this product was added to 10 μl (Hi Di) Formamide (Applied Biosystems) and 0.15 μl ROX-500 size standard (Applied Biosystems) in a 96-well plate. This was heated at 95°C for 5 min, followed by immediate cooling on ice. The samples were analyzed on an ABI 3700 capillary sequencer (Applied Biosystems).

Data Analysis

Data were analyzed using the programs Gene Scan (Applied Biosystems) and Excel (Microsoft). Peaks were considered unreliable if they were outside predefined thresholds (upper and lower limits of 12,000 and 150 units, respectively).

Male samples were initially visually assessed, to detect any deletions. Presence of a peak that corresponds to a Y chromosome–specific probe confirmed the sex of the sample. Absence of one or more *DMD* peaks was taken to be a deletion, and no calculations were performed.

Samples were analyzed using a combination of methods described for analysis of MAPH (Armour et al. 2000; Sismani et al. 2001) and array-based comparative genomic hybridization (Hodgson et al. 2001) experiments. In the original MAPH publication (Armour et al. 2000),

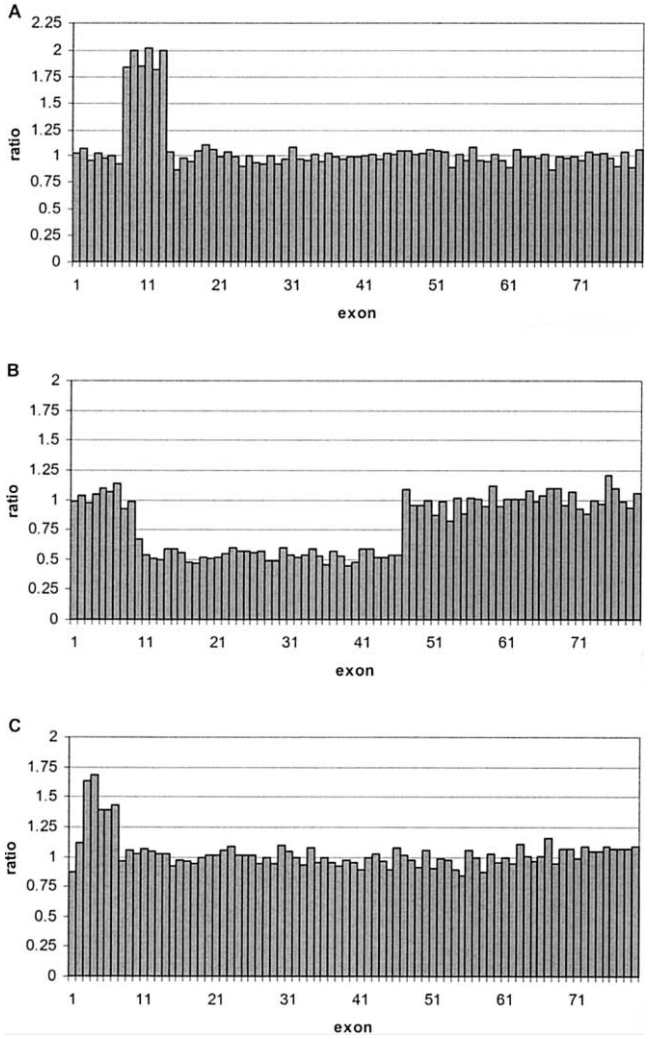

Figure 4 Analysis of different patient samples. *A,* Male patient L1's duplicated exons 8–13, with SD 0.05. *B,* Female carrier D36's deleted exons 10–46, with SD 0.05 *C,* Female carrier D70's duplicated exons 3–7, with SD 0.06.

each peak was compared with the two nearest peaks, for normalization. Since each *DMD* exon could potentially be altered in copy number, we also added probes from exons of autosomal genes unrelated to *DMD*. For each dystrophin exon, the peak height was divided by the sum of the peak heights of the two nearest unlinked probes, to give a ratio. Within one hybridization, the median of the ratio for each exon was calculated and was used as a reference value against which all exons were compared. Each exon was divided by this number, thereby normalizing all unaffected exons to 1.0. For each sample, initial estimates for deletions or duplications were performed visually, by setting arbitrary thresholds on the basis of expected ratios (Hodgson et al. 2001). Wild-type exons were expected to fall in the ranges of 1.0 ± 0.5 for male patients and 1.0 ± 0.25 for female carriers. The median and SD of the exons that fell within this range were calculated, and each exon was divided by the median to correct for variations between samples. Any exon that was outside 3 SDs of the "normal" exons was assumed to be altered in copy number. Samples that showed an SD >15% over the unaffected exons or that appeared to show noncontiguous deletions or duplications were deemed to be unreliable.

Results

Probes were initially tested by hybridization to control DNA from 24 healthy individuals, as well as to DNA from a patient with a deletion encompassing the entire *DMD* gene. From the control samples, all the probes could be recovered (fig. 2), whereas only control probes from outside *DMD* were recovered from the patient sample (data not shown). Thus, none of the probes hybridize to other regions in the genome, which would lead to false-positive signals.

A total of 125 samples were screened in a semiblind manner (table 1 and figs. 3 and 4). These were a mixture of fully and partially characterized cases, as well as samples from cases in which no mutation had been found. In several cases, the DNA was from a potential carrier in whom the mutation sought was already known. With a threshold of 3 SDs and the assumption that the unaffected ratios are normally distributed, a false-positive result should only occur ~0.3% of the time, which is the equivalent of one exon per four *DMD* genes tested. This is approximately the ratio of false-positive results seen among samples that were not excluded for other reasons (e.g., peak height being outside the boundaries or SD being >15%). Therefore, all samples that showed a single-exon rearrangement were tested at least twice. Following these criteria, we found no sample that showed evidence of more than one mutational event.

False-negative results are more difficult to assess. An estimate can be made by looking at patient samples in which more than two exons are deleted or duplicated. A result would be considered to be false negative when one or more exons within a mutated series was found to be normal. In no patient sample was this seen. Although this does not exclude the possibility that false-negative calls will occur, it does suggest that they will happen very rarely.

The exon 75 probe was the probe that showed the highest variation among the 79 *DMD* probes used. This appeared to be due to slight variations in PCR/washing conditions, rather than variations in a polymorphic sequence, since no sample consistently showed a duplication of exon 75. For any hybridization in which exon 75 showed such variation, the results for that exon were ignored, and the exon was retested in a subsequent ex-

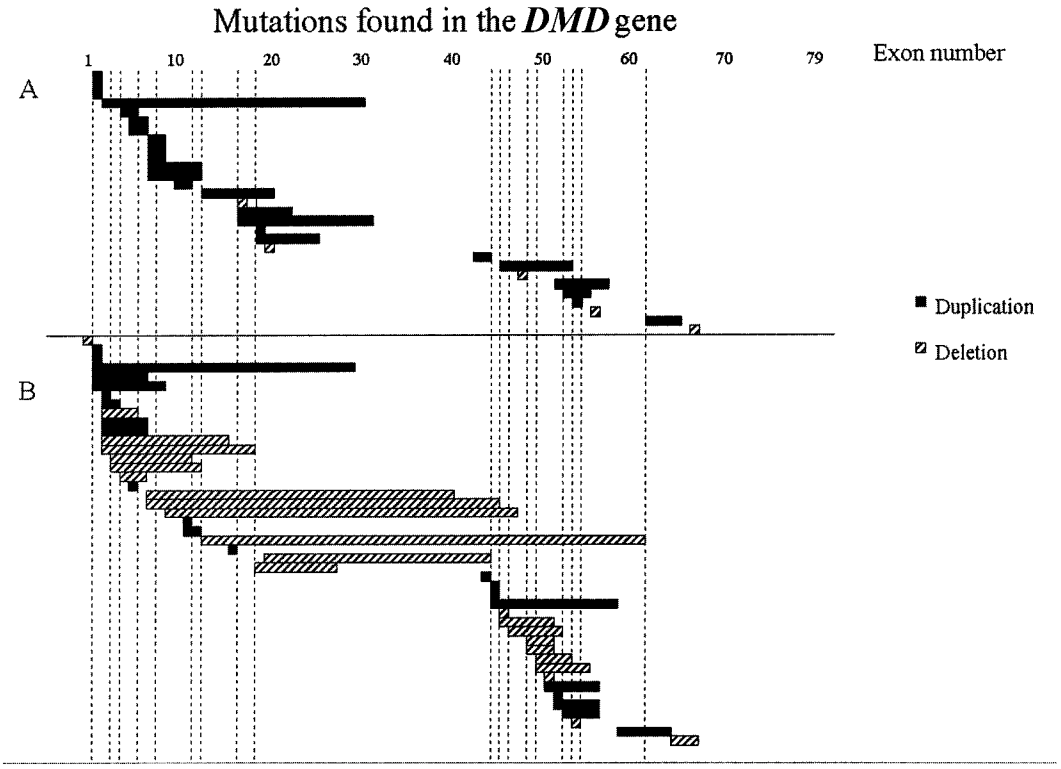

Figure 5 Independent mutations detected during the present study. Vertical bars represent the 18 exons tested using the Chamberlain et al. (1988) and Beggs et al. (1990) kits. *A*, Mutations detected in samples in which point mutations and deletions had been excluded, mainly by multiplex PCR. *B*, All other mutations detected.

periment. In no cases could an exon 75 duplication be confirmed.

Initially, some probes could not be recovered. Close examination of the sequences revealed that all had a relatively low GC content (<40%). In some cases—for example, exon 2—it was not possible to raise this percentage, since the entire region was extremely AT rich. To solve this problem, we made the probes longer. In this manner, we were able to use a 602-bp exon 2 probe with a GC content as low as 30%.

Discussion

Of the 24 mutations previously characterized in our laboratory, all were detected using the MAPH technique (table 1). In one case, the breakpoints did not match exactly. This was in a male patient (D61) that, with MAPH, was seen as having a deletion of exons 3–6 but had previously been diagnosed as having a deletion of exons 3–7. Southern blot analysis showed a junction fragment for exon 7, suggesting that the breakpoint may be within the exon. In a hybridization-based technique, a breakpoint may be misdiagnosed if the deletion occurs within the sequence to which it is bound by the probe and if there is enough of the exon remaining for the probe to hybridize; this is likely to occur rarely however, since it has been calculated that, in ~99% of cases of DMD and/or BMD, the breakpoints are outside the coding exon (den Dunnen et al. 1989). By contrast, PCR from genomic DNA may lead to false-negative results if there is a polymorphism within the priming site (Abbs et al. 1991).

Of the 72 male samples in which no mutation had previously been found, 37 (51%) had mutations that were detectable by use of MAPH (table 1). When only those samples that had been checked for deletions and point mutations were included, the frequency was 74% (29/39). These were composed of five deletions (all of one exon) and, strikingly, but not unexpectedly, 24 duplications. To present an unbiased view of duplication distribution, we have depicted those mutations detected in samples that were also screened for point mutations and deletions (fig. 5*A*) separately from an overview of all the mutations detected (fig. 5*B*).

Samples LA1–LA27 (table 1) were from a point-mutation screening by use of the DOVAM-S (detection of virtually all mutations–SSCP) technique (Mendell et al. 2001). A total of 141 samples were tested, and 108 point

mutations were found. We analyzed 27 of the remaining 33 samples, finding two deletions and 20 duplications. Samples M1 and M2 (table 1) remained from a study, using denaturing high-performance liquid chromatography, that screened eight patients with DMD for point mutations (Bennett et al. 2001) but found no obvious pathological mutations in two of these samples. By use of MAPH, both samples showed a duplication, thereby completing the mutation study.

A total of 23 female potential or proven carriers were tested. Of the 15 samples in which no mutation had previously been found, two deletions and eight duplications were found. Analysis of potential carriers was facilitated when it was known what mutation to expect.

Newly found mutations could often be confirmed using other methods. Small duplications, such as that in sample D52 (with an exon 51 duplication), could be confirmed by retrospective examination of Southern blots that had been previously prepared and analyzed in our laboratory. The DNA in sample D45 showed an exon 2 duplication by use of MAPH, as did DNA from the mother (sample D44). The result from sample D44 was confirmed by quantitative PCR, yet was not evident on a Southern blot.

The exon 2 (samples D25 and D120.7) and exons 58–63 (samples D10 and DL33.2) duplications, which have been described elsewhere (den Dunnen et al. 1989), are interesting cases. Pulsed-field gel electrophoresis analysis indicated that there were rearrangements of ~150 kb, at the 5′ end of the gene, and 30–50 kb, around exon 60. Despite a focused analysis of both regions, no duplications could be detected using Southern blotting.

Duplication of exon 2 alone is extremely difficult to detect by Southern blotting, since the band is very weak. This may be due to the very low GC content (~30%) of exon 2 and its surrounding region, leading, under stringent conditions, to weak hybridization. Given the extremely large size (190 and 170 kb) of the introns flanking exon 2, it is not surprising that a deletion or duplication of exon 2 by itself is a mutation that has been found more than once. In fact, it was the single most common duplication found, occurring five times. Interestingly, however, no deletion of exon 2 alone has so far been reported (Leiden Muscular Dystrophy Pages).

Our results show that, even when the *DMD* gene is screened for deletions, duplications, and point mutations (DOVAM-S or denaturing gradient gel electrophoresis), a small number of samples remain in which no disease-causing mutation can be detected. There are several possible explanations why no mutation was found in these samples. When RNA has not yet been analyzed in a patient, mutations that affect splicing are the most plausible candidates. Indeed, RNA-based techniques, such as the protein-truncation test, detect mutations that would be missed using DNA-based techniques (Roest et al. 1996; Whittock et al. 1997). It is also possible that the disease was misdiagnosed and that the mutation lies in a gene responsible for other muscular disorders. Germline mosaicism has been reported elsewhere (Bakker et al. 1987; Wood and McGillivray 1988) and would not necessarily be detectable by use of the methods described herein. Another, less likely reason is mutations in a gene that is involved in the regulation of dystrophin expression.

Although mutation detection obviously is critical for diagnosis, it may also be important for future therapeutic purposes. Recent reports have showed the potential use of read-through protein synthesis (Gentamycin) (Barton-Davis et al. 1999) and exon skipping (with antisense oligoribonucleotides) (van Deutekom et al. 2001) in the restoration of the reading frame of the dystrophin transcript. In particular, single-exon duplications, as detected in 12 cases in this study, would make an ideal target for exon skipping. The presence of two targets not only would double the efficiency but also should produce a normal transcript, leading to a wild-type protein.

The MAPH approach's primary advantages over Southern blotting and quantitative PCR are the relative simplicity, speed, and completeness of coverage of all 79 exons. Although 90%–95% of the deletions can be detected using multiplex PCR, the breakpoints are often not determined, and rare mutations outside the hotspots will be missed. In previously published reports on MAPH (Armour et al. 2000; Sismani et al. 2001), recovered probes were radioactively labeled and were separated on a polyacrylamide gel. For speed and convenience, we chose to use a combination of fluorescent labeling and capillary electrophoresis. Capillary electrophoresis is becoming more widely used in mutation detection, since it provides greater sensitivity and has high-throughput capabilities (Bosserhoff et al. 2000). We used the ABI 3700 (Applied Biosystems), which allows the simultaneous analysis of 96 samples. One run of 96 samples takes ~4 h, with the data analyzed by software provided with the machine.

There are several ways in which the current system can be further enhanced. In the present study, only two (blue [FAM sample] and red [ROX size standard]) of the four available colors were used. By use of up to three sets of primers, each labeled with a different fluorophore, it should be possible to expand the potential number of probes by threefold. Hybridizing the PCR products to a microarray composed of each individual probe could further increase the number of probes tested, with the additional advantage that they would no longer need to be differentiated in length.

In contrast to many other methods, this technique should be easy to implement in a standard diagnostic laboratory, since no new technology needs to be introduced. The critical techniques are hybridization and

PCR, and the products can be analyzed on any apparatus that is used for sequence analysis. Furthermore, it can easily be applied to any disease gene of interest, and the resolution provided and the potential of array implementation may even allow future genomewide screening.

Acknowledgments

We would like to thank R. Hofstra, E. Hoffman, S. Sommer, and R. Bennett for providing DNA samples. This work was supported by grants from ZorgOnderzoek Nederland (9607.031.1) and the Prinses Beatrix Fonds.

Electronic-Database Information

The accession number and URLs for data presented herein are as follows:

BLAST, http://www.ncbi.nlm.nih.gov/BLAST/ (for nr and htgs databases)
GenBank, http://www.ncbi.nlm.nih.gov/Genbank/
Leiden Muscular Dystrophy Pages, http://www.dmd.nl/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for DMD [MIM #310200])

References

Abbs SJ, Yau SC, Mathew CG, Bobrow M (1991) A convenient multiplex PCR system for the detection of dystrophin gene deletions: a comparative analysis with cDNA hybridisation shows mistypings by both methods. J Med Genet 28:304–311

Armour JA, Sismani C, Patsalis PC, Cross G (2000) Measurement of locus copy number by hybridisation with amplifiable probes. Nucleic Acids Res 28:605–609

Bakker E, Van Broeckhoven C, Bonten EJ, Van De Vooren MJ, Veenema H, Van Hul W, van Ommen GJB, Vandenberghe A, Pearson PL (1987) Germline mosaicism and Duchenne muscular dystrophy mutations. Nature 329:554–556

Barton-Davis ER, Cordier L, Shoturma DI, Leland SE, Sweeney HL (1999) Aminoglycoside antibiotics restore dystrophin function to skeletal muscles of *mdx* mice. J Clin Invest 104:375–381

Beggs AH, Koenig M, Boyce FM, Kunkel LM (1990) Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction. Hum Genet 86:45–48

Bennett RR, Dunnen J, O'Brien KF, Darras BT, Kunkel LM (2001) Detection of mutations in the dystrophin gene via automated DHPLC screening and direct sequencing. BMC Genet 2:17

Bosserhoff AK, Buettner R, Hellerbrand C (2000) Use of capillary electrophoresis for high throughput screening in biomedical applications: a minireview. Comb Chem High Throughput Screen 3:455–466

Boyce FM, Beggs AH, Feener CA, Kunkel LM (1991) Dystrophin is transcribed in brain from a distant upstream promoter. Proc Natl Acad Sci USA 88:1276–1280

Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT (1988) Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Res 16:11141–11156

Darras BT, Blattner P, Harper JF, Spiro AJ, Alter S, Francke U (1988) Intragenic deletions in 21 Duchenne muscular dystrophy (DMD)/Becker muscular dystrophy (BMD) families studied with the dystrophin cDNA: location of breakpoints on *Hin*dIII and *Bgl*II exon-containing fragment maps, meiotic and mitotic origin of the mutations. Am J Hum Genet 43:620–629

den Dunnen JT, Grootscholten PM, Bakker E, Blonden LAJ, Ginjaar HB, Wapenaar MC, van Paassen HMB, van Broeckhoven C, Pearson PL, van Ommen GJB (1989) Topography of the Duchenne muscular dystrophy (DMD) gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. Am J Hum Genet 45:835–847

Forrest SM, Cross GS, Speer A, Gardner-Medwin D, Burn J, Davies KE (1987) Preferential deletion of exons in Duchenne and Becker muscular dystrophies. Nature 329:638–640

Gillard EF, Chamberlain JS, Murphy EG, Duff C, Smith B, Burghes AHM, Thompson MW, Sutherland J, Oss I, Bodrug SE, Klamut HJ, Ray PN, Worton RG (1989) Molecular and phenotypic analysis of patients with deletions within the deletion-rich region of the Duchenne muscular dystrophy (DMD) gene. Am J Hum Genet 45:507–520

Hodgson G, Hager JH, Volik S, Hariono S, Wernick M, Moore D, Albertson DG, Pinkel D, Collins C, Hanahan D, Gray JW (2001) Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. Nat Genet 29:459–464

Hu X, Ray PN, Murphy E, Thompson MW, Worton RG (1990) Duplicational mutation at the Duchenne muscular dystrophy locus: its frequency, distribution, origin and phenotype/genotype correlation. Am J Hum Genet 46:682–695

Ioannou P, Christopoulos G, Panayides K, Kleanthous M, Middleton L (1992) Detection of Duchenne and Becker muscular dystrophy carriers by quantitative multiplex polymerase chain reaction analysis. Neurology 42:1783–1790

Koenig M, Beggs AH, Moyer M, Scherpf S, Heindrich K, Bettecken T, Meng G, et al (1989) The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. Am J Hum Genet 45:498–506

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener CA, Kunkel LM (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

Mansfield ES, Robertson JM, Lebo RV, Lucero MY, Mayrand PE, Rappaport E, Parrella T, Sartore M, Surrey S, Fortina P (1993) Duchenne/Becker muscular dystrophy carrier detection using quantitative PCR and fluorescence-based strategies. Am J Med Genet 48:200–208

Mendell JR, Buzin CH, Feng J, Yan J, Serrano C, Sangani DS, Wall C, Prior TW, Sommer SS (2001) Diagnosis of Duchenne dystrophy by enhanced detection of small mutations. Neurology 57:645–650

Monaco AP, Bertelson CJ, Liechti-Gallati S, Moser H, Kunkel LM (1988) An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. Genomics 2:90–95

Morgan NV, Tipping AJ, Joenje H, Mathew CG (1999) High frequency of large intragenic deletions in the Fanconi anemia group A gene. Am J Hum Genet 65:1330–1341

Petrij-Bosch A, Peelen T, van Vliet M, van Eijk R, Olmer R, Drusedau M, Hogervorst FBL, Hageman S, Arts PJ, Ligtenberg MJ, Meijers-Heijboer H, Klijn JG, Vasen HF, Cornelisse CJ, van 't Veer LJ, Bakker E, van Ommen GJB, Devilee P (1997) BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. Nat Genet 17:341–345

Roest PAM, Bout M, Van Der Tuijn AC, Ginjaar HB, Bakker E, Hogervorst FBL, van Ommen GJB, den Dunnen JT (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

Sismani C, Armour JA, Flint J, Girgalli C, Regan R, Patsalis PC (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

van Deutekom JC, Bremmer-Bout M, Janson AAM, Ginjaar HB, Baas F, den Dunnen JT, van Ommen GJB (2001) Antisense-induced exon skipping restores dystrophin expression in DMD patient derived muscle cells. Hum Mol Genet 10: 1547–1554

Whittock NV, Roberts RG, Mathew CG, Abbs SJ (1997) Dystrophin point mutation screening using a multiplexed protein truncation test. Genet Test 1:115–123

Wijnen J, van der Klift H, Vasen H, Khan PM, Menko F, Tops C, Meijers HH, Lindhout D, Moller P, Fodde R (1998) *MSH2* genomic deletions are a frequent cause of HNPCC. Nat Genet 20:326–328

Wood S, McGillivray BC (1988) Germinal mosaicism in Duchenne muscular dystrophy. Hum Genet 78:282–284

Worton RG, Thompson MW (1988) Genetics of Duchenne muscular dystrophy. Annu Rev Genet 22:601–629

Yamagishi H, Kato S, Hiraishi Y, Ishihara T, Hata J, Matsuo N, Takano T (1996) Identification of carriers of Duchenne/Becker muscular dystrophy by a novel method based on detection of junction fragments in the dystrophin gene. J Med Genet 33:1027–1031

Yau SC, Bobrow M, Mathew CG, Abbs SJ (1996) Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis. J Med Genet 33:550–558

Chapter 2.2

White S.J., Aartsma-Rus A., Flanigan K., Lalic T., Janson A.A.M., Breuning M.H., den Dunnen J.T. Duplications in the *DMD* gene are non-randomly distributed and can be complex rearrangements. Manuscript in preparation.

Duplications in the *DMD* gene are non-randomly distributed and can be complex rearrangements

White S.J., Aartsma-Rus A., Flanigan K., Lalic, T., Janson, A.A.M., Breuning, M.H., den Dunnen J.T.

Abstract

Using Multiplex Amplifiable Probe Hybridization (MAPH) and Multiplex Ligation-dependent Probe Amplification (MLPA) we have screened different cohorts of Duchenne/Becker Muscular Dystrophy (DMD/BMD) patients for duplications. In an unselected series the duplication frequency was 8%; in a group of patients already screened for deletions and point mutations we found a duplication in 64% of cases. The majority were simple, contiguous duplications, however we detected 4 non-contiguous duplications, with two also including a triplication. In two instances the 3' end of the gene was affected, a region not usually screened by multiplex PCR. These mutations would therefore go undetected, whilst potentially disturbing the reading frame of the mRNA. This emphasizes the importance of screening the entire gene for rearrangements.

More than 50% of the duplications found were at the 5` end of the gene, whereas most deletions are found in the middle of the gene. A more detailed comparison of the regions affected showed that a duplication of exon 2 only was the single most common duplication found. Analysis of the breakpoints of 11 such cases revealed two recombination hotspots within intron 2, whereas the breakpoints within intron 1 were scattered. We propose that unequal crossing over between sister chromatids is not responsible for exon 2 duplications. Instead, a mechanism such as synthesis-dependent non-homologous end joining may be responsible. Assuming this also applies to other duplications within the gene, this may explain the different distributions seen for deletions and duplications.

Introduction

The Duchenne Muscular Dystrophy (*DMD*) gene is the largest known, spanning ~2.4 Mb of genomic sequence on Xp21 (Den Dunnen et al., 1992; Mandel, 1989). Mutations in the gene cause DMD, the most commonly inherited neuromuscular disorder, and Becker Muscular Dystrophy (BMD), the milder allelic form of the disease. The mutation spectrum within the gene is unusual in that deletions of one or more exons are found in ~65% of cases (Den Dunnen et al., 1989). These deletions are known to cluster in hot spot regions (Oudet et al., 1992; Forrest et al., 1987), and can easily be detected in males using simple PCR reactions. Two multiplex PCR kits of 9 exons each were developed (Beggs et al., 1990; Chamberlain et al., 1988), and together detect ~98% of all deletions. Duplication analysis and the determination of carrier status, however, require a quantitative method of analysis. Until recently the most commonly applied have been Southern blotting (Forrest et al., 1987) and quantitative multiplex PCR (Ioannou et al., 1992; Abbs and Bobrow, 1992). Southern blotting can cover all exons of the gene, but requires 6-8 blots to do this. This makes it time consuming, and it can be difficult to get blots of the necessary quality. Quantitative multiplex PCR has proven difficult to apply, and will miss mutations outside the hotspot regions. In addition, it implicitly assumes that the majority of duplications will be found within the deletion hotspots.

We recently reported the use of Multiplex Amplifiable Probe Hybridization (MAPH) for screening all 79 exons of the *DMD* gene for deletions and duplications in 2 reactions. Using this approach we were able to identify several small deletions and duplications not detected with other techniques, and showed that a duplication of exon 2, previously undescribed, was the single most common duplication found (White et al., 2002).

A similar technique, Multiplex Ligation-dependent Probe Amplification (MLPA) (Schouten et al., 2002), has become widely used for the detection of deletions and duplications in a variety of diseases (Taylor et al., 2003; Rooms et al., 2004; Slater et al., 2003). In comparison to MAPH, removal of any unbound probes is unnecessary, making the approach easier to perform.

In this report we have used both MAPH and MLPA to screen for duplications in different series of DMD/BMD patients.

Materials and Methods

Patients

All patients were diagnosed with either DMD or BMD by a medical specialist. DNA samples were obtained from blood samples using standard procedures.

Mutation Detection

All samples were screened for deletions and duplications with either MLPA or MAPH. The MLPA reaction was performed with the P034 and P035 kits from MRC-Holland (Amsterdam, the Netherlands). These sets contain probes for all 79 exons of the *DMD* gene as well as the cortical promoter (Dp427c). The reaction was performed following the protocol described in Schouten et al 2002. After the PCR amplification (33-35 cycles) 2 µl of product was mixed with 10 µl Hi Di formamide and 0.1 µl ROX 500 size standard, and separated on the ABI 3700 capillary sequencer (Applied Biosystems). MAPH was performed as described in White et al 2002. Peak data was derived with GeneScan and exported into Excel for analysis. Both MAPH and MLPA analysis was performed as described for MAPH in White et al 2002.

Determination of exon 2 duplication breakpoints.

MLPA probes were designed in intron 1 and intron 2, based on criteria outlined in White et al 2004. Probes within intron 1 ended with sequences allowing amplification with the MAPH primers; the intron 2 probes used the MLPA amplification sequences. All probes were combined in a single mix, with the reaction and analysis being performed as described previously (White et al., 2004).

Based on the estimated duplication borders primers for PCR amplification were designed using the Primer 3 program, with the forward primers in intron 2 and the reverse primers in intron 1. Long range PCR was performed using the Expand Long Template PCR system (Roche), and the resulting PCR products were separated on a 0.8% agarose gel by electrophoresis. Bands of interest were excised and purified using the QIAquick PCR purification kit (Qiagen), and sequenced by the Leiden Genome Technology Center (www.LGTC.nl).

MyoD infection, RNA isolation and RT-PCR analysis

Amniocytes from sample LM7 were forced into myogenesis by infection with a MyoD-containing adenovector as described previously (Havenga et al., 2002; Aartsma-Rus et al., 2003) . RNA isolation and RT-PCR analysis were performed as described (Aartsma-Rus et al., 2002). The full list of primers used is available on request.

Results

Different patient cohorts were screened for deletions and duplications in the *DMD* gene using either MLPA or MAPH. In the first series 120 patients had been screened for deletions within the hot spot regions using modified multiplex PCR kits. Rescreening these revealed 9 duplications, as well as several smaller deletions falling outside the hot spots (Lalic et al, manuscript in preparation). One of the duplications was a complex rearrangement, consisting of 2 distinct duplications and a triplication (Figure 1).

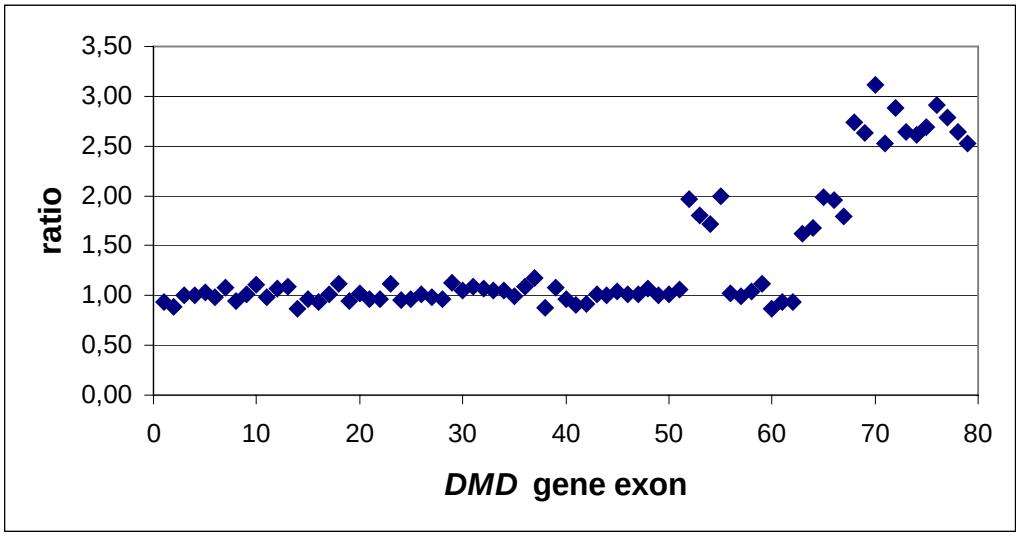

Figure 1. Graphical results based on MLPA analysis of a non-contiguous duplication/triplication.

The second series consisted of 50 patients that had been screened for point mutations and deletions. In this cohort we found 32 (64%) duplications. Non-contiguous rearrangements were seen in 3 samples, and a duplication of exon 2 was the single most common duplication seen, in this study making up 28% of all duplications found. The distribution of the remainder of the mutations was relatively even throughout the gene, but overall the mutations tended to cluster more in the 5' end of the gene. The duplications found in both study populations are listed in table 1.

To see if the large number of exon 2 duplications was due to specific rearrangement hotspots we attempted to define the breakpoints in 11 samples, 10 from these studies and one identified previously (White et al., 2002). MLPA analysis with probes initially spaced ~20 kb apart throughout introns 1 and 2 showed that the breakpoints in intron 1 were scattered, whereas 10 of the 11 breakpoints in intron 2 were found in the first 40 kb. The largest duplication was maximally 220 kb, whereas the smallest duplication was at least 40 kb.

Sample	Mutation
D1	dup 02
D2	dup 02
D3	dup 02
D4	dup 02
D5	dup 02
D6	dup 02
D7	dup 02
D8	dup 02
D9	dup 02
D10	dup 02
D11	dup 02-11
D12	dup 03-04
D13	dup 03-11
D14	dup 03-44
D15	dup 03-06
D16	dup 5-18; trip 19-41; dup 42; trip 43-44
D17	dup 5-19; dup 38-41
D18	dup 06-07
D19	dup 08-13
D20	dup 08-13
D21	dup 08-29
D22	dup 08-44
D23	dup 12-30
D24	dup 17-18
D25	dup 22-25
D26	dup 29-43
D27	dup 37-43
D28	dup 43
D29	dup 45-55; 65-79
D30	dup 45-65
D31	dup 46-47
D32	dup 46-60
D33	dup 49-60
D34	dup 50-59
D35	dup 51
D36	dup 52-55; dup 63-67; trip 68-79
D37	dup 56-63
D38	dup 61
D39	dup 61-62
D40	dup 61-63
D41	dup 61-63
LM7	dup 52-62

Table 1. The duplications found in this study. The numbers refer to exons of the *DMD* gene; non-contiguous rearrangements are separated with a semi colon.

Given the apparent clustering of breakpoints within intron 2, more MLPA probes were designed within the first 40 kb of this intron. Retesting the 10 samples with apparently clustered intron 2 breakpoints showed that five of the samples had the intron 2 breakpoint within a 7 kb area, with the other five breakpoints clustering in a separate region of ~4 kb (Figure 2).

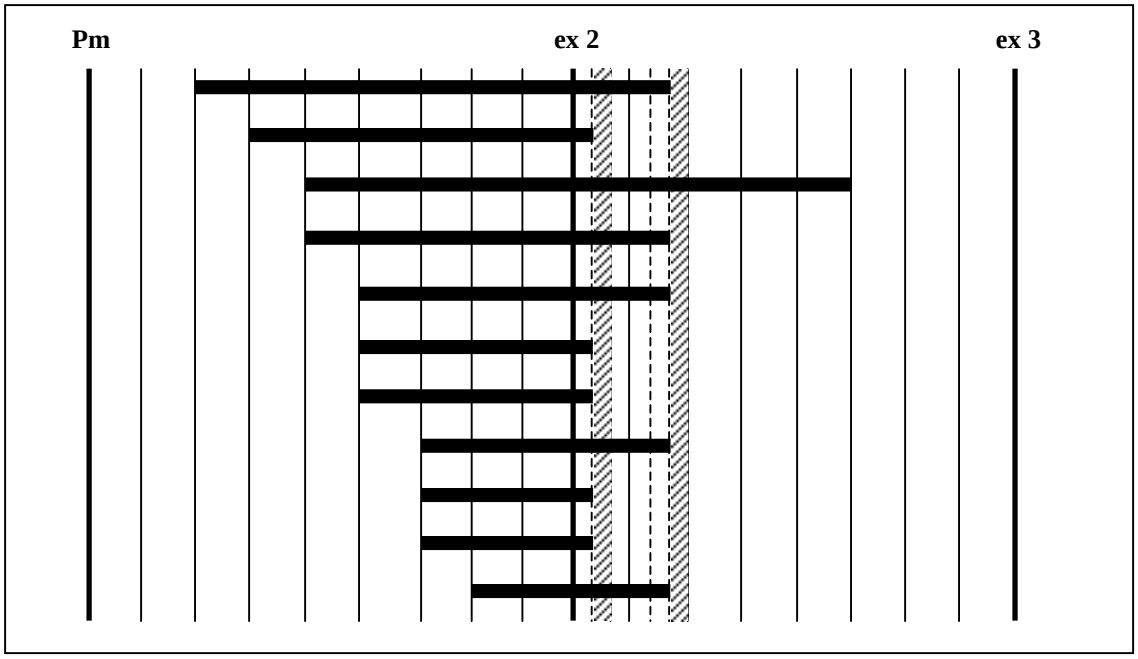

Figure 2. The minimum extent of the different exon 2 duplications, as determined by MLPA and long range PCR. The vertical bars indicate the approximate position of the MLPA probes; the shaded columns indicate the hotspots.

Using long range PCR (see materials and methods) it was possible to precisely define the sequence at the breakpoint in 4 of the samples, indicating that these rearrangements were in tandem. All of the breakpoints had an insertion or deletion of one or more nucleotides at the breakpoint junction.

An isolated sample tested was derived from amniocytes, and showed a deletion of exon 63 at the mRNA level. Exon 63 was present in the genomic DNA, and sequence analysis did not reveal any mutation that might affect splicing. Analysis with MLPA showed that the mutation at DNA level was in fact a duplication of exons 52-62. In order to gain a greater insight as to how such a rearrangement could lead to the skipping of an exon we performed further RT-PCR analysis. By amplifying different fragments it was possible to show that the rearrangement juxtaposed the duplicated exon 62 onto the original exon 64 (Figure 3).

Discussion

We have screened DMD/BMD patients for duplications using the MAPH and MLPA techniques. The patients could be divided into different cohorts, the first unselected and the second where deletions and

point mutations had already been excluded. Duplications were found in 8% of the first group, an overall duplication frequency comparable with what other studies have found (White et al., 2002; Den Dunnen et al., 1989). In patients that had already been screened for deletions and point mutations we detected duplications in 64% of cases. As these samples have not been tested at the RNA level, it is likely that the majority of the remaining patients have mutations affecting splicing. Studies have shown that mutations deep in intronic sequences can activate cryptic splice sites (Beroud et al., 2004), changes that can only be detected by mRNA analysis (Roest et al., 1993; Tuffery-Giraud et al., 2004).

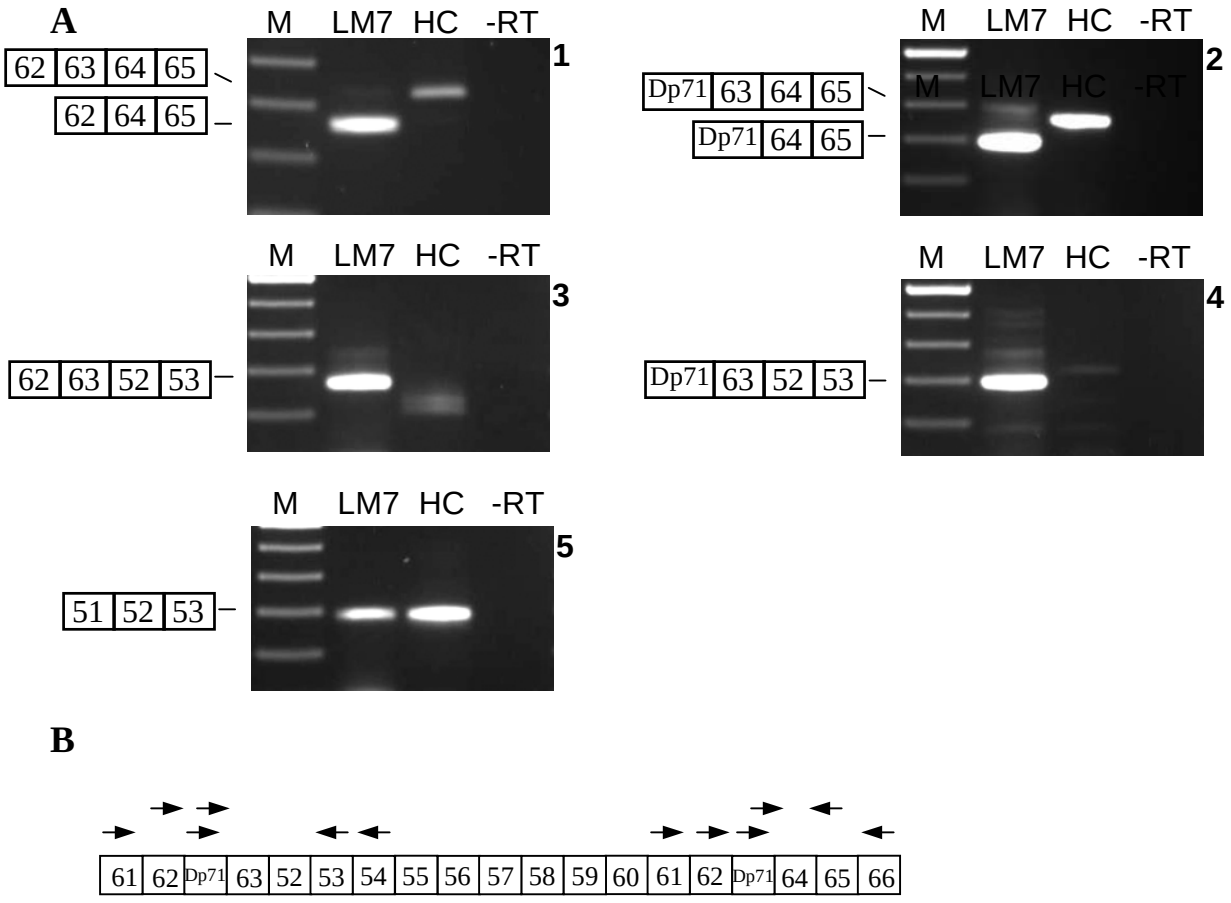

Figure 3 – An example of how a duplication at DNA level can lead to an apparent deletion at RNA level. A. RT-PCR analysis of LM7 and a healthy control (HC) encompassing the duplication breakpoints of LM7 in full-length dystrophin (Dp427m) and Dp71 transcripts. Using primers flanking exon 63 a shorter fragment lacking exon 63 was detected for LM7 when compared to the control, both in the Dp427m and Dp71 transcripts (upper two panels). Using a reversed primer in exon 53 in combination with forward primers in exon 62 (third panel) or in the first exon of Dp71 (fourth panel) a clear band containing exon 62 and 63, or the Dp71 first exon spliced to exon 52 could be produced for LM7, whereas only aspecific bands were observed for the control. Using the same cDNA and reversed primer, a fragment could be amplified for both LM7 and the control with forward primers in exon 51 (lower panel). M is 100 bp size marker, -RT is negative control.
B. The order of the exons in the duplicated region, with the arrows indicating the different primers that were used in determining the borders of the rearrangement.

Combining the data in this report with all other duplications previously described, and comparing these with the most common deletions, shows a marked difference in distribution of the two types of rearrangements (table 2). More than 50% of duplications are located in the 5' end of the gene, whereas the most common deletions cluster almost exclusively between exons 45-52. The reason for this difference is not clear, but the disparity, coupled with the far greater incidence of deletions, suggests that there are fundamental differences in the origins of the two types of rearrangements in the *DMD* gene.

It has been previously described for several genes on the X chromosome, including the *DMD* gene, that deletions are predominantly maternally inherited, whereas duplications mostly originate in the male germline (Hu et al., 1990; Grimm et al., 1994). It has also been shown that the distribution of mutations in the DMD gene when comparing mosaic and nonmosaic cases is significantly different (Passos-Bueno et al., 1992). It is likely that different chromosomal regions show varying levels of susceptibility to rearrangement, depending on the type or stage of cell division. This has been described for rearrangements in the NF1 gene, where the extent of the deletion is dependent on whether the rearrangement occurs during mitosis or meiosis (Kehrer-Sawatzki et al., 2004). The specific mechanism responsible for the rearrangement may also be dependent on factors such as cell type and position in the cell cycle.

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. Data from The Leiden Muscular Dystrophy pages (LMDp), based on the DMD_deldup database at 2004/10/26 (Data kindly provided by Ivo Fokkema).

Several different mechanisms are known to cause rearrangments (Shaffer and Lupski, 2000; Helleday, 2003; Hu et al., 1991). For example, intrachromatid looping can occur through the interaction of palindromic sequences, with the intervening sequence being deleted. This mechanism is not expected to produce a duplication. In contrast, unequal crossing over is expected to produce deletions and duplications at an equal frequency. This mechanism has been demonstrated to be responsible for many genomic disorders, via nonallelic homologous recombination between low copy repeats (Emanuel and Shaikh, 2001; Ji et al., 2000; Stankiewicz and Lupski, 2002). Unequal crossing over between Alu repeats has also been described within several genes, including *DMD* (Prior et al., 1997; Deininger and Batzer, 1999; Hu et al., 1991). The 3 duplication breakpoints in the *DMD* gene that had previously been described showed that all were tandem duplications. In one case homologous recombination between Alu elements was seen, the other 2 cases appeared to be due to non-homologous end joining (NHEJ), possibly mediated by topoisomerases. It was suggested that the rearrangements were the result of unequal crossing over between sister chromatids (Hu et al., 1991; Hu et al., 1990; Hu et al., 1989) .

The most frequently occurring duplication described to date in the *DMD* gene is of exon 2. Our initial analysis of the size and location of 11 unrelated exon 2 duplication cases showed that they differed in size. Sequencing of four of the breakpoints of the exon 2 duplications showed the removal or addition of one or more nucleotides at the junction, consistent with NHEJ. Scattered breakpoints is another characteristic feature of NHEJ, yet two distinct hotspots of ~4 kb and ~7 kb were seen, each containing five of the intron 2 breakpoints. This suggests that the event initiating the duplication occurred in intron 2.

It is interesting to note that a deletion of only exon 2 has never been reported. There are several possible reasons for this. It could be that such deletions do occur, but do not result in a severe DMD phenotype. It is also possible that such a mutation is somehow embryonic lethal. These explanations seem unlikely, as there is no obvious reason why an out of frame deletion should not lead to a DMD-like phenotype, when other deletions in the same region do. Likewise, larger deletions encompassing exon 2 have been described in DMD patients, making it unlikely that an individual with a smaller deletion in the same region would be non-viable before birth. A more feasible explanation is that the mechanism responsible for the duplication does not involve unequal crossing over. An example of such a mechanism is synthesis dependent NHEJ (Helleday, 2003). This will result in a tandem duplication at the site of a double stranded break, without unequal crossing over taking place. If the repair can procede in either direction, then one would also expect a similar number of duplications starting in intron 2, i.e. extending from exon 3. As can be seen in table 2, this is in fact the case, where duplications starting at exon 3 occur at a similar frequently to exon 2 duplications. Alternative mechanisms for duplication and deletion formation may therefore underly the differences seen in distribution and parent-of-origin of the two mutation types.

We found 4 cases involving non-contiguous duplications and/or triplications, suggesting multiple mutational events. The chance of 2 independent duplications is low, and it is more likely that the initial duplication was larger, covering the full extent of the different rearrangments. It is known that duplications can be unstable, with reversion to the normal situation (Helleday et al., 1998; Monnat et al., 1992; Hu et al., 1990). It is possible that this reversion will not always be absolute, leading to unusual, non-contiguous rearrangements. Notably in two of these cases a duplication would have been detected in both samples using probes located in the hotspot regions, yet the complete extent of the rearrangements would not have been identified. Indeed, in the case of sample D29, the duplication of exons 45-55 is expected to be an in-frame mutation, which is presumed to lead to a BMD-type phenotype. This patient in fact has a DMD phenotype, presumably as the more 3' duplication disrupts the reading frame. Despite non-contiguous rearrangements being rare events, the fact that they have been shown to occur reinforces the importance of screening the entire gene, in particular when the reading frame rule is going to be used prognistically.

The application of the reading frame rule to duplications assumes that the duplication is in tandem. Although this has been the case in all the *DMD* gene duplication breakpoints described to date (Hu et al., 1991), the consequence of a duplication on the RNA molecule may not always be as simple. An example of this was a duplication of exon 52-62 seen in sample LM7. The actual effect of this mutation at RNA level was an apparent skip of exon 63. This is of importance not only in the predictive diagnosis of disease progression, but also for targetted gene therapy (van Deutekom et al., 2001; Lu et al., 2003). In addition, it is theoretically possible that the apparent duplication is in fact a transposition, which, if outside the *DMD* gene, would have no effect on the reading frame.

In conclusion, we show here that duplications within the *DMD* gene are distributed differently compared to deletions, and can be complex rearrangements. In addition, we provide evidence that the mechanism involved in generating the exon 2 duplications, and by extension other duplications as well, does not involve unequal crossing over between sister chromatids.

References

1. Aartsma-Rus A., Bremmer-Bout M., Janson A.A., Den Dunnen J.T., van Ommen G.J., and van Deutekom J.C. (2002). Targeted exon skipping as a potential gene correction therapy for Duchenne muscular dystrophy. *Neuromuscul.Disord.* **12 Suppl 1**, S71-S77.

2. Aartsma-Rus A., Janson A.A., Kaman W.E., Bremmer-Bout M., Den Dunnen J.T., Baas F., van Ommen G.J., and van Deutekom J.C. (2003). Therapeutic antisense-induced exon skipping in cultured muscle cells from six different DMD patients. *Hum.Mol.Genet.* **12**, 907-914.

3. Abbs S.J. and 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.

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

5. Beroud C., Carrie A., Beldjord C., Deburgrave N., Llense S., Carelle N., Peccate C., Cuisset J.M., Pandit F., Carre-Pigeon F., Mayer M., Bellance R., Recan D., Chelly J., Kaplan J.C., and Leturcq F. (2004). Dystrophinopathy caused by mid-intronic substitutions activating cryptic exons in the DMD gene. *Neuromuscul.Disord.* **14**, 10-18.

6. Chamberlain J.S., Gibbs R.A., Ranier J.E., Nga Nguyen P.N., and Caskey C.T. (1988). Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucl.Acids Res.* **23**, 11141-11156.

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

8. Den Dunnen J.T., Grootscholten P.M., Bakker E., Blonden L.A.J., Ginjaar H.B., Wapenaar M.C., Van Paassen H.M.B., Van Broeckhoven C., Pearson P.L., and Van Ommen G.J.B. (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.

9. Den Dunnen J.T., Grootscholten P.M., Dauwerse J.D., Monaco A.P., Walker A.P., Butler R., Anand R., Coffey A.J., Bentley D.R., Steensma H.Y., and Van Ommen G.J.B. (1992). Reconstruction of the 2.4 Mb human DMD-gene by homologous YAC recombination. *Hum.Mol.Genet.* **1**, 19-28.

10. Emanuel B.S. and Shaikh T.H. (2001). Segmental duplications: an 'expanding' role in genomic instability and disease. *Nat.Rev.Genet.* **2**, 791-800.

11. Forrest S.M., Cross G.S., Speer A., Gardner-Medwin D., Burn J., and Davies K.E. (1987). Preferential deletion of exons in Duchenne and Becker Muscular Dystrophies. *Nature* **329**, 638-640.

12. Grimm T., Meng G., Liechti-Gallati S., Bettecken T., Muller C.R., and Muller B. (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.

13. Havenga M.J., Lemckert A.A., Ophorst O.J., van Meijer M., Germeraad W.T., Grimbergen J., van Den Doel M.A., Vogels R., Van Deutekom J., Janson A.A., de Bruijn J.D., Uytdehaag F., Quax P.H., Logtenberg T., Mehtali M., and Bout A. (2002). Exploiting the natural diversity in adenovirus tropism for therapy and prevention of disease. *J.Virol.* **76**, 4612-4620.

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

15. Helleday T., Arnaudeau C., and Jenssen D. (1998). A partial hprt gene duplication generated by non-homologous recombination in V79 Chinese hamster cells is eliminated by homologous recombination. *J.Mol.Biol.* **279**, 687-694.

16. Hu X., Burghes A.H.M., Bulman D.E., Ray P.N., and Worton R.G. (1989). Evidence for mutation by unequal sister chromatid exchange in the Duchenne muscular dystrophy gene. *Am.J.Hum.Genet.* **44**, 855-863.

17. Hu X., Ray P.N., Murphy E., Thompson M.W., and Worton R.G. (1990). Duplicational mutation at the Duchenne muscular dystrophy locus: its frequency, distribution, origin and phenotype/genotype correlation. *Am.J.Hum.Genet.* **46**, 682-695.

18. 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.

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

20. Ji Y., Eichler E.E., Schwartz S., and Nicholls R.D. (2000). Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res.* **10**, 597-610.

21. Kehrer-Sawatzki H., Kluwe L., Sandig C., Kohn M., Wimmer K., Krammer U., Peyrl A., Jenne D.E., Hansmann I., and Mautner V.F. (2004). High frequency of mosaicism among patients with neurofibromatosis type 1 (NF1) with microdeletions caused by somatic recombination of the JJAZ1 gene. *Am.J.Hum.Genet.* **75**, 410-423.

22. Lu Q.L., Mann C.J., Lou F., Bou-Gharios G., Morris G.E., Xue S.A., Fletcher S., Partridge T.A., and Wilton S.D. (2003). Functional amounts of dystrophin produced by skipping the mutated exon in the mdx dystrophic mouse. *Nat.Med.* **9**, 1009-1014.

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

24. Monnat R.J.jr., Chiaverotti T.A., Hackmann A.F., and Maresh G.A. (1992). Molecular structure and genetic stability of human hypoxanthine phosphoribosyltransferase (HPRT) gene duplications. *Genomics* **13**, 788-796.

25. Oudet C., Hanauer A., Clemens P., Caskey C.T., and Mandel J.L. (1992). Two hot spots of recombination in the DMD-gene correlate with the deletion prone regions. *Hum.Mol.Genet.* **1**, 599-603.

26. Passos-Bueno M.R., Bakker E., Kneppers A.L.J., Takata R.I., Rapaport D., Den Dunnen J.T., Zatz M., and Van Ommen G.J.B. (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.

27. Prior T.W., Bartolo C., Papp A.C., Snyder P.J., Sedra M.S., Burghes A.H., Kissel J.T., Luquette M.H., Tsao C.Y., and Mendell J.R. (1997). Dystrophin expression in a Duchenne muscular dystrophy patient with a frame shift deletion. *Neurol.* **48**, 486-488.

28. Roest P.A.M., Roberts R.G., Van Der Tuijn A.C., Heikoop J.C., Van Ommen G.J.B., and Den Dunnen J.T. (1993). Protein truncation test (PTT) to rapidly screen the DMD-gene for translation-terminating mutations. *Neuromusc.Disord.* **3**, 391-394.

29. Rooms L., Reyniers E., van Luijk R., Scheers S., Wauters J., Ceulemans B., Van Den E.J., Van Bever Y., and Kooy R.F. (2004). Subtelomeric deletions detected in patients with idiopathic mental retardation using multiplex ligation-dependent probe amplification (MLPA). *Hum.Mutat.* **23**, 17-21.

30. Schouten J.P., McElgunn C.J., Waaijer R., Zwijnenburg D., Diepvens F., and Pals G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30**, e57.

31. Shaffer L.G. and Lupski J.R. (2000). Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annu.Rev.Genet.* **34**, 297-329.

32. Slater H.R., Bruno D.L., Ren H., Pertile M., Schouten J.P., and Choo K.H. (2003). Rapid, high throughput prenatal detection of aneuploidy using a novel quantitative method (MLPA). *J.Med.Genet.* **40**, 907-912.

33. Stankiewicz P. and Lupski J.R. (2002). Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74-82.

34. Taylor C.F., Charlton R.S., Burn J., Sheridan E., and Taylor G.R. (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.

35. Tuffery-Giraud S., Saquet C., Chambert S., Echenne B., Marie C.J., Rivier F., Cossee M., Philippe C., Monnier N., Bieth E., Recan D., Antoinette V.M., Perelman S., Lambert J.C., Malcolm S., and Claustres M. (2004). The role of muscle biopsy in analysis of the dystrophin gene in Duchenne muscular dystrophy: experience of a national referral centre. *Neuromuscul.Disord.* **14**, 650-658.

36. van Deutekom J.C., Bremmer-Bout M., Janson A.A.M., Ginjaar H.B., Baas F., Den Dunnen J.T., and Van Ommen G.J.B. (2001). Antisense-induced exon skipping restores dystrophin expression in DMD patient-derived muscle cells. *Hum.Mol.Genet.* **10**, 1547-1554.

37. White S., Kalf M., Liu Q., Villerius M., Engelsma D., Kriek M., Vollebregt E., Bakker B., van Ommen G.J., Breuning M.H., and Den Dunnen J.T. (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.

38. White S.J., Vink G.R., Kriek M., Wuyts W., Schouten J., Bakker B., Breuning M.H., and Dunnen J.T. (2004). Two-color multiplex ligation-dependent probe amplification: Detecting genomic rearrangements in hereditary multiple exostoses. *Hum.Mutat.* **24**, 86-92.

Chapter 2.3

White S.J., Sterrenburg E., van Ommen G.J., Den Dunnen J.T., Breuning, M.H. (2003). An alternative to FISH: detecting deletion and duplication carriers within 24 hours. *J.Med.Genet.* 40 (10):e113.

An alternative to FISH: detecting deletion and duplication carriers within 24 hours

S J White, E Sterrenburg, G-J B van Ommen, J T den Dunnen, M H Breuning

J Med Genet 2003;**40**:e113 (http://www.jmedgenet.com/cgi/content/full/40/10/e113

A range of genetic disorders has been revealed to be caused by deletions and duplications within the genome.[1–3] In addition, computational analysis of the recently completed human genome sequence[4] suggests that many more rearrangements might exist. Such rearrangements are either directly involved in genetic disease or may play an important, but yet to be determined, role in human variation and multifactorial diseases. Efficient methods are thus required to screen for and detect such rearrangements.

While changes of several megabases are usually cytogenetically visible, smaller changes require other methods of analysis. Many techniques have been applied, including dinucleotide repeat polymorphism analysis,[5] array comparative genomic hybridisation,[6] fluorescent in situ hybridisation (FISH),[7 8] quantitative multiplex PCR,[9 10] and Southern blotting.[11 12] The last three mentioned are the most commonly applied techniques,[13] with FISH analysis preferred as the method of choice in many clinical centres. FISH has the advantage that the analysis is visual, with the number of fluorescent signals determining the copy number of the region examined. However, the method is rather laborious, with cell culturing and preparation of metaphase spreads being necessary, but difficult and time consuming steps. FISH is thus expensive and not suitable for high throughput analysis. In addition, as FISH probes are usually artificial chromosomes or cosmids, it precludes the analysis of small rearrangements, and duplications can be difficult to detect.

Quantitative multiplex PCR seems an attractive alternative. It can co-amplify up to 15 products per sample, with the amount of each product corresponding to the copy number of the locus. However, achieving consistent results has proven to be technically challenging, and the method requires fluorescent labels and sophisticated equipment.

Southern blotting is more flexible and does not require sophisticated equipment. Its disadvantages are that it is laborious, requiring several blots if multiple loci are to be examined, and its accuracy critically depends on the quality of the blot, with duplications being particularly difficult to detect.

We have applied an alternative method, based on multiplex amplifiable probe hybridisation (MAPH).[14] MAPH facilitates the quantitative recovery of probes hybridised to immobilised genomic DNA, and thus the detection of deletions and duplications. Previous studies have separated the resultant PCR products on acrylamide gels or with a capillary sequencer, using a radioactively[14] or fluorescently[15] labelled primer respectively. To speed up the analysis, we used a chip based gel electrophoresis system (Lab-on-a-chip; Agilent, Palo Alto, CA, USA) to analyse and quantify the reaction products. This system analyses 12 unlabelled samples in ~30 min, with quantitative data being generated automatically by the accompanying software.

We have tested the efficacy and reliability of this methodology by performing carrier detection in Duchenne muscular dystrophy (DMD). This lethal disease is caused by a deletion or duplication of one or more of the 79 exons of the DMD gene in ~70% of cases.[11 16] As the DMD gene is located on the X chromosome, deletion screening in male DMD patients is relatively simple.[17 18] Detecting duplications or carrier status in females, however, requires a quantitative method of analysis. By selecting probes for exons within and outside the rearranged regions, it is possible to compare the relative ratios for the two groups. As multiple probes in parallel hybridisations are used, a high level of redundancy, and thus reliability, can be obtained.

In this paper, we show the validity of this approach by analysing 17 potential carriers for deletion/duplication mutations.

Key points

- When a deletion or duplication mutation has been detected in an index case, relatives may wish to be analysed for carrier status. Methods currently applied are either technically demanding, time consuming or not always applicable.
- We have previously described multiplex amplifiable probe hybridisation (MAPH) as a versatile method for the detection of deletions and duplications, applied to the analysis of Duchenne muscular dystrophy patients.
- Here we show that MAPH is a reliable, quick, and inexpensive alternative for fluorescent in situ hybridisation as a method for carrier detection of deletion/duplication mutations. Following MAPH-based hybridisation and PCR, the amplification products are separated using ''Lab-on-a-chip'' electrophoresis, which quantitatively processes 12 samples in 30 minutes.
- The method is very rapid, taking less than 24 h. Moreover, as several independent probes and duplicates can be run in parallel, it is also very reliable. This approach is an attractive alternative for current FISH-based screens, and should especially facilitate genetic counselling in situations where a rapid diagnosis is important.

METHODS

Probe preparation and the MAPH protocol used have been described previously.[15] Based on the mutation to be tested, a specific set of probes were selected. Where possible, at least

Abbreviations: FISH, fluorescent in situ hybridisation; MAPH, multiplex amplifiable probe hybridisation; MLPA, multiplex ligation dependant probe amplification

two probes within the rearrangement were included, with a minimum of 1 exon from an unaffected region of the gene. In addition, at least two control probes were chosen from a set of autosomal probes. A minimum of two hybridisations were performed on each sample; if the mutation was of a single exon, then three separate hybridisations with the specific probe were carried out.

Following hybridisation and washing, the PCR reaction was performed as previously described,[15] with both primers being unlabelled. Bioanalyzer 2100 (Agilent) analysis was carried out according to the manufacturer's instructions (http://www.chem.agilent.com). Briefly, the DNA500 chip was preloaded with a gel matrix containing a DNA dye. From each PCR sample, 1 µl (~10 ng) of product was added, with a maximum of 12 samples loaded per chip. The samples were then separated, with the data being subsequently exported to Excel (Microsoft Corp.).

Exon specific peaks were normalised within each sample to unlinked probes, with each exon subsequently being normalised to 1.0 based on those samples known to be unaffected at the respective loci.

Ratios derived from probes outside the rearranged regions were compared with those from probes within the rearranged regions with an independent samples Student's *t* test. An individual was considered to be a carrier of the mutation if the difference between the two groups was statistically significant ($p<0.01$). Confidence intervals of 99% were calculated, giving a predicted error rate of 1%. Statistical analysis was performed using SPSS 10.0.7 (SPSS Inc., Palo Alto, CA, USA).

RESULTS

Analysis started with the selection of the probes to be tested. After hybridisation and subsequent amplification, the PCR products were separated on the Lab-on-a-chip. In the resulting trace pattern, each peak corresponded to a specific probe. As shown in fig 1, changes in peak height and area correspond to a deletion or duplication at that specific locus. Although most mutations could be detected visually, quantitative analysis was always performed. The area underneath each peak was calculated by the Bioanalyzer software and subsequently tabulated in Excel. A typical example is shown in table 1. In this analysis, six samples were tested: two deletion and four duplication carriers. Based on the exons known to be affected, four DMD exon probes were chosen, ensuring that for each sample at least one exon gave a normalised ratio of ~1.0. This probe represents the control for hybridisation quality. As can be seen in this example, deletions and duplications could be detected as ratios of around 0.5 and 1.5 respectively. All samples were screened at least twice, with the data from each sample being collated.

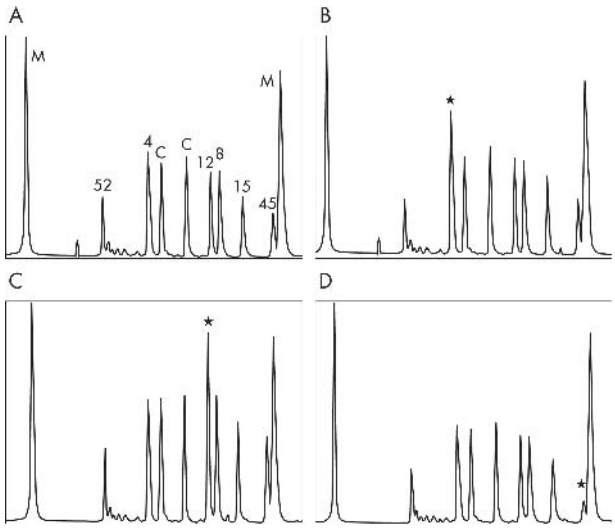

Figure 1 An example of the trace patterns obtained from the Bioanalyzer software. Changes in the peak height and area correspond to changes in copy number of the specific probe. The numbers refer to DMD exons, with autosomal control probes indicated with C. M indicates the two marker alignment peaks, at 15 and 600 bp. These are used by the software for lane to lane alignment. Four different cases are shown here: A, no mutation; B, duplication exon 4; C, duplication exon 12; D, deletion exon 45. In each case, the affected exon is indicated with an asterisk.

Table 1 Lab-on-a-chip analysis

	Sample number						
	1	2	3	4	5	6	Mean
A							
Exon 52	5.31	6.86	8.08	14.81	7.27	6.40	—
Exon 62	12.66	8.14	13.68	12.16	5.71	5.08	—
C1	16.77	12.11	14.50	18.43	8.84	7.06	—
C2	16.12	8.86	11.54	13.42	7.03	5.69	—
C3	17.74	12.14	13.73	18.08	8.61	6.82	—
Exon 54	6.68	8.71	10.93	18.75	9.51	6.76	—
Exon 49	7.32	4.34	10.18	12.46	6.60	4.88	—
B							
Exon 52	0.10	0.21	0.20	0.30	0.30	0.33	0.21
Exon 62	0.25	0.25	0.34	0.24	0.23	0.26	0.25
Exon 54	0.13	0.26	0.27	0.38	0.39	0.35	0.27
Exon 49	0.14	0.13	0.26	0.25	0.27	0.25	0.25
C							
Exon 52	*0.48*	1.00	0.95	**1.43**	**1.43**	**1.57**	—
Exon 62	1.00	1.00	**1.40**	0.96	0.92	1.04	—
Exon 54	*0.48*	0.96	1.00	**1.41**	**1.44**	**1.30**	—
Exon 49	*0.56*	*0.52*	1.04	1.00	1.08	1.00	—

Following electrophoresis, the peak data areas from six samples (1–6) were imported into Excel from the Bioanalyzer software (Section A).
By dividing the area under each exon (specific peak is divided by the sum of the area of the control peaks), a ratio for each exon was obtained (Section B).
These ratios are were then normalised to 1.0 based on the mean ratio of samples known to be unaffected at that specific locus (Section C).
The normalised ratios of the exons that are duplicated are shown in **bold**, those of the deleted exons are in *italics*.

In total, 17 potential DMD carriers were analysed, with the results summarised in table 2. The extent of the mutations varied, ranging from a deletion or duplication of a single exon to a deletion of 37 exons. Of the 17 samples tested, 13 were shown to be mutation carriers. This agreed completely with the results found with other methods, namely FISH, Southern blotting or by MAPH analysed by capillary electrophoresis.

Although duplications are known to be more difficult to detect than deletions, the results were unequivocal in all cases. All carriers had a p value of <0.001, whereas the four non-carriers had p values ≥ 0.10.

DISCUSSION

We describe a novel method for the clinical diagnosis of deletion/duplication mutations, which we consider an attractive alternative for FISH analysis. Based on prior knowledge as to where a mutation might be (index patient), a set of probes is selected, of which some are located inside the rearranged region, some directly flanking and some from other, unrelated regions in the genome. Rapid, quantitative analysis of the reaction products is possible using the Lab-on-a-chip from Agilent. This chip allows the electrophoretic separation of 12 samples in ~30 min, providing a detailed analysis of each peak.

Unless the suspected mutation was of a single exon, at least two probes within the region of interest were chosen,

Table 2 The 17 samples examined

Case no.	Mutation in son	Mean ratio within rearrangement (n)	Mean ratio outside rearrangement (n)	99% CI of the difference	p Value	Carrier?
1	dup 58-63	**1.44** (3)	1.01 (11)	−0.58 to −0.28	<0.001	Yes
2	del 10-46	*0.47* (10)	0.97 (13)	0.30 to 0.70	<0.001	Yes
3	dup 44-57	**1.51** (13)	1.07 (24)	−0.58 to −0.31	<0.001	Yes
4	dup 50-55	**1.39** (6)	0.98 (19)	−0.51 to −0.30	<0.001	Yes
5	dup 52-55	**1.48** (6)	1.03 (13)	−0.61 to −0.29	<0.001	Yes
6	dup 51-55	**1.60** (7)	0.99 (18)	−0.94 to −0.26	<0.001	Yes
7	del 45	*0.39* (3)	1.02 (15)	0.46 to 0.80	<0.001	Yes
8	del 49-54	*0.51* (10)	1.00 (19)	0.39 to 0.59	<0.001	Yes
9	del 48-50	*0.53* (5)	1.01 (12)	0.41 to 0.55	<0.001	Yes
10	dup 2-9	1.01 (4)	0.98 (11)	−0.16 to 0.11	0.63	No
11	dup 3-7	**1.43** (6)	0.94 (20)	−0.65 to −0.32	<0.001	Yes
12	dup 12-13	**1.47** (4)	1.03 (17)	−0.64 to −0.23	<0.001	Yes
13	dup 2-6	**1.28** (4)	1.01 (17)	−0.42 to −0.11	<0.001	Yes
14	dup 2-7	1.07 (4)	0.94 (8)	−0.35 to 0.12	0.13	No
15	del 52	*0.55* (3)	0.96 (12)	0.10 to 0.63	<0.001	Yes
16	del 8-43	1.00 (4)	0.96 (6)	−0.24 to 0.15	0.47	No
17	dup 12	1.10 (3)	1.00 (12)	−0.28 to 0.07	0.10	No

Listed are the ratios derived from probes within and outside the rearrangements.
The mean ratio for each sample is given (duplicated in **bold**, deleted in *italics*), with the figure in brackets being the number of probes tested.
The p values were determined with Student's *t* test, and the associated 99% confidence intervals (CI) of the differences are also shown.

and all samples were tested in at least two hybridisations (three hybridisations for single exon mutations). Due to the simplicity of the technique, it is little extra effort to perform these hybridisations in parallel, and no time is lost. Data derived from the different hybridisations for each sample were collated, and the ratios were separated into two groups based on whether the probes were localised within or outside the potential breakpoints. By combining the data, the potential influence of any false positives and negatives was minimised. Previous studies have used different methods of assessing a positive result, ranging from setting arbitrary boundaries of 0.75 and 1.25,[19] to bivariate analysis for each affected probe.[20] We have taken advantage of the fact that the potential mutation was already known, by comparing the ratios derived from probes within and outside the rearranged region. If the difference was not statistically significant ($p>0.01$) then it was assumed that the individual was not a carrier. Conversely, a significant difference was taken to indicate the presence of the suspected mutation. This was confirmed by the results obtained. As can be seen by the 99% confidence intervals, the actual error rate will be considerably lower than the 1% predicted.

In some cases, the mother may be a mosaic, meaning that the mutation will not be present in all cells. This makes the analysis more difficult. Whether such cases would be detected by the described method depends on several factors, including the standard deviation of the probes, the number of different probes that can be used, and the degree of mosaicism. Due to the influence of the unaffected cells, a p value between 0.01 and 0.1 may occur, prompting further analysis.

There are several advantages to using MAPH in combination with the Lab-on-a-chip. It can be broadly applied, as a variety of probes can be chosen and all can be used under identical PCR conditions. The resolution is limited only by the size of the probes, which can be as short as 100 base pairs. Analysis is rapid, simple and can be readily automated, as data can be exported to Excel. The DNA chip can measure DNA fragments at less than 1 ng, meaning that unlabelled samples can be directly loaded on the chip without any prior concentration.

The advantages described here for MAPH based analysis also apply to a similar technique, multiplex ligation dependant probe amplification (MLPA).[21] MLPA is based on the specific hybridisation and subsequent ligation of two oligonucleotides, with only ligated end products generating a target for PCR amplification. MLPA has the advantage of being a ''single tube'' assay, and requiring less input DNA. However, compared to MAPH, probe preparation for MLPA is more time consuming. The method of choice would be based on the exact goal and probe availability.

Many probes for MAPH/MLPA have already been developed[15 19–22] and as more probes become available, the possibility of screening other regions of the genome increases (Kriek *et al*, manuscript in preparation). The combination of these techniques with a rapid and simple method of analysis should allow diagnostic laboratories to implement this as a broadly applicable, robust, and readily automated method for high resolution copy number determination.

ACKNOWLEDGEMENTS

We would like to thank Professor Bert Bakker (LUMC) for providing the DNA samples. This work was financially supported by Zon MW (Grant 2100.0026).

Authors' affiliations

S J White, E Sterrenburg, G-J B van Ommen, J T den Dunnen, M H Breuning, Human and Clinical Genetics, Leiden University Medical Center, Wassenarrseweg 72, Leiden, the Netherlands

Correspondence to: Dr Johan T den Dunnen , Human and Clinical Genetics Leiden University Medical Center Wassenarrseweg 72 Leiden, the Netherlands; ddunnen@lumc.nl

REFERENCES

1 **Petrij-Bosch A**, Peelen T, van Vliet M, van Eijk R, Olmer R, Drusedau M, Hogervorst FB, Hageman S, Arts PJ, Ligtenberg MJ, Meijers-Heijboer H, Klijn JG, Vasen HF, Cornelisse CJ, van 't Veer LJ, Bakker E, van Ommen GJ, Devilee P. BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. *Nat Genet* 1997;**17**:341–5.
2 **Patel PI**, Garcia C, Montes de Oca-Luna R, Malamut RI, Franco B, Slaugenhaupt S, Chakravarti A, Lupski JR. DNA duplication associated with Charcot-Marie-Tooth disease type 1a. *Cell* 1991;**66**:219–32.
3 **Kurotaki N**, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, Ohashi H, Naritomi K, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Hasegawa T, Chinen Y, Tomita Ha HA, Kinoshita A, Mizuguchi T, Yoshiura Ki K, Ohta T, Kishino T, Fukushima Y, Niikawa N, Matsumoto N. Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat Genet* 2002;**30**:365–6.

4 **Bailey JA**, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. Recent segmental duplications in the human genome. *Science* 2002;**297**:1003–7.

5 **Clemens PR**, Fenwick RG, Chamberlain JS, Gibbs RA, de Andrade M, Chakraborty R, Caskey CT. Carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophy families, using dinucleotide repeat polymorphisms. *Am J Hum Genet* 1991;**49**:951–60.

6 **Pinkel D**, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998;**20**:207–11.

7 **Rosenberg C**, Navajas L, Vagenas DF, Bakker E, Vainzof M, Passos-Bueno MR, Takata RI, Van Ommen GJ, Zatz M, Den Dunnen JT. Clinical diagnosis of heterozygous dystrophin gene deletions by fluorescence *in situ* hybridization. *Neuromusc Disord* 1998;**8**:447–52.

8 **Petrij F**, Dauwerse HG, Blough RI, Giles RH, van der Smagt JJ, Wallerstein R, Maaswinkel-Mooy PD, van Karnebeek CD, van Ommen GJ, van Haeringen A, Rubinstein JH, Saal HM, Hennekam RC, Peters DJ, Breuning MH. 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* 2000;**37**:168–76.

9 **Duponchel C**, Di Rocco C, Cicardi M, Tosi M. Rapid detection by fluorescent multiplex PCR of exon deletions and duplications in the C1 inhibitor gene of hereditary angioedema patients. *Hum Mutat* 2001;**17**:61–70.

10 **Yau SC**, Bobrow M, Mathew CG, Abbs SJ. Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis. *J Med Genet* 1996;**33**:550–8.

11 **Den Dunnen JT**, Grootscholten PM, Bakker E, Blonden LA, Ginjaar HB, Wapenaar MC, van Paassen HM, van Broeckhoven C, Pearson PL, van Ommen GJ. Topography of the DMD gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. *Am J Hum Genet* 1989;**45**:835–47.

12 **van Overveld PG**, Lemmers RJ, Deidda G, Sandkuijl L, Padberg GW, Frants RR, van der Maarel SM. Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. *Hum Mol Genet* 2000;**9**:2879–84.

13 **Armour JA**, Barton DE, Cockburn DJ, Taylor GR. The detection of large deletions or duplications in genomic DNA. *Hum Mutat* 2002;**20**:325–37.

14 **Armour JA**, Sismani C, Patsalis PC, Cross G. Measurement of locus copy number by hybridisation with amplifiable probes. *Nucl Acids Res* 2000;**28**:605–9.

15 **White S**, Kalf M, Liu Q, Villerius M, Engelsma D, Kriek M, Vollebregt E, Bakker B, van Ommen GJ, Breuning MH, den Dunnen JT. Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. *Am J Hum Genet* 2002;**71**:365–74.

16 **Koenig M**, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* 1987;**50**:509–17.

17 **Chamberlain JS**, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucl Acids Res* 1988;**23**:11141–56.

18 **Beggs AH**, Koenig M, Boyce FM, Kunkel LM. Detection of 98-percent DMD/BMD gene deletions by polymerase chain reaction. *Hum Genet* 1990;**86**:45–8.

19 **Sismani C**, Armour JA, Flint J, Girgalli C, Regan R, Patsalis PC. 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* 2001;**9**:527–32.

20 **Hollox EJ**, Atia T, Cross G, Parkin T, Armour JA. High throughput screening of human subtelomeric DNA for copy number changes using multiplex amplifiable probe hybridisation (MAPH). *J Med Genet* 2002;**39**:790–5.

21 **Schouten JP**, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 2002;**30**:e57.

22 **Akrami SM**, Winter RM, Brook JD, Armour JA. Detection of a large TBX5 deletion in a family with Holt-Oram syndrome. *J Med Genet* 2001;**38**:E44.

Chapter 3

White S.J., Uitte de Willige S., Verbove D., Politano L., Ginjaar H.B., Breuning M.H., den Dunnen J.T. Sarcoglycanopathies and the risk of undetected deletion alleles in diagnosis. Submitted for publication.

Sarcoglycanopathies and the risk of undetected deletion alleles in diagnosis

Stefan J. White, Shirley Uitte de Willige, Dennis Verbove, Luisa Politano*, Ieke Ginjaar, Martijn H. Breuning and Johan T. den Dunnen.

Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; * Department of Clinical and Experimental Medicine and Surgery, 2nd Naples University, Naples, Italy

Abstract

We have designed Multiplex Amplifiable Probe Hybridization (MAPH) probes for each exon of the sarcoglycan genes *SGCA*, *SGCB*, *SGCG* and *SGCD* (33 in total). The set was used to screen DNA from limb-girdle muscular dystrophy (LGMD) patients for the presence of pathogenic deletion or duplication mutations. An unexpected heterozygous deletion of *SGCG* exon 7 was detected in a patient from a consanguineous family segregating a known c.525delT mutation. The exon 7 deletion was inherited from the father, who was part of the consanguineous c.525delT branch of the family but who screened negative for the c.525delT mutation. The deletion breakpoint was mapped, isolated and sequenced. Identical breakpoints were detected in 2 unrelated LGMD patients from Southern Italy. Haplotype analysis showed identical alleles segregating with the mutation in all three patients, suggesting a common ancestor. Exonic deletions in sarcoglycanopathies appear to be rare events. However, we recommend screening for exonic deletions / duplications in patients where a mutation has not been identified in both alleles, as well as in seemingly homozygous cases where segregation of the mutations can not be confirmed in the parents.

Introduction

The Limb-Girdle Muscular Dystrophies (LGMD) are a heterogeneous group of neuromuscular disorders, characterized by progressive wasting of the limb-girdle musculature. The proteins involved show a wide range of functions and cellular localization (1). A subset of these are α, β, γ and δ sarcoglycan, which are transmembrane glycoproteins that form part of the dystrophin-associated glycoprotein complex (DAG)(2,3). The absence of any one of these subunits leads to the reduction or loss of the other sarcoglycans, which in turn disrupt the DAG complex.

The sarcoglycanopathies are recessive disorders, however there are often cases where a mutation can be identified in only one allele. In addition, it is not always possible to confirm homozygous mutations through analysis of the parents. As the majority of sarcoglycan mutations described have been point mutations, screening at the DNA level has been almost exclusively performed by direct sequencing. This method however is not quantitative, meaning that deletions or duplications will not be detected.

Although several different methods have been described for detecting copy number changes (4), these have not yet been been applied for the sarcoglycanopathies in a routine setting. The most commonly used methods include quantitative Southern blotting (5), quantitative multiplex PCR (6) and fluorescent in situ hybridization (FISH) (7), but these are either time consuming, technically challenging or of insufficient resolution to find small rearrangements. We have used Multiplex Amplifiable Probe Hybridization (MAPH) (8) to look for copy number changes in the sarcoglycan α, β, γ and δ genes in 5 sarcoglycanopathy patients diagnosed as being homozygous for a point mutation, as well as a patient where a mutation in only one allele could be found.

Methods.

Patient details.

All patients were diagnosed by a neurologist as having a neuromuscular disorder, based on their clinical features. Immunohistochemical analysis was performed as previously described (9), and had shown in each case the presence of dystrophin and abnormal levels of the sarcoglycan proteins. Based on these results, sequence analysis of the gene(s) showing the lowest levels of expression was performed as previously described (9). This revealed an apparent homozygous point mutation in one of the sarcoglycan genes in 5 cases (LG 1-5), whereas a mutation was detected in only one allele in patient LG 6 (Table 1).

Patient	Gene Affected	Mutations detected by sequencing	exon(s) of gene
L1	*SGCG*	[c.525delT]+ [c.525delT]	6
L2	*SGCG*	[c.525delT]+ [c.525delT]	6
L3	*SGCG*	[c.525delT]+ [c.525delT]	6
L4	*SGCA*	[c.229C>T]+ [c.229C>T]	3
L5	*SGCB*	[c.275C>T]+ [c.275C>T]	3
L6	*SGCG*	[c.525delT]+ [c.579-4730_702+1721delinsACACTA]	6 and 7

Table 1. The patients tested in this study. The mutations are described based on the following reference sequences; *SGCA*:NM_000023.1, *SGCB*:NM_000232.3, *SGCG*:NM_000231.1. In each case +1 is taken to be the A of the ATG initiation codon.

This was a 5-year-old boy, born of non-consanguineous North African parents. He showed tip-toe walking and progressive lower limb weakness from the age of 2 years. Motor skills such as stair climbing and cycling were progressively lost. Clinical examination at the age of 5 years showed muscular weakness of the upper and lower limb girdle and Gower's sign, with no overt cognitive impairment. The patient has three healthy siblings. Further investigation showed a family history of neuromuscular disorders (Figure 1). The brother of the patient's paternal grandfather has three grandchildren (one boy and two girls) with a DMD-like phenotype: these grandchildren born from consanguineous Moroccan parents were diagnosed as LGMD2C patients, and one has been described in a previous paper (9). The siblings were homozygous for a 1 bp deletion (525delT) in the γ-sarcoglycan gene (*SGCG*), and both parents were demonstrated to be carriers for this mutation.

MAPH

To allow screening for whole exon rearrangements MAPH probes for the exons of the α, β, γ and δ sarcoglycan genes were developed (33 in total). Probe preparation and the MAPH protocol have been described (10). Briefly, a DNA sequence covering or in close proximity to each exon was PCR amplified and cloned into the pGEM-T easy vector (Promega). This sequence was reamplified with vector-specific primers. In this way all probes could be co-amplified with only one pair of primers. The probes were combined and hybridized to immobilized genomic DNA. Following stringent washing the remaining probes were recovered from the filter and amplified by PCR. Using a fluorescent primer and capillary electrophoresis the peaks corresponding to each

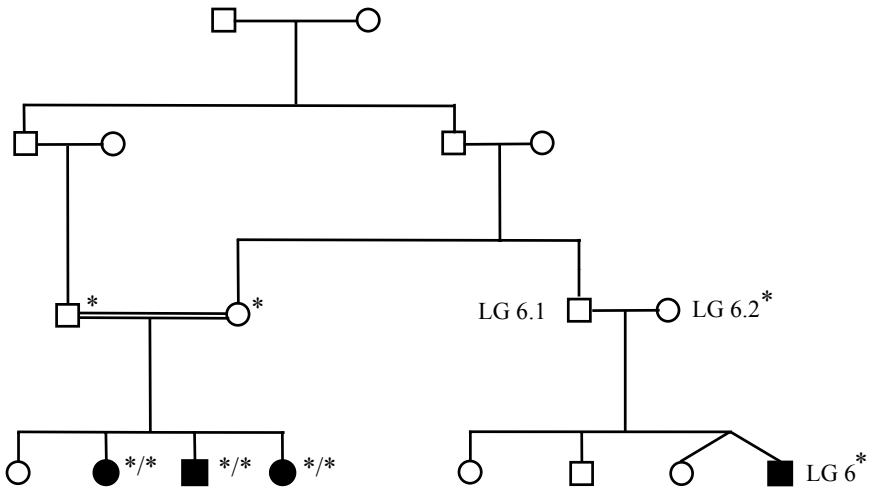

Figure 1. Extended pedigree of patient LG 6. Individuals affected by the sarcoglycanopathy are shown in black. All individuals indicated with a * are heterozygous for the point mutation c.525delT in the *SGCG* gene, */* indicates homozygosity for the same mutation. The mutation is described based on reference sequence NM_000231.1, with +1 taken to be the A of the ATG initiation codon.

probe were separated and quantified. Probes were normalized against the 4 nearest probes not from the same gene. A deletion was defined as a normalized ratio below 0.75, found in duplicate.

Long range PCR

Primers were designed using the Primer 3 program (11), based on sequence derived from the human genome working draft (April 2003). PCR was performed on 200-300 ng genomic DNA using the Expand Long Template PCR system (Roche), with products analyzed on a 0.7% agarose gel. Any samples that were to be sequenced were purified with the Qiagen PCR purification kit (Qiagen). Sequencing was performed at the Leiden Genome Technology Center using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) on an ABI 3700 Sequencer (Applied Biosystems).

Haplotype Analysis

The following polymorphic markers were used; D13S232, D13S292 and D13S115. The sequences used for amplification were obtained from the UNISTS site of the NCBI (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists), with the forward primer of each pair being labeled with the fluorescent labels FAM or HEX. After PCR amplification the three marker products were pooled and analyzed on an ABI 3700 Sequencer (Applied Biosystems).

Results

MAPH analysis

MAPH analysis was initially performed on genomic DNA from 12 healthy control samples to determine probe variation. In total 28 of the 33 probes gave consistent ratios, and were included in subsequent screening tests. The five probes excluded (*SGCA* exons 5 and 7, *SGCG* exon 3 and *SGCD* exons 2 and 4), did not cover any of the homozygous point mutations to be analyzed. Following MAPH analysis no copy number mutations were detected in any of the 5 patients diagnosed as having a homozygous point mutation. Analysis of DNA from patient LG 6 showed that exon 7 of the *SGCG* gene was heterozygously deleted, i.e. the height of the corresponding peak was ~50% of that in normal controls, whereas the other 27 probes were within normal limits. Testing DNA from the parents showed that the father (LG 6.1) was a carrier for the deletion, whereas the mother (LG 6.2) did not have the deletion. As there were several relatives of the father who also had been diagnosed with γ-sarcoglycanopathy, genomic DNA from these patients was also tested for an exon 7 deletion. No deletion was detected (data not shown).

Long Range PCR

To confirm the deletion and determine the breakpoints, long range PCR was initially performed on genomic DNA from both LG 6 and LG 6.1. As exon 8 of the *SGCG* gene was known to be present, the reverse primer was located within this sequence. Using the DNA sequence from the Human Genome Working Draft (April 2003) several forward primers were designed in intron 6. These primers were spaced at ~5kb intervals, extending from the 5' end of exon 6, until exon 7. With one particular pair of primers (forward – TGGATGTGGTATATATGCGATCGTG; reverse - GCTTGGGTAAGCACACAGTTTCAG) a 9kb fragment was expected, but a product of ~2.5 kb was amplified from both the patient and his father. Sequence analysis showed that the PCR product spanned the deletion breakpoint, and that the entire exon was deleted. It was not a simple rearrangement, with an insertion of 6 bp between the breakpoints (Figure 2).

Two unrelated patients (N66 and N950) from Italy, previously reported as being homozygous for a deletion of exon 7 of the *SGCG* gene (12), were also tested with the same primer pair. Both DNA samples gave a similar size PCR junction fragment, and sequence analysis revealed the identical deletion breakpoint as seen in LG 6. Analysis of the parents of N950 showed that both were heterozygous carriers of the deletion. As the identical mutation suggested either a recurrent mutation or a common ancestor, haplotype analysis was performed on the 3 patients and 25 CEPH control samples. The haplotype D13S115 (165 bp), D13S232 (112 bp) and D13S292 (205 bp) was found in 5/6 chromosomes in the affected individuals, and in a maximum of 4/50

chromosomes in the control samples. Analysis with Fisher's exact test showed this difference to be significant ($p<0.001$), strengthening the argument for a common ancestor.

```
22788031 CCAACTGGTGACCTCAGCAGAGAAGGATTTTAATAATCAAGTG 22788073
         ||||||||||||||||-------------------------
         CCAACTGGTGACCTCAGACACTAGCTCCGATGTCCTTCACAGA
         -----------------------|||||||||||||||||||
22794597 GTGTTGCAAGGGATGGGCCAGGCGCTCCGATGTCCTTCACAGA 22794639
```

Figure 2. The sequence across the deletion breakpoint (middle row). The six nucleotides underlined indicate the insertion. The numbers indicate the nucleotide position based on the sequence of chromosome 13 from the Human Genome working draft (May 2004).

Discussion.

We report here the identical single exon deletion in three unrelated individuals diagnosed with γ-sarcoglycanopathy. The extended family of patient LG 6 has been described (9). Several individuals had been diagnosed with γ-sarcoglycanopathy, but were all concluded to be homozygous for the single nucleotide deletion c.525delT. Unexpectedly, these were all on the paternal side, and the index patient inherited the point mutation c.525delT from his mother. We found that the father has an exon 7 deletion in one allele, with no other affected relatives tested having the same deletion.

Exonic deletions in the *SGCG* gene have been rarely described (table 2). There was a single report in the literature of a deletion of exon 7 of *SGCG*, which had been found homozygously in two unrelated individuals in a village in Southern Italy (12).

Comparison of the deletion junction in our patient (of North African heritage) and those of Italian origin showed that they were identical. Analysis of the deletion breakpoint showed an insertion of 6 bp, suggesting that the mechanism involved was non-homologous end-joining. Such rearrangements are associated with scattered breakpoints and insertions at the junction site (13,14), making it less likely that the deletions had occurred independently of each other. Indeed, haplotype analysis showed that the mutation segregated with the same haplotype in all 3 patients, supporting the hypothesis of a common ancestor.

There have been founder mutations described for several LGMD genes, although these have all been small mutations that were detected by sequencing. The c.525delT mutation in *SGCG*, seen

heterozygously in LG 6 and homozygously in LG 1, 2 and 3, is common within the North African population, segregating with the rare 122 bp allele of marker D13S232 (15). The mutation has also been found in Brazilians of African heritage, again segregating with the same allele (16). Another change within the *SGCG* gene, c.87insT was found in patients in Northern Italy (17). This mutation has not been described in other populations, and it was proposed that the mountainous surroundings in that region of Italy contributed to the relative genetic isolation. In contrast, it is likely that villages in Southern Italy had more contact with other regions via sea travel. It is therefore not surprising that a mutation in this region spreads across the Mediterranean.

***SGCG* deletion**	**Times Reported**	**References**
-124-?_1500+?del (entire gene)	2	(20,21)
c.506-?_578(702)+?del (exon 6*)	1	(22)
c.579-4730_702+1721delinsACACTA (exon 7)	2	(12), this report

Table 2. Deletions of one or more exons described for the *SGCG* gene. The mutations are described based on reference sequence NM_000231.1, with +1 taken to be the A of the ATG initiation codon. * In this case it was not determined whether exon 7 was also deleted.

Partial gene deletions or duplications as founder effects have been described, although less commonly than mutations that can be detected by sequencing. Within the Dutch population deletions of one or more exons of the *BRCA1* gene occur in ~30% of breast cancer cases (18). There was a report of a partial gene duplication within the *SCL3A* gene, found in several individuals of German descent (19). Although no haplotype analysis was performed, it was a complex rearrangement with identical breakpoints.

The relatively low number of partial gene rearrangements detected as a founder mutation is probably due, at least partly, to the use of sequencing as the predominant method for mutation detection, and it is likely that the number of such cases will increase as more genes are screened with quantitative methods.

It has already been the suggested that exonic deletions may be present in γ-sarcoglycanopathy (20), particularly since many cases were reported to be homozygous. If parental DNA is not available it is difficult to distinguish with certainty whether the mutation is truly homozygous or

is a compound heterozygote, with the loss of the second allele masking the true situation. In addition, when parental DNA is unavailable, the fact that one of either parent does not carry the mutation may be erroneously explained as either non-paternity or a de novo occurrence of the mutation. An associated danger in the latter situation is that the de novo change suggested is taken as proof that this change has pathogenic consequences.

As exonic deletions have been shown to occur in at least some sarcoglycanopathy patients it suggests that screening for such mutations should be routinely implemented in a diagnostic setting. Our report shows that such analysis is highly recommended in cases where a mutation can be found in one allele only, or when segregation from the parents can not be confirmed. An accurate diagnosis is not only important for the patient but also for subsequent counseling of family members.

Acknowledgements

We would like to thank the patients and their families for their cooperation, Annemieke Aartsma-Rus for assistance with the long-range PCR, and Peter de Knijff for providing the CEPH samples.

Reference List

1. Bushby,K.M. (1999) Making sense of the limb-girdle muscular dystrophies. *Brain*, **122 (Pt 8)**, 1403-1420.

2. Matsumura,K., Campbell,K.P. (1994) Dystrophin-glycoprotein complex: its role in the molecular pathogenesis of muscular dystrophies. *Muscle Nerve*, **17**, 2-15.

3. Cohn,R.D., Campbell,K.P. (2000) Molecular basis of muscular dystrophies. *Muscle Nerve*, **23**, 1456-1471.

4. Armour,J.A., Barton,D.E., Cockburn,D.J., and Taylor,G.R. (2002) The detection of large deletions or duplications in genomic DNA. *Hum.Mutat.*, **20**, 325-337.

5. Den Dunnen,J.T., Grootscholten,P.M., Bakker,E., Blonden,L.A.J., Ginjaar,H.B., Wapenaar,M.C., Van Paassen,H.M.B., Van Broeckhoven,C., Pearson,P.L., and Van Ommen,G.J.B. (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.

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

7. Petrij,F., Dauwerse,H.G., Blough,R.I., Giles,R.H., van der Smagt,J.J., Wallerstein,R., Maaswinkel-Mooy,P.D., van Karnebeek,C.D., van Ommen,G.J., Van Haeringen,A., Rubinstein,J.H., Saal,H.M., Hennekam,R.C., Peters,D.J., and Breuning,M.H. (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.

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

9. Ginjaar,H.B., Van Der Kooi,A.J., Ceelie,H., Kneppers,A.L., Van Meegen,M., Barth,P.G., Busch,H.F., Wokke,J.H., Anderson,L.V., Bonnemann,C.G., Jeanpierre,M., Bolhuis,P.A., Moorman,A.F., De Visser,M., Bakker,E., and Ommen,G.J. (2000) Sarcoglycanopathies in Dutch patients with autosomal recessive limb girdle muscular dystrophy. *J.Neurol.*, **247**, 524-529.

10. White,S., Kalf,M., Liu,Q., Villerius,M., Engelsma,D., Kriek,M., Vollebregt,E., Bakker,B., van Ommen,G.J., Breuning,M.H., and Den Dunnen,J.T. (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.

11. Rozen,S., Skaletsky,H. (2000) In Krawetz S, Misener S (eds), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 365-386.

12. Politano,L., Nigro,V., Passamano,L., Petretta,V., Esposito,M., Papparella,S., Ricci,E., Giugliano,M., Servidei,S., Comi,L., and Nigro,G. (1998) Clinical and genetic findings in sarcoglycanopathies. *Acta Myologica*, **2**, 33-40.

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

14. Shaw,C.J., Lupski,J.R. (2004) Implications of human genome architecture for rearrangement based disorders: the genomic basis of disease. *Hum.Mol.Genet.*.

15. Noguchi,S., McNally,E.M., Ben Othmane,K., Hagiwara,Y., Mizuno,Y., Yoshida,M., Yamamoto,H., Bonnemann,C.G., Gussoni,E., Denton,P.H., Kyriakides,T., Middleton,L., Hentati,F., Ben Hamida,M., Nonaka,I., Vance,J.M., Kunkel,L.M., and Ozawa,E. (1995) Mutations in the dystrophin-associated protein gamma-sarcoglycan in chromosome 13 muscular dystrophy. *Science*, **270**, 819-822.

16. McNally,E.M., Passos-Bueno,M.R., Bonnemann,C.G., Vainzof,M., Moreira,E.S., Lidov,H.G.W., Ben Othmane,K., Denton,P.H., Vance,J.M., Zatz,M., and Kunkel,L.M. (1996) Mild and severe muscular dystrophy caused by a single gamma-sarcoglycan mutation. *Am.J.Hum.Genet.*, **59**, 1040-1047.

17. Fanin,M., Hoffman,E.P., Angelini,C., and Pegoraro,E. (2000) Private beta- and gamma-sarcoglycan gene mutations: evidence of a founder effect in Northern Italy. *Hum.Mutat.*, **16**, 13-17.

18. Petrij-Bosch,A., Peelen,T., Van Vliet,M., Van Eijk,R., Olmer,R., Drusedau,M., Hogervorst,F.B.L., Hageman,S., Arts,P.J., Ligtenberg,M.J., Meijers-Heijboer,H., Klijn,J.G., Vasen,H.F., Cornelisse,C.J., Van 't Veer,L.J., Bakker,E., Van Ommen,G.J.B., and Devilee,P. (1997) BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. *Nat.Genet.*, **17**, 341-345.

19. Schmidt,C., Vester,U., Wagner,C.A., Lahme,S., Hesse,A., Hoyer,P., Lang,F., Zerres,K., and Eggermann,T. (2003) Significant contribution of genomic rearrangements in SLC3A1 and SLC7A9 to the etiology of cystinuria. *Kidney Int.*, **64**, 1564-1572.

20. Bonnemann,C.G., Wong,J., Jones,K.J., Lidov,H.G., Feener,C.A., Shapiro,F., Darras,B.T., Kunkel,L.M., and North,K.N. (2002) Primary gamma-sarcoglycanopathy (LGMD 2C): broadening of the mutational spectrum guided by the immunohistochemical profile. *Neuromuscul.Disord.*, **12**, 273-280.

21. Nowak,K.J., Walsh,P., Jacob,R.L., Johnsen,R.D., Peverall,J., McNally,E.M., Wilton,S.D., Kakulas,B.A., and Laing,N.G. (2000) Severe gamma-sarcoglycanopathy caused by a novel missense mutation and a large deletion. *Neuromuscul.Disord.*, **10**, 100-107.

22. Takano,A., Bonnemann,C.G., Honda,H., Sakai,M., Feener,C.A., Kunkel,L.M., and Sobue,G. (2000) Intrafamilial phenotypic variation in limb-girdle muscular dystrophy type 2C with compound heterozygous mutations. *Muscle Nerve*, **23**, 807-810.

Chapter 4

Kriek M., White S.J., Bouma M.C. Dauwerse H.G., Hansson K.B., Nijhuis J.V., Bakker B., van Ommen G.J., den Dunnen J.T., Breuning, M.H. (2004). Genomic imbalances in mental retardation. *J.Med.Genet.* 41 (4):249-255.

(Colour images from this chapter can be seen in the appendix)

Genomic imbalances in mental retardation

M Kriek, S J White, M C Bouma, H G Dauwerse, K B M Hansson, J V Nijhuis, B Bakker, G-J B van Ommen, J T den Dunnen, M H Breuning

J Med Genet 2004;**41**:249–255. doi: 10.1136/jmg.2003.014308

See end of article for authors' affiliations

Correspondence to: Professor Dr M H Breuning, Leiden University Medical Center, Center for Human and Clinical Genetics, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands; m.h.breuning@lumc.nl

Received 8 September 2003
Accepted for publication 2 November 2003

Introduction: It has been estimated that cytogenetically visible rearrangements are present in ~1% of newborns. These chromosomal changes can cause a wide range of deleterious developmental effects, including mental retardation (MR). It is assumed that many other cases exist where the cause is a submicroscopic deletion or duplication. To facilitate the detection of such cases, different techniques have been developed, which have differing efficiency as to the number of loci and patients that can be tested.
Methods: We implemented multiplex amplifiable probe hybridisation (MAPH) to test areas known to be rearranged in MR patients (for example, subtelomeric/pericentromeric regions and those affected in microdeletion syndromes) and to look for new regions that might be related to MR.
Results: In this study, over 30 000 screens for duplications and deletions were carried out; 162 different loci tested in each of 188 developmentally delayed patients. The analysis resulted in the detection of 19 rearrangements, of which ~65% would not have been detected by conventional cytogenetic analysis. A significant fraction (46%) of the rearrangements found were interstitial, despite the fact that only a limited number of these loci have so far been tested.
Discussion: Our results strengthen the arguments for whole genome screening within this population, as it can be assumed that many more interstitial rearrangements would be detected. The strengths of MAPH for this analysis are the simplicity, the high throughput potential, and the high resolution of analysis. This combination should help in the future identification of the specific genes that are responsible for MR.

The evolution of the human genome has resulted in a mixture of large and small interspersed and tandem segmental duplications throughout the genome. Such duplications provide substrates for homologous recombination, and consequently, the intervening regions show a considerable rate of rearrangement.[1–3] Many of these rearrangements occur in regions where a change in gene dosage does not affect human health. However, after the description by Lejeune of trisomy 21 in Down's syndrome,[4] and the many subsequent publications on different aneuploidies, it became clear that the genome contains many loci for which the correct copy number is critical for normal development. Change in genetic dosage of one or more genes is one of the most common causes of mental retardation (MR). Examples of known important loci include the subtelomeric regions and the areas involved in microdeletion syndromes.

The subtelomeric regions, localised proximal to the telomeres, have been found to be especially susceptible to copy number changes, owing to repeat rich sequences that show a high frequency of recombination.[1] It has been hypothesised that about 6% of the patients with idiopathic MR will have a subtelomeric rearrangement,[5] a figure confirmed in several studies that have reported a frequency of 2–9% of cryptic rearrangements in MR patients.[6 7]

The cause for MR is only established in approximately 50% of cases, limiting the efficiency of genetic counselling, detection of carriers, and prenatal diagnosis in these families. This rather low percentage of diagnosis may have several explanations. A routine cytogenetic analysis gives a minimum resolution of only 4–10 Mb. Fluorescent in situ hybridisation (FISH) largely overcomes this limitation of resolution; however, it can only be applied to simultaneously test a limited number of chromosome regions. FISH is therefore mostly used to confirm well recognised microdeletion syndromes in patients who present a suggestive phenotype. Another potential explanation is that the genome contains undiscovered loci that are involved in the aetiology of MR. New technologies, such as multiplex amplifiable probe hybridisation (MAPH),[8] multiplex ligation dependent probe amplification (MLPA),[9] and array based comparative genomic hybridisation (array CGH),[10] have recently been developed to search for such undiscovered regions. We chose to implement a high resolution, high throughput, rapid, and simple method, MAPH,[8] which allows the simultaneous screening at the exon level for copy number changes of 40–50 different chromosomal loci in up to 96 patients in one assay.

Hollox *et al*[11] previously described subtelomeric screening using MAPH of patients with a developmental delay. In our study, we screened loci known to be involved in MR (subtelomeric/pericentromeric regions and genes involved in microdeletion syndromes) as well as interstitial genes randomly spaced throughout the genome. A total of 30 000 gene dosage screens were performed from 188 cases with unexplained developmental delay that were each scanned for copy number changes at 162 loci. We were able to detect subtelomeric, pericentromeric, and interstitial rearrangements in a group of patients with MR and dysmorphic features and/or multiple congenital abnormalities, as well as in patients selected solely on the basis of developmental delay.

SUBJECTS AND METHODS

Probe design and MAPH

The probe design has been previously described,[12] using unique sequences only. The primers of the chosen sequences

Abbreviations: BAC, bacterial artificial chromosome; CGH, comparative genomic hybridisation; FISH, fluorescent in situ hybridisation; MAPH, multiplex amplifiable probe hybridisation; MCA, multiple congenital abnormalities; MLPA, multiplex ligation dependent probe amplification; MR, mental retardation; SMS, Smith-Magenis syndrome

were designed using Prophet (http://www.basic.nwu.edu/biotools/prophet.html), and supplied by Invitrogen Life Technologies. Products were amplified from genomic DNA by PCR and cloned into the pGEM-T easy vector (Promega). The correct insert was confirmed by sequencing with the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) at the Leiden Genome Technology Center, using an ABI 3700 Sequencer (Applied Biosystems).

MAPH was performed as described by White *et al*[12] (see also Leiden Muscular Dystrophy Pages (http://www.dmd.nl/DMD_MAPH.html)).

Study population

The DNA of 188 patients (110 males and 78 females) from the Center for Human and Clinical Genetics Leiden (a DNA diagnostic laboratory) was analysed. The patients had been seen by a clinical geneticist or a paediatrician and diagnosed with developmental delay. The study population was divided into two groups. The first group contained 123 coded patients who had been referred for fragile X screening. Before testing, information about the results of additional tests, such as karyotyping, was not known to the investigators. The second study group (n = 65) was known to have a normal karyotype and had tested negative for fragile X screening. All patients had (multiple) congenital malformations or dysmorphic features in addition to psychological developmental delay.

Data analysis

The data were analysed with GeneScan Analysis and Genotyper Software (Applied Biosystems). These programs provide information about the length, peak height, and peak area of the DNA fragments. Peaks were not used for analysis if they were outside predefined thresholds (upper and lower limits of 12 000 and 150 units, respectively). To obtain a ratio, the height of a given peak was divided by the sum of the heights of the four nearest peaks. As it is not likely that all four probes from diverse regions of the genome are altered in one patient, adding unrelated standards was not necessary in most of the probe sets. For the chromosome 22 probe set, however, unrelated probes, containing sequences from other chromosomes, were used as references.

The median ratio for each probe within a single hybridisation (minimum number of samples 8; maximum number 12) was determined and used to calculate a normalised ratio for each patient. Within each patient, initial ''normal'' thresholds were set as 0.75 and 1.25. The standard deviation from the ratios within these limits was calculated, and three times this standard deviation was used as the threshold for any given patient. Any probe that was outside these limits was retested, and samples that showed an apparent copy number change in duplicate were examined further using other techniques. Samples that showed a standard deviation of >10% over probes within the normal thresholds were retested.

Verifying the MAPH results

Copy number changes detected by MAPH were verified using another technique, primarily FISH with a bacterial artificial chromosome (BAC) or cosmid probe covering the appropriate genomic region. The BACs used were designed by Flint,[13] or supplied by Vysis Abbott Laboratories (TV, Telvysion, LSI, locus specific identifiers) or selected from the RPCI human BAC library. The FISH experiments were performed following standard operating procedures as described in Dauwerse *et al*.[14] Some MAPH results were verified using MLPA.[9]

RESULTS

Genotyping

We designed several probe sets covering both the subtelomeric/pericentromeric and interstitial regions, including genes involved in microdeletion syndromes, genes on chromosome 22, and genes spread across all chromosomes (table A, supplemental). The subtelomeric probe set is composed of probes corresponding to the 41 subtelomeric regions, preferably an exon of a gene within 1 Mb from the telomere, five genes near the centromere on the q arm of the acrocentric chromosomes, a sequence in the pseudoautosomal region of chromosome Xq and Yq, and an exon of a Yp specific gene. The microdeletion probe set was made up of 27 probes from 21 different genes involved in microdeletion syndromes (Williams, Prader Willi, Angelman, Smith-Magenis, Sotos, 22q11, Alagille, and Wolf-Hirschhorn syndromes). The chromosome 22 probe set included 19 probes from genes on chromosome 22 with approximately 1 Mb spacing. Finally, we used two probe sets containing a total of 68 interstitial genes spread throughout the genome.

We applied these probe sets following two methods of validation. Firstly, a probe was considered to be reliable when the standard deviation over 12 unaffected samples (one hybridisation) was <15%. Secondly, where possible, we verified the unique and correct localisation of the probes using DNA from patients with known aberrations (42% of the subtelomeric probes, 70% of the microdeletion probes).

Overall, 188 patients were screened for deletions and duplications at 162 loci, resulting in the detection of 19 copy number changes. Of these, four aberrations turned out to be cytogenetically visible, namely an isochromosome 18p (karyotype 47, XY, +i(18p)), a marker chromosome (karyotype 47, XY, +mar.ish der(22)t(8;22)(q24.1;q11.2)), a triple X female (karyotype 47, XXX) and a Turner syndrome (karyotype 45, X), because the outcome of additional investigations had not been made known to the investigators before testing. These patients and their corresponding aberrations were not included in the calculation of the percentage of rearrangements found by MAPH; however, they emphasise the usefulness of MAPH for detecting copy number changes.

In total, eight subtelomeric/pericentromeric rearrangements were found (table 1; upper part). Five of these mutations were detected in the group of MR patients with additional dysmorphic features or additional congenital malformations (5/65 = 7.7%) and the remaining three subtelomeric aneusomies were diagnosed in the group selected on the basis of developmental delay only (3/123 = 2.4%). The smallest mutation found was a deletion of 110 kb maximum present in chromosome band 7p22.3 (table 1, F; and data not shown). Seven rearrangements were interstitial mutations. These are summarised in the lower part of table 1. Where possible, the DNA of both parents of these patients was tested; 75% (9/12) were shown to be de novo. The duplication of 14q11.2 (table 1, O) and the 7ptel deletion (table 1, F) were also found in the parental DNA, and one of the parents of patient E was a balanced translocation carrier.

As the number of cytogenetically detectable aberrations is highly dependent on the banding resolution, the karyograms of all 15 patients with a MAPH detected rearrangement were re-examined. At a resolution of 500–550 bands per haploid set, the karyograms showed that two subtelomeric copy number changes should have been detected cytogenetically (table 1; A, C). The detection of a 1ptel deletion (table 1, H) was doubtful; however, the duplication of 1ptel (table 1, H) was picked up. This implies that although the presence of the copy number change was known, 63% (12/19) of these genomic changes found in this study were cytogenetically

Table 1 An overview of all 15 patients (A–O) with MAPH detected subtelomeric/pericentromeric and interstitial aneusomies. After the verification of these imbalances by FISH or MLPA, the karyograms of the patients were re-examined at a resolution of 500–550 bands. The results obtained are shown in the column 'cytogenically visible'. The clinical features known to be related to the rearrangement found by MAPH are highlighted. The presence or absence of a genotype–phenotype correlation is summarised under ''Pathogenic''.

Case		Aneusomy	Group	Gender	Confirmed by	Cytogenetically visible	Clinical features	Pathogenic	References
Subtelomeric/Pericentromeric									
A	1	Deletion 18q22.1	MR++‡	Female	FISH¶ clone ID:TV 18q	Yes: 500–550 bands	MR, small stature, hearing loss, TAPVR,†† mild facial dysmorphism, tapering fingers	Yes	Many: latest are[15 33]
B	2	Deletion 16p13.3	DD only§	Male	FISH clone ID;COS15A	No	Moderate MR, mild facial dysmorphism, mild alpha thalassemia	Yes	Many: latest is[17]
C		Deletion 6p25	DD only	Male	FISH clone ID: TV 6p	Yes: 500–550 bands	Moderate MR, iris dysplasia, excentric pupil, hypertelorism, hearing loss	Yes	[34 35]
D		Deletion pericentromeric region of chr. 22, duplication of 22q11.2	MR++	Male	FISH clone ID: RP11_3018K1	No	Mild MR, hearing loss, palatoschisis, cataract, microcephaly, double set of teeth	?	Kriek *et al*†
E		Deletion 6qtel, duplication 20qtel	MR++	Male	FISH clone ID:57H24 (6q), 81F12 (20q) MLPA**	No	MR, hypotonicity, microcephaly, brain anomalies, mild facial dysmorphism.	*	
F		Deletion 7ptel	DD only	Male	MLPA	No	Mild developmental delay in early childhood, mild facial dysmorphism	No/?	[36]
G		Duplication 1ptel	MR++	Female	FISH clone ID: 785P20, 37J18	Yes: 500–550 bands	Psychomotor developmental delay, double sided ptosis, parasis of VI cranial nerve, strabismus	?	[37]
H		Deletion 1ptel	MR++	Female	FISH clone ID: 465B22, 37J18	Doubtful: 500–550 bands	Psychomotor developmental delay, dysmorphic features, hirsutism, epilepsy	Yes	[38]
Interstitial									
I		Duplication 17p11.2	MR++	Female	FISH clone ID: LSI-SMS	No	MR, microcephaly, retrognathia, tapering acra, hypertelorism, synophrys, epilepsy	?	[39]
J	3	Deletion 17p11.2	DD only	Male	FISH clone ID: LSI-SMS, MLPA	No	Psychomotor developmental delay (speech delay), infantile hypotonicity, tent shaped mouth	Yes	Many: latest is[40]
K	4	Deletion 4q34.1	DD only	Male	FISH clone ID: RP11-475B2	No	Mild learning disability, short stature, severe delay of bone maturation, aberrant hand shape	Yes	[19]
L	5	Duplication 20p12.2	DD only	Male	MLPA	No	Mild MR, psychiatric disorder	?	[21]
M	6	Duplication 22q11.2	MR++	Female	FISH clone ID: LSI TUPLE1	No	Severe psychomotor retardation, short stature, microcephaly, facial dysmorphism, epilepsy, brain anomalies, renal aplasia	?	[41 42]
N		Deletion 22q11.2	MR++	Female	FISH clone ID LSI TUPLE1	No	Developmental delay, tetralogy of Fallot, absent pulmonary valve, respiratory complications	Yes	Many: latest is[43]
O		Duplication 14q11.2	DD only	Male	MLPA	No	MR, mild facial dysmorphism, short hands and feet, shawl scrotum	No/?	

*The rearrangement is probably causative, as a sibling with a similar phenotype has the same aberration.
†Manuscript in preparation.
Group of patients ‡with mental retardation and additional features, §selected solely on the basis of developmental delay.
¶Fluorescent in situ hybridisation, **multiplex ligation dependent probe amplification, ††total anomalous pulmonary venous return.
No/?: one of the parents also has the aberration; however, imprinting, variable expression and low penetrance have not been excluded; TAPVR, total anomalous pulmonary venous return.
Cases 1–6 are described in more detail in the text.

undetectable using karyotyping at a resolution of 500–550 bands.

Case descriptions

Case 1

This 15 year old girl was diagnosed with total anomalous pulmonary venous return, hearing loss in combination with a narrow external auditory meatus, and MR. Physical examination at the age of 14 years showed a short stature (−3 SD) and some facial dysmorphic features (small palpebrae, broad mouth, thin upper lip). Karyotyping at a resolution of 400 bands and FISH studies of the 22q11 region did not detect any rearrangements. MAPH study showed a de novo deletion of the subtelomeric region of 18q, which was confirmed by FISH using probe TV18q. The clinical features of this patient are consistent with those of the 18q syndrome phenotype.[15]

Case 2

A male patient, who had previously tested negative for Williams syndrome, was diagnosed with a de novo deletion of 16ptel by MAPH. FISH analysis confirmed this finding and limited the proximal breakpoint to chromosome band 16p13.3, distal to the *PKD1–TSC2* (LocusLink 5310–7249) gene cluster[16] using probe COS15A. As expected, owing to the location of the alphaglobin gene (*HBA1*; LocusLink 3039) in this region (16p13.3),[17] further investigation showed that this patient had mild anaemia (alpha thalassaemia heterozygosity) in addition to his moderate mental handicap and dysmorphic features.

Case 3

This boy was seen by a clinical specialist at the age of 2.5 years for his psychomotor retardation and joint hyperflexibility. Physical examination showed few dysmorphic features (a tent shaped mouth), hypotonia, and hypermobility. MAPH analysis revealed a de novo deletion within chromosome band 17p11.2 corresponding to the Smith-Magenis syndrome (SMS) region, using a probe for the *DRG2* gene (LocusLink: 1819). The more distally located *COPS3* gene (LocusLink: 8533) showed two copies (fig 1a). Additional MLPA testing showed that the *RAI1* gene (LocusLink: 10743) was also deleted in this patient (fig 1b), and FISH analysis (probe LSI-SMS) verified the deletion of part of chromosome band 17p11.2 (fig 1c). Recently, three dominant frameshift mutations in *RAI1* have been identified in three patients with phenotypic characteristics of SMS but no cytogenetically detectable deletion of chromosome band 17p11.2.[18] The authors argue that mutations in *RAI1* are responsible for most of the characteristic features of SMS and that further variation is caused by hemizygosity of the other genes in the chromosome region.

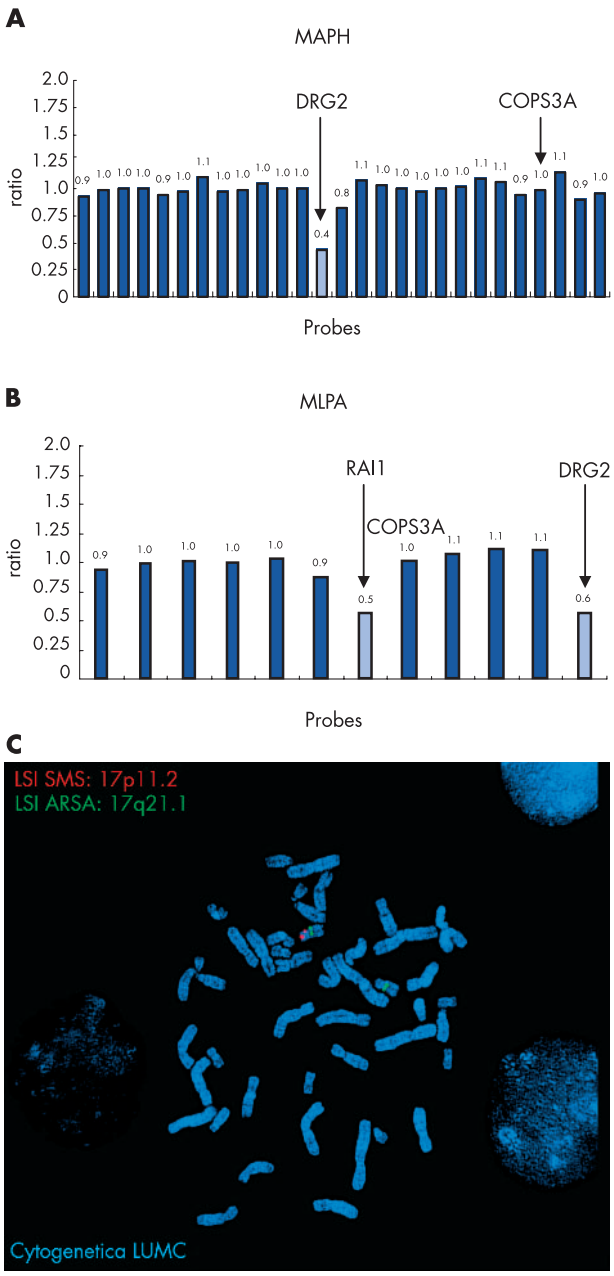

Figure 1 The plots correspond to the MAPH results showing (A) a deletion of the *DRG2* gene, two normal copies of *COPS3A* (*RAI1* not present), and the MLPA results; and (B) a deletion of *RAI1*, a deletion of *DRG2*, and a normal ratio of *COPS3A*. (C) The additional FISH analysis using the LSI-SMS probe specific for the Smith Magenis chromosomal region shows a normal signal on the short arm of only one copy of chromosome 17.

Case 4

This male patient showed at the age of 12 years a mild learning disability, a low voice, a disproportionally short stature (height −2 SD, span −3 SD for height, sitting height −0.5 SD, head circumference −2 SD), limited elbow extension, a permanently extended, inflexible fifth digit of both hands with a ram's horn shaped nail and hypotrophy of the hypothenar muscles (fig 2), and a short broad great toe on both feet. The hand *x* ray revealed short metacarpals I and V, short distal phalange V, and a delay of bone maturation. In this patient, a de novo deletion of 4q34.1 was detected and confirmed by FISH (probe RP11-475B2). Analysis with a more distally located MAPH probe at chromosome band 4q35.1 showed that this latter region was still present, indicating an interstitial rearrangement. Additional FISH experiments using different BAC probes limited the deletion to a maximum of 3 Mb (data not shown).

Patients with an interstitial 4q deletion have been described with a range of features, depending on the proximal and distal breakpoints of the deletion.[19] As it is known that fifth finger anomalies and short stature are found in patients with an interstitial deletion of 4q including 4q34,[20] as well as in patients with a terminal deletion of 4q, it is possible that the genes responsible for these features are located within this region.

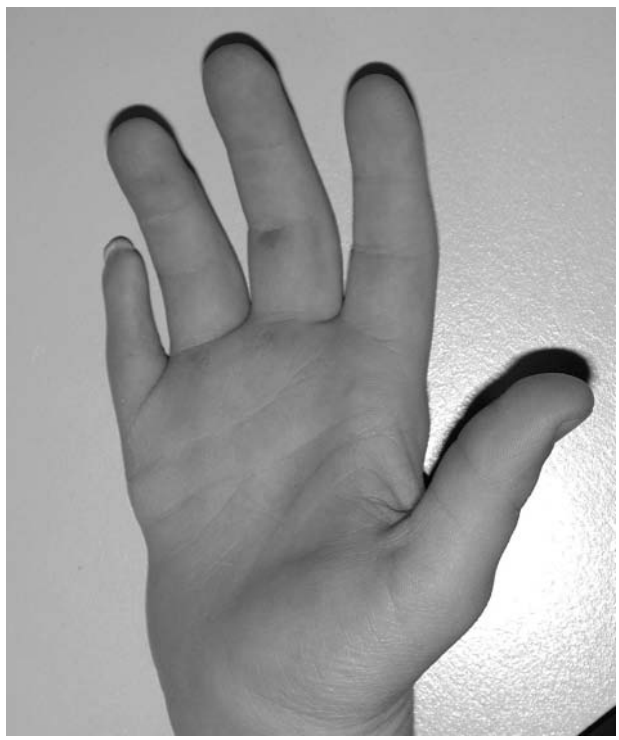

Figure 2 The right hand of case 4 showing a short, inflexible fifth digit with a ram's horn shaped nail and hypotrophy of the hypothenar muscles.

Case 5

This mildly retarded man, with a de novo duplication within chromosome band 20p12.2, containing the *Jagged1* gene (*JAG1*; LocusLin: 182), died at the age of 60 years from multiple myeloma. He had been institutionalised for over 40 years in a psychiatric hospital because of aggressive behaviour, and was diagnosed as schizophrenic. To the best of our knowledge there has been only one previous report[21] of a duplication of 20p11.21–p11.23, in four members of a family with clinical signs of Alagille syndrome. As our patient is not available for further investigation, it remains unclear whether he had such features.

Case 6

After 41 weeks of gestation, this child was born with a birth weight of 1995 g ($\leqslant$2.5 SD) and a head circumference of 28.5 cm ($\leqslant$2.5 SD). At the age of 25 months, her psychomotor development was severely delayed and she suffered from epilepsy. Physical examination showed growth retardation (length $\leqslant$2 SD; weight −6 SD), microcephaly (head circumference −6 SD), hypertonicity, dystonic movements, facial dysmorphisms (ptosis of the left eye, flat philtrum, thin upper lip; fig 3) ear pits, café au lait spots, and absence of the labia minora. Further investigation revealed corpus callosum hypoplasia and deformed gyri, the presence of only one kidney and mildly increased urinary glutaric acid.

Using the microdeletion probe set, a duplication of 22q11.2 was detected by MAPH, and FISH analysis in interphase nuclei confirmed this finding (LSI TUPLE1). The patient's mother did not carry the duplication, and the father was unavailable for testing. We plan to use polymorphic markers to determine the parental origin of the aberrant chromosome 22.

DISCUSSION

Using MAPH analysis, we performed a high resolution duplication/deletion screening of 188 patients with a developmental delay; 162 loci per patient were tested, amounting to over 30 000 typings. The MAPH probes designed for this study can be broadly divided into two groups: (*a*) subtelomeric and pericentromeric probes (n = 48) and (*b*) interstitial probes (n = 114), containing sequences located in regions previously found to be rearranged in mentally retarded individuals, and genes randomly spaced through out the genome.

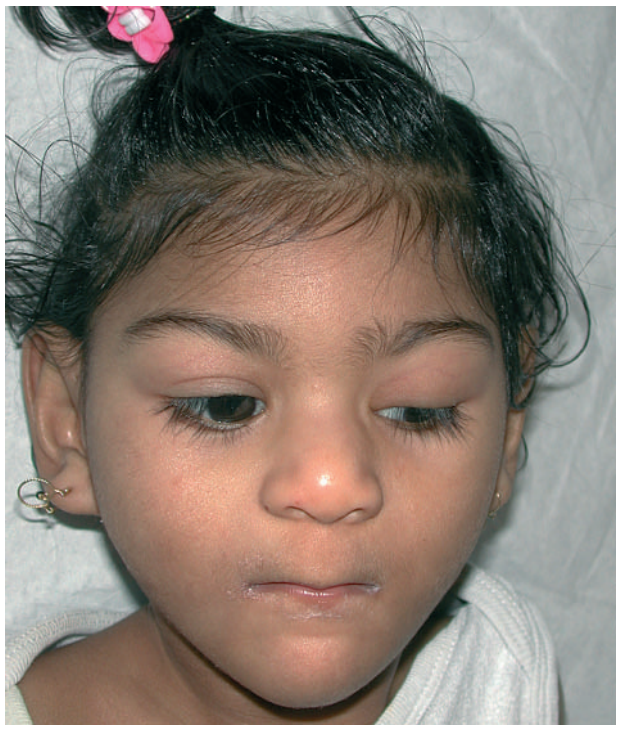

Figure 3 Facial dysmorphism of case 6. Note the microcephaly, ptosis of the left eye, flat philtrum, and thin upper lip.

We detected 4.3% (8/184) subtelomeric/pericentromeric rearrangements (six deletions, one duplication, and one subtelomeric deletion/duplication in one patient), using 48 MAPH probes. A subdivision of subtelomeric aberrations over our two study populations agrees with the findings of Knight *et al*[22] and Yasseen *et al*.[23] The percentage of subtelomeric mutations detected was higher in a group of MR patients with additional malformations (7.7%) than in a group selected on the basis of developmental delay only (2.5%). This supports the suggestion of De Vries *et al* that pre-selection of patients for subtelomeric screening is worthwhile. However, pre-selection of these patients for subtelomeric rearrangements is difficult, as only two clinical features (perinatal onset growth retardation and a positive family history) differed significantly between patients with subtelomeric aneusomies and patients with idiopathic MR.[24] Our overall percentage is similar to that reported in a recent paper that summarised all previous subtelomeric publications.[7] A total of 131 subtelomeric imbalances were found using several different methods among 2582 MR patients, resulting in an overall frequency of 5.1%. A review of the corresponding clinical aspects of these subtelomeric rearrangements has been published recently.[25] After re-examining the karyogram of our patients at a banding resolution of 500–550 bands, it showed that five MAPH detected subtelomeric imbalances were not cytogenetically visible, despite the knowledge of a copy number change present. This means that the percentage of ''true'' submicroscopic subtelomeric/pericentromeric findings is ~3% (5/184) in this study.

Previous reports by Sismani *et al*[26] and Hollox *et al*[11] had already shown the ability of MAPH to detect subtelomeric copy number changes. Hollox *et al* found a copy number change in 5 of 37 male patients (13.5%) who had been referred for fragile X screening. The higher percentage of mutations found by this group may be due to differences in selection criteria for fragile X screening.

We also screened the subtelomeric/pericentromeric regions in eight newborns suffering multiple congenital abnormalities (MCA). Among these patients, one deletion of the subtelomeric region of chromosome 15 was detected and subsequently confirmed by FISH (data not shown).[27] To determine whether it is worthwhile to test this group for submicroscopic mutations, more newborns with MCA should be examined. The ease and relatively low cost of the MAPH technique means that such analysis is feasible. Moreover, new techniques such as MAPH/MLPA and array CGH provide the possibility of genetic diagnosis at a younger age. As the suggestive phenotype for some microdeletion syndromes emerge only later in life, this diagnosis would be very important for providing appropriate healthcare.

In addition to the reports published by Sismani *et al*[25] and Hollox *et al*,[11] we also examined interstitially localised genes, including genes involved in several microdeletion syndromes, genes on chromosome 22 (as this was the first chromosome to be completely sequenced), and genes that are spread throughout the genome and might be involved in cognitive development. Recently, Bailey *et al*[3] argued that regions between highly similar duplications (low copy repeats) are prone to recombination and consequently, copy number changes occur at a higher frequency in these regions compared with other loci in the genome. Several of the areas

described were also tested in this study, mostly corresponding to chromosomal regions involved in microdeletion syndromes. In total, seven interstitial deletions and duplications were detected, of which five were diagnosed in three different regions known to be involved in the microdeletion syndromes and flanked by segmental duplications. Three of these interstitial rearrangements detected include duplications of regions that are usually deleted (the chromosome regions of Smith Magenis (17p11.2), DiGeorge (22q11.2), and Alagille syndromes (20p12.2)). This observation supports the theory that the regions between low copy repeats can both be deleted and duplicated, and implies that the number of patients suffering from a microduplication syndrome is currently probably underestimated. The phenotype (if any) of a microduplication syndrome might, however, be less severe, and under standard diagnostic conditions, the detection of duplications is more problematical. It should be noted that in the second study group, the cases with a distinctive phenotype for a specific microdeletion syndrome were not included.

As has been the case during the development of every new technique, the genomic variations detected can be divided into the following subclasses: (*a*) genetic changes that are clearly pathogenic, (*b*) rearrangements that may or may not be causal to the patient's problem, and (*c*) polymorphic changes. In some cases, extensive clinical studies will be needed to determine to which category a newly detected aberration belongs. In two of our cases, we could detect the rearrangement in one of the parents (the duplication of chromosome band 14q11.2 and the deletion of 7ptel on chromosome band 7p22.3). One explanation is that these imbalances are polymorphic, and that the phenotype of the patient is not related to the copy number change. However, other explanations are possible: (*a*) the affected region is imprinted, and the parental origin of this region is critical in causing the deleterious phenotype;[28] (*b*) allelic variation in the expression of the genes may influence the phenotype;[29] and (*c*) low penetrance of the rearrangement—that is, a genetic defect does not always lead to a phenotypic effect. The detection of such rearrangements will increase as high resolution techniques are applied, and this will pose new problems for genetic counselling. Therefore, it is important to map these familial imbalances in further detail to allow a genotype–phenotype correlation in larger populations of individuals with the same copy number change. In this way, the understanding of any clinical consequence of such a rearrangement should be improved.

Based on previous publications, seven rearrangements found in this study were considered to be pathogenic (table 1). In the remaining cases, the data available in literature were insufficient to support a conclusion that the aneusomy detected is related to the phenotype of the patient. It should be noted that the fact that a rearrangement is de novo is not in itself proof that it is causally related to the deleterious phenotype.

Several different methologies have been described to identify changes using MAPH and MLPA. These include visual comparison of traces from controls and patients,[30] the setting of arbitrary thresholds,[26] and bivariate analysis.[11] We observed that the standard deviations for each probe varied slightly between hybridisations, and could be normalised only within a single hybridisation. The standard deviation of ''normal'' probes within each patient was calculated, with 3 times this figure defining the threshold for a potential rearrangement, thus minimising the effect of any genuine copy number changes on the analysis. As false negative results are, by definition, mutations that were not detected, it is difficult to determine the percentage. To gain an estimate as to the actual false negative rate, we looked at a number of samples where a mutation was previously known. We tested 30 samples that had aberrations at loci corresponding to 39 of the probes used. The appropriate copy number changes were detected in all cases. Using the LaPlace formula $p = (x+1)/(n+2)$ to provide a false negative rate from our data yields an expected value of ~2.5%. This figure suggests that the true false negative rate would be, at least for the 39 probes examined, comparable to the 2% theoretically predicted by Hollox *et al*.[11] Of course, it would be desirable to test all the probes on known mutations in the future.

The number of interstitial aneusomies found in this report strengthens the arguments for genomewide screening for copy number changes in developmentally delayed patients. In most clinical laboratories, deletions and duplications are detected by FISH. This usually focuses on only one region per hybridisation, and is therefore relatively slow and expensive. Several new technologies have emerged that facilitate large scale and genomewide screening of deletion and duplication mutations. For genomewide screening, array CGH currently seems to be the most attractive, with recent publications describing screening with approximately 2000 BAC-PAC clones at an average resolution of 1.5 Mb.[31 32] This is impressive, but inherently means that 90% of the genome is not screened. In addition, probes in array CGH are 100–200 kb BAC clones, often covering more than one gene and thus able to pick up large multi-gene deletions/duplications only—that is, those >100 kb, while it is probable that a significant proportion of deletion/duplication mutations are smaller than this. In contrast, it is possible to detect rearrangements of only 100 bp using MAPH and MLPA technology. By applying a high resolution method, however, the percentage of the genome that can be screened using the same number of probes will be much less compared with array CGH. Using MAPH/MLPA, it is not possible to screen the whole genome for copy number changes at this moment, unless a very large number of probes are included. For this reason, a different approach is required. We consider array CGH to be an excellent tool for finding large regions in the genome where genes involved in particular diseases reside. As soon as these areas have been identified, targeted and much cheaper assays can be designed, zooming in on these regions only. For these reasons, we believe that gene specific screening is ultimately more attractive. With that in mind, MAPH/MLPA have an important role in such analyses, as they are able to pick up both large and small deletions/duplications.

ACKNOWLEDGEMENTS

We would like to thank all physicians of the Center for Human and Clinical Genetics Leiden for selecting patients and gathering blood samples, Dr J P Schouten (MRC-Holland, Amsterdam) for providing the MLPA probes, the Leiden Genome Technology Center for technical assistance, Dr P Eilers and E Sterrenburg for giving statistical advice, Dr C Rosenberg for critical reading of the manuscript, Dr E Peeters for her efforts, and the patients and families for their cooperation. M Kriek is funded by ZON-MW (AGIKO fellowship 940–37–032).

Authors' affiliations

M Kriek, S J White, M C Bouma*, H G Dauwerse, K B M Hansson, J V Nijhuis, B Bakker, Gert-J B van Ommen, J T den Dunnen, M H Breuning, Center for Human and Clinical Genetics, Leiden University Medical Center, The Netherlands

*Department of Clinical Genetics, University Hospital Groningen, The Netherlands

The first two authors contributed equally to this work

REFERENCES

1 **Mefford HC**, Trask BJ. The complex structure and dynamic evolution of human subtelomeres. *Nat Rev Genet* 2002;**3**:91–102.

2 **Lupski JR**. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet* 1998:417–22.
3 **Bailey JA**, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. Recent segmental duplications in the human genome. *Science* 2002;**297**:1003–7.
4 **Lejeune J**, Turpin MR, Gautier M. Etude des chromosomes de neuf enfants mongoliens. *Contes Rendus Acad Sci* 1959;**248**:1721–2.
5 **Flint J**, Wilkie AO, Buckle VJ, Winter RM, Holland AJ, McDermid HE. The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nat Genet* 1995;**9**:132–40.
6 **Biesecker LG**. The end of the beginning of chromosome ends. *Am J Med Genet* 2002;**107**:263–6.
7 **Flint J**, Knight S. The use of telomere probes to investigate submicroscopic rearrangements associated with mental retardation. *Curr Opin Genet Dev* 2003;**13**:310–16.
8 **Armour JA**, Sismani C, Patsalis PC, Cross G. Measurement of locus copy number by hybridisation with amplifiable probes. *Nucl Acids Res* 2000;**28**:605–9.
9 **Schouten JP**, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 2002;**30**:e57.
10 **Pinkel D**, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998;**20**:207–11.
11 **Hollox EJ**, Atia T, Cross G, Parkin T, Armour JA. High throughput screening of human subtelomeric DNA for copy number changes using multiplex amplifiable probe hybridisation (MAPH). *J Med Genet* 2002;**39**:790–5.
12 **White S**, Kalf M, Liu Q, Villerius M, Engelsma D, Kriek M, Vollebregt E, Bakker B, van Ommen GJ, Breuning MH, Den Dunnen JT. Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. *Am J Hum Genet* 2002;**71**:365–74.
13 **Flint J**, Wilkie AO, Buckle VJ, Winter RM, Holland AJ, McDermid HE. The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nat Genet* 1995;**9**:132–40.
14 **Dauwerse JG**, Jumelet EA, Wessels JW, Saris JJ, Hagemeijer A, Beverstock GC, van Ommen GJ, Breuning MH. Extensive cross-homology between the long and the short arm of chromosome 16 may explain leukemic inversions and translocations. *Blood* 1992;**79**:1299–304.
15 **Cody JD**, Ghidoni PD, DuPont BR, Hale DE, Hilsenbeck SG, Stratton RF, Hoffman DS, Muller S, Schaub RL, Leach RJ, Kaye CI. Congenital anomalies and anthropometry of 42 individuals with deletions of chromosome 18q. *Am J Med Genet* 1999;**85**:455–62.
16 The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. The European Polycystic Kidney Disease Consortium. *Cell* 1994;**77**:881–94.
17 **Wilkie AO**, Buckle VJ, Harris PC, Lamb J, Barton NJ, Reeders ST, Lindenbaum RH, Nicholls RD, Barrow M, Bethlenfalvay NC. Clinical features and molecular analysis of the alpha thalassemia/mental retardation syndromes. I. Cases due to deletions involving chromosome band 16p13.3. *Am J Hum Genet* 1990;**46**:1112–26.
18 **Slager RE**, Newton TL, Vlangos CN, Finucane B, Elsea SH. Mutations in RAI1 associated with Smith-Magenis syndrome. *Nat Genet* 2003;**33**:466–8.
19 **Lin AE**, Garver KL, Diggans G, Clemens M, Wenger SL, Steele MW, Jones MC, Israel J. Interstitial and terminal deletions of the long arm of chromosome 4: further delineation of phenotypes. *Am J Med Genet* 1988;**31**:533–48.
20 **Tsai CH**, Van Dyke DL, Feldman GL. Child with velocardiofacial syndrome and del (4)(q34.2): another critical region associated with a velocardiofacial syndrome-like phenotype *Am J Med Genet* 1999;**82**:336–9.
21 **Moog U**, Engelen J, Albrechts J, Hoorntje T, Hendrikse F, Schrander-Stumpel C. Alagille syndrome in a family with duplication 20p11. *Clin Dysmorphol* 1996;**5**:279–88.
22 **Knight SJ**, Regan R, Nicod A, Horsley SW, Kearney L, Homfray T, Winter RM, Bolton P, Flint J. Subtle chromosomal rearrangements in children with unexplained mental retardation. *Lancet* 1999;**354**:1676–81.
23 **Yasseen AA**, Al Musawi TA. Cytogenetics study in severely mentally retarded patients. *Saudi Med J* 2001;**22**:444–9.
24 **De Vries BB**, White SM, Knight SJ, Regan R, Homfray T, Young ID, Super M, McKeown C, Splitt M, Quarrell OW, Trainer AH, Niermeijer MF, Malcolm S, Flint J, Hurst JA, Winter RM. Clinical studies on submicroscopic subtelomeric rearrangements: a checklist. *J Med Genet* 2001;**38**:145–50.
25 **De Vries BB**, Winter R, Schinzel A, Ravenswaaij-Arts C. Telomeres: a diagnosis at the end of the chromosomes. *J Med Genet* 2003;**40**:385–98.
26 **Sismani C**, Armour JA, Flint J, Girgalli C, Regan R, Patsalis PC. 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* 2001;**9**:527–32.
27 **Roback EW**, Barakat AJ, Dev VG, Mbikay M, Chretien M, Butler MG. An infant with deletion of the distal long arm of chromosome 15 (q26.1–qter) and loss of insulin-like growth factor 1 receptor gene. *Am J Med Genet* 1991;**38**:74–9.
28 **Pollack JR**, Iyer VR. Characterizing the physical genome. *Nat Genet* 2002;**32**(Suppl):515–21.
29 **Yan H**, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. Allelic variation in human gene expression. *Science* 2002;**297**:1143.
30 **Duponchel C**, Di Rocco C, Cicardi M, Tosi M. Rapid detection by fluorescent multiplex PCR of exon deletions and duplications in the C1 inhibitor gene of hereditary angioedema patients. *Hum Mutat* 2001;**17**:61–70.
31 **Veltman JA**, Fridlyand J, Pejavar S, Olshen AB, Korkola JE, DeVries S, Carroll P, Kuo WL, Pinkel D, Albertson D, Cordon-Cardo C, Jain AN, Waldman FM. Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. *Cancer Res* 2003;**63**:2872–80.
32 **Snijders AM**, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, Albertson DG. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* 2001;**29**:263–4.
33 **Brkanac Z**, Cody JD, Leach RJ, DuPont BR. Identification of cryptic rearrangements in patients with 18q-deletion syndrome. *Am J Hum Genet* 1998;**62**:1500–6.
34 **Zurcher VL**, Golden WL, Zinn AB. Distal deletion of the short arm of chromosome 6. *Am J Med Genet* 1990;**35**:261–5.
35 **Palmer CG**, Bader P, Slovak ML, Comings DE, Pettenati MJ. Partial deletion of chromosome 6p: delineation of the syndrome. *Am J Med Genet* 1991;**39**:155–60.
36 **Chotai KA**, Brueton LA, van Herwerden L, Garrett C, Hinkel GK, Schinzel A, Mueller RF, Speleman F, Winter RM. Six cases of 7p deletion: clinical, cytogenetic, and molecular studies. *Am J Med Genet* 1994;**51**:270–6.
37 **Heilstedt HA**, Shapira SK, Gregg AR, Shaffer LG. Molecular and clinical characterization of a patient with duplication of 1p36.3 and metopic synostosis. *Clin Genet* 1999;**56**:123–8.
38 **Heilstedt HA**, Ballif BC, Howard LA, Lewis RA, Stal S, Kashork CD, Bacino CA, Shapira SK, Shaffer LG. Physical map of 1p36, placement of breakpoints in monosomy 1p36, and clinical characterization of the syndrome. *Am J Hum Genet* 2003;**72**:1200–12.
39 **Potocki L**, Chen KS, Koeuth T, Killian J, Iannaccone ST, Shapira SK, Kashork CD, Spikes AS, Shaffer LG, Lupski JR. DNA rearrangements on both homologues of chromosome 17 in a mildly delayed individual with a family history of autosomal dominant carpal tunnel syndrome. *Am J Hum Genet* 1999;**64**:471–8.
40 **Vlangos CN**, Yim DK, Elsea SH. Refinement of the Smith-Magenis syndrome critical region to approximately 950 kb and assessment of 17p11.2 deletions. Are all deletions created equally? *Mol Genet Metab* 2003;**79**:134–41.
41 **Edelmann L**, Pandita RK, Spiteri E, Funke B, Goldberg R, Palanisamy N, Chaganti RS, Magenis E, Shprintzen RJ, Morrow BE. A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum Mol Genet* 1999;**8**:1157–67.
42 **Ensenauer RE**, Adeyinka A, Flynn HC, Michels VV, Lindor NM, Dawson DB, Thorland EC, Lorentz CP, Goldstein JL, McDonald MT, Smith WE, Simon-Fayard E, Alexander AA, Kulharya AS, Ketterling RP, Clark RD, Jalal SM. Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients *Am J Hum Genet* 2003;**73**:1027.
43 **Bartsch O**, Nemeckova M, Kocarek E, Wagner A, Puchmajerova A, Poppe M, Ounap K, Goetz P. DiGeorge/velocardiofacial syndrome: FISH studies of chromosomes 22q11 and 10p14, and clinical reports on the proximal 22q11 deletion. *Am J Med Genet* 2003;**117A**:1–5.

Chapter 5.1

White S.J., Vink G.R., Kriek M., Wuyts W., Schouten J.P., Bakker B., Breuning M.H., den Dunnen J.T. (2004). Two-colour MLPA; detecting genomic rearrangements in hereditary multiple exostoses. *Hum. Mutat.* 24 (1):86-92.

(Colour images from this chapter can be seen in the appendix)

Two-Color Multiplex Ligation-Dependent Probe Amplification: Detecting Genomic Rearrangements in Hereditary Multiple Exostoses

Stefan J. White,[1]* Geraldine R. Vink,[1] Marjolein Kriek,[1] Wim Wuyts,[2] Jan Schouten,[3] Bert Bakker,[1] Martijn H. Breuning,[1] and Johan T. den Dunnen[1]

[1]Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; [2]Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; [3]MRC-Holland, Amsterdam, The Netherlands

Communicated by Graham R. Taylor

Genomic deletions and duplications play an important role in the etiology of human disease. Versatile tests are required to detect these rearrangements, both in research and diagnostic settings. Multiplex ligation-dependent probe amplification (MLPA) is such a technique, allowing the rapid and precise quantification of up to 40 sequences within a nucleic acid sample using a one-tube assay. Current MLPA probe design, however, involves time-consuming and costly steps for probe generation. To bypass these limitations we set out to use chemically synthesized oligonucleotide probes only. The inherent limitations of this approach are related to oligonucleotide length, and thus the number of probes that can be combined in one assay is also limited. This problem was tackled by designing a two-color assay, combining two sets of probes, each amplified by primers labeled with a different fluorophore. In this way we successfully combined 28 probes in a single reaction. The assay designed was used to screen for the presence of deletions and duplications in patients with hereditary multiple exostoses (HME). Screening 18 patients without detectable point mutations in the EXT1 and EXT2 genes revealed five cases with deletions of one or more exons: four in EXT1 and one in EXT2. Our results show that a two-color MLPA assay using only synthetic oligonucleotides provides an attractive alternative for probe design. The approach is especially suited for cases in which the number of patients to be tested is limited, making it financially unattractive to invest in cloning. Hum Mutat 24:86–92, 2004.

KEY WORDS: MLPA; EXT1; EXT2; hereditary multiple exostoses; HME; mutation detection

DATABASES:
EXT1 – OMIM: 608177; GenBank: NM_000127.1
EXT2 – OMIM: 608210; GenBank: NM_000401.1

INTRODUCTION

Intragenic rearrangements are a common cause of human disease. As mutation screening is usually based on sequence analysis of PCR-amplified fragments, deletions and duplications of complete exons will be missed unless quantitative methods are applied. Many alternatives have been described [reviewed in Armour et al., 2002]. Southern blotting [Den Dunnen et al., 1989], quantitative multiplex PCR [Yau et al., 1996], and fluorescent in situ hybridization (FISH) [Petrij et al., 2000] have been most commonly used, but all have limitations that hinder routine implementation in a flexible and high-throughput manner.

A quick and simple technique for quantitative analysis has recently been described, termed multiplex ligation-dependent probe amplification (MLPA) [Schouten et al., 2002]. This method is based around the hybridization and ligation of two adjacently-annealing probes. Only if these half-probes are ligated can they serve as a template for PCR amplification. The different probes in a set are designed to have common ends, meaning all can be simultaneously amplified with one primer pair. By using a fluorescently-labeled primer, the resulting products can be separated according to size and quantified. This method, which can be performed in a one-tube format, has been successfully applied to several genes in which deletions and duplications are known to frequently occur [Gille et al., 2002; Hogervorst et al., 2003; Taylor et al., 2003].

A significant drawback of the method is the time-consuming nature of probe production. Following the original protocol [Schouten et al., 2002], the generation

Received 18 December 2003; accepted revised manuscript 3 March 2004.

*Correspondence to: S. J. White, Center for Human and Clinical Genetics, Leiden University Medical Center, Wassenaarseweg 72, 2333AL Leiden, The Netherlands. E-mail: s.white@lumc.nl

Grant sponsor: Leiden University Medical Center.

DOI 10.1002/humu.20054
Published online in Wiley InterScience (www.interscience.wiley.com).

of single-stranded DNA fragments of several hundred nucleotides requires cloning into, and subsequent isolation from, a specifically modified M13 vector.

Theoretically, it is possible to use chemically-synthesized oligonucleotides for both of the half-probes, but length limitations mean that relatively fewer probes can be used within the size range available ($2 \times \sim 40$–60 nt). This can be partially circumvented by making use of the increased resolution of capillary electrophoresis, as well as the ability to use multiple colors for detection. By designing the probes such that two different fluorophores can be used simultaneously, it is possible to combine twice as many probes within a single reaction.

To test the efficacy of this approach we designed probe sets to screen for deletions and duplications in the EXT1 (MIM# 608177) and EXT2 (MIM# 608210) genes, in which mutations cause hereditary multiple exostoses (HME). This is a genetically heterogeneous disorder, characterized by multiple bony outgrowths (osteochondromas) on the ends of the long bones, having an incidence of $\sim$1 out of 50,000. EXT1 is found on chromosome 8q24 [Ahn et al., 1995], and is an 11-exon gene, spanning 250 kb. EXT2 is a smaller gene, composed of 14 exons and covering 110 kb on 11p11.2 [Stickens et al., 1996]. Both genes code for glycosyltransferases, which are involved in heparan sulfate synthesis.

The disease shows a dominant pattern of inheritance, and mutations are found in either EXT1 or EXT2 in 70 to 80% of all cases [Wuyts et al., 1998; Wuyts and Van Hul, 2000]. The mutations found to date have been mostly truncating mutations or missense mutations, most probably leading to loss of EXT function. Linkage analysis has implicated a third region (on chromosome 19p) in this disease [Le Merrer et al., 1994], but no gene has been identified to date.

Several cases remain, however, where no mutations could be found. As mutation screening was performed almost exclusively at the sequence level, quantitative (deletions, duplications), and positional (inversions, translocations) changes will not have been detected. Entire gene deletions have been seen involving EXT1 (in Langer Gideon syndrome [Ludecke et al., 1995]) and EXT2 (in P11pDS [Bartsch et al., 1996; Wuyts et al., 2001]) as part of contiguous gene syndromes, but to date there has only been one suggestion of a partial gene deletion, in EXT2 [Stickens et al., 1996]. Using the two-color MLPA assay, we detected single- and multi-exon deletions in 5 out of 18 HME cases, with exon 1 of EXT1 being deleted in three unrelated cases.

MATERIALS AND METHODS

Patients

The DNA of 18 unrelated HME patients was studied to identify mutations in the EXT1 or EXT2 genes. The entire coding sequence of the EXT1 and EXT2 genes had been previously analyzed with direct sequence analysis, with no mutations being detected. All patients showed multiple osteochondromas, and 11 were known to have no family history of HME.

Probe Design

Probes were designed for each coding exon of EXT1 and EXT2 (Table 1). To allow simultaneous probe amplification, each set of probes were designed to allow amplification with one pair of primers. For each EXT1 probe, the common ends corresponded to the MLPA primers described in Schouten et al. [2002], and the EXT2 probes used the multiplex amplifiable probe hybridization (MAPH) amplification primers described in White et al. [2002]. To ensure specific hybridization, the presence of repetitive sequences was excluded using the BLAT program from the University of California Santa Cruz (UCSC) website (http://genome.ucsc.edu) [Kent, 2002]. Probes within each set were designed to produce PCR products with a minimum separation of 2 bp, with the products ranging in size from 80 to 125 bp. The hybridizing regions of the probes had a Tm of at least 65°C (defined using the RAW program (MRC-Holland, Amsterdam, The Netherlands), with a GC% between 35 and 60%.

Oligonucleotides were ordered from either Sigma Genosys (UK, www.sigma-genosys.com) or Illumina, Inc. (San Diego, CA). The oligonucleotides from Sigma Genosys were desalted without further purification, whereas the oligonucleotides from Illumina were synthesized in a salt-free environment and were unpurified. All oligonucleotides were synthesized at a starting scale of 50 nmol. The downstream oligonucleotide of each pair was 5′ phosphorylated to allow ligation to occur.

Probe mixes were prepared by combining each oligonucleotide so that all were present at a final concentration of 4 fmol/μl. The EXT MLPA mixes are available on request (www.LGTC.nl).

MLPA Reaction

All reagents for the MLPA reaction and subsequent PCR amplification were purchased from MRC-Holland (Amsterdam, The Netherlands), with the exception of the MAPH-F and MAPH-R primers (Sigma Genosys). The MLPA reactions were performed essentially as described in Schouten et al. [2002]. Briefly, 50–200 ng of genomic DNA (concentration determined using a UV spectrophotometer) in a final volume of 5 μl was heated at 98°C for 5 minutes. After cooling to room temperature, 1.5 μl probe mix and 1.5 μl SALSA hybridization buffer were added to each sample, heat denatured at 95°C for 2 minutes, followed by hybridization for 16 hr at 60°C.

Ligation was performed at 54°C by adding 32 μl ligation mix. After 10–15 minutes, the reaction was stopped by heat inactivation at 95°C for 5 minutes.

PCR amplification was carried out for 30–33 cycles in a final volume of either 25 μl or 50 μl. In addition to the reagents described [Schouten et al., 2002], MAPH-F and MAPH-R were added to each PCR reaction to a final concentration of 100–200 nM, with MAPH-F being fluorescently labeled with either HEX or ROX. The MLPA primers were labeled with FAM. From each PCR reaction, 1–2 μl of product was mixed with 10 μl (Hi Di) formamide in a 96 well plate. For reactions performed with HEX labeled MAPH-F, 0.1 μl ROX 500 size standard (Applied Biosystems, www.appliedbiosystems.com) was also added to each well. Product separation was performed using capillary electrophoresis on the ABI 3700 (Applied Biosystems).

Data Analysis

For quantitative analysis, trace data were retrieved using the accompanying software (GeneScan; Applied Biosystems). These data were then exported to Excel (Microsoft; www.microsoft.com) for further calculations. Within each probe set, two probes for unlinked loci were included as a reference, and all calculations were performed within one probe set. The height of each exon-specific peak was divided by the sum of the heights of the two reference peaks, to give a ratio. The median ratio for each probe across all samples was calculated, and this value was used for

TABLE 1. The EXT1 and EXT2 Exonic Probes Used in This Analysis*

Probe	Standard deviation (range)	Upstream hybridising sequence	Downstream hybridising sequence
EXT1 exon 1	0.06 (0.88–1.04)	GCATGGCAAAGACTGGCAAAAGCACAAGGAT	TCTCGCTGTGACAGAGACAACACCGAGTATGAGAAGTAA
EXT1 exon 2	0.04 (0.92–1.07)	GTATGATTATCGGGAAATGCTGCACAAT	GCCACTTTCTGTCTGGTTCCTCGTGGTCGC
EXT1 exon 3	0.03 (0.93–1.03)	CCCTGTGATGCTCAGCAATGGATGGGAGTTGCCATTCTCT	GAAGTGATTAATTGGAACCAAGCTGCCGTCATAGGCGATG
EXT1 exon 4	0.05 (0.95–1.08)	GGTCTATTCATCAGGATAAAATCCTAGCACTTAGACA	GCAGACACAATTCTTGTGGGAGGCTTATT
EXT1 exon 5	0.03 (0.92–1.04)	CCACAGTATTCATCTTATCTGGGAGATTTTCCTT	ACTACTATGCTAATTTAGGTAAGTGAATTTCCTCCAGGG
EXT1 exon 6	0.02 (0.94–1.02)	GTACTGTGCCCAGGTGAGCGGGAAGT	TGACAGAGAAGCCCCTGCCTGCT
EXT1 exon 7	0.06 (0.94–1.13)	CAGCCATCTAATGAGCCCCATCCCTTTCAGATCATAGTTCT	ATGGAATTGTGACAAGCCCCTACCAGCCAAACACCG
EXT1 exon 8	0.05 (0.96–1.10)	CGAGGACACGGTGCTTTCAACAACAG	AGGTAAGAACCCATGCCTGAGGAGCA
EXT1 exon 9	0.05 (0.95–1.14)	CTCCATGGTGTTGACAGGAGCTGCTATT	TACCACAAGTGAGGAATCTGGACATGT
EXT1 exon 10	0.03 (0.93–1.03)	CCTGAAGAACATGGTGGACCAATTGGCCAATTGTGAGGACAT	TCTCATGAACTTCCTGGTGTCTGCTGTGACAAAATTGCCTC
EXT1 exon 11	0.04 (0.96–1.02)	GAGCTGCATGAATACGTTTGCCA	GCTGGTTTGGCTACATGCCGCTG
EXT2 exon 2	0.04 (0.96–1.08)	CAGTTGCAGAATGCACACGTGTTTTG	ATGTCTATCGCTGTGGCTTCAAC
EXT2 exon 3	0.03 (0.94–1.05)	CTTGACAGGTGGGATCGAGGT	ACGAATCACCTGTTGTTCAACATGTTGCCTG
EXT2 exon 4	0.03 (0.93–1.04)	CTATAGTCCACTGTCAGCTGAGGTGGATCTTCCA	GAGAAAGGACCAGGGTAAGGTACATTCATCCCA
EXT2 exon 5	0.06 (0.94–1.14)	CAAACATGGAGAGTCAGTGTTAGTACTCGAT	AAATGCACCAACCTCTCAGAGGGTGTCCTTT
EXT2 exon 6	0.05 (0.96–1.12)	GCAGTATTGAGCGATGTGTTACAAGCTGGCTGT	GTCCCGGTTGTCATTGCAGACTCCTATATTTT
EXT2 exon 7	0.12 (0.82–1.21)	CAGAGCATCTGTGGTTGTACCAGAAGAAAAGATGTCAGATGT	GTACAGTATTTTGCAGAGCATCCCCCAAAGACAGATT
EXT2 exon 8	0.03 (0.96–1.06)	GCTGCCATCTCCTATGAAGAATGGAATGACCCTCCT	GCTGTGGTAAGTGAATTCCAGTGCTAGCCACATGA
EXT2 exon 9	0.03 (0.94–1.02)	GTTCACCGCCATAGTCCTCACCT	ACGACCGAGTAGAGAGCCTCTTC
EXT2 exon 10	0.03 (0.96–1.05)	CCAGATTCTCTCTGGCCCAAAATCCGGGT	TCCATTAAAAGTTGTGAGGACTGCTGAA
EXT2 exon 11	0.04 (0.95–1.10)	GCATCTCTGGGACCATGAGATGAATAAGTGGAAGT	ATGAGTCTGAGTGGACGAATGAAGTGTCCATGGT
EXT2 exon 12	0.04 (0.95–1.08)	GGTGGCCAACGTCACGGGAAAAGCAGTT	ATCAAGGTAGGAGGCTCTGCCACTCAC
EXT2 exon 13	0.19 (0.69–1.32)	CAGCCATAGATGGGCTTTCACT	AGACCAAACACACATGGTGGAG
EXT 2 exon 14	0.03 (0.94–1.05)	GTTAAGGGTGGAAGGTTGACCTACTTGGATCTTGGCAT	GCACCCACCTAACCCACTTTCTCAAGAACAAGAACCTA

*For each exon the unique sequence of the two half-probes is given, along with the SD and range of normalized ratios obtained when analyzing 12 unaffected control samples.
Mixes of these probes are available upon request (www.LGTC.nl).

normalizing each probe to 1.0 (corresponding to a copy number of two). Thresholds for deletions and duplications were set at 0.75 and 1.25, respectively, meaning that the adjusted ratios within each sample needed to be normalized to 1.0. The normalizing factor was calculated by determining the mean value of the unaffected probes within a sample (defined as falling between 0.8 and 1.2), and dividing all values within that sample by this value. All samples were tested at least twice.

Confirmation of Single-Exon Mutations

Because sequence changes at the ligation site of the two half-probes can also appear as deletions, all single-exon changes were confirmed using another technique. Two of the EXT1 exon 1 deletions were confirmed using MAPH. The sequences for amplification of the probe were forward; AGATGCAGG-GATTTGTGAGG, reverse; CATCTTTGGGTTGCACAATG. Further probe preparation and MAPH was carried out as previously described [White et al., 2002].

The third EXT1 exon 1 deletion was confirmed by FISH analysis. This was performed using standard protocols, with the following probes: D822 (orange), 90D8 (red), and 46F10 (green). D822 is the reference probe for chromosome 8, 90D8 matches exon 1 and the 5′ upstream region of EXT1, and 46F10 covers exons 6–11 of EXT1 [Bernard et al., 2001].

The exon 2 deletion in EXT2 was confirmed by long-range PCR and sequencing across the breakpoints. The PCR reaction was performed using the Expand Long Template PCR System (Roche, www.roche-applied-science.com), with the primers used being forward; CATGATGGGTGCTCAATAATGGTTT, reverse; GCTGTGTTATAATCTGGGGGACCTC. The sequencing reaction used the nested primer, ATTATGTAAGTGCTACGAG-GAGGTG, and was analyzed by the Leiden Genome Technology Center on an ABI 3730 capillary sequencer.

RESULTS

To maximize the number of loci that can be analyzed in a single MLPA assay, we chose to test whether different primer sets can be efficiently coamplified under the same PCR conditions. Testing showed that the primer sequences we used for MAPH analysis [White et al., 2002] were also effective under the MLPA conditions. The probes for EXT1 were designed with the MLPA primer sequences attached, and the probes for EXT2 used MAPH primer sequences. To circumvent the laborious cloning step, we decided to use synthetic oligonucleotides, ranging in size from 39–64 nt, including amplification sequence. The probes were tested on 12 control samples to assess their reliability and consistency, as well as to determine the influence of the two primer pairs on the amplification. The signal strength between the two colors was not always equal, which complicated analysis. Titration experiments showed that adding the MAPH primers at half the concentration of the MLPA primers resolved this issue, usually yielding similar peak heights for both probe sets. The accuracy of analysis, however, was not affected when equimolar amounts of MLPA and MAPH primers were added, even though up to a 10-fold difference in peak height between the two probe sets was occasionally observed.

Of the 24 exonic probes tested, two (EXT2 exon 7 and EXT2 exon 13) gave a standard deviation of greater than 10% (Table 1). These probes were considered to be unreliable, and were not included in further calculations. Of note, the smallest standard deviations were obtained when comparisons were only performed between samples from the same source.

To see if any deletion or duplication mutations could be detected in patients suffering from HME, a total of 18 samples were examined, in which previous sequence analysis was unable to identify any mutations. We identified five rearrangements (Table 2; Figs. 1 and 2)—four in EXT1 and one in EXT2. These mutations were seen irrespective of whether the two probe sets were used separately or combined. The most common deletion was exon 1 of EXT1, which was seen in three unrelated individuals. The deletion was confirmed in one of the samples using FISH (Fig. 3). A probe covering exon 1 and the 5′ upstream region was deleted on one copy of chromosome 8. A probe covering exon 6–11 was present on both copies. In addition, heterozygosity for a single nucleotide polymorphism (SNP) in exon 3 confirmed that the deletion did not extend past exon 2 (data not shown).

Additional analysis of the sample with the deletion of exon 2 of EXT2 showed that the deletion did not include exon 1 (data not shown). Long-range PCR and sequencing defined the deletion to be 422 bp, with one of the breakpoints being in exon 2. The last five nucleotides before the upstream breakpoint (ctccc) are also the last five nucleotides of the deleted sequence, but no further sequence homology was seen.

DISCUSSION

We describe here a further development of MLPA, using synthetic oligonucleotides and two colors. In the original description, one of the two half-probes was generated by cloning into an M13 vector. This approach allows the generation of single-stranded DNA molecules several hundred nucleotides long. The cloning and subsequent restriction digestion, however, is time-consuming and expensive. Using chemically synthesized

TABLE 2. A Summary of the Mutations Found*

Sample	Gene	Mutation at DNA-level	Description of mutation	Confirmed by
1	EXT1	c.-772-?_962+?del	EX1del	MAPH
2	EXT1	c.-772-?_962+?del	EX1del	MAPH
3	EXT1	c.-772-?_962+?del	EX1del	FISH
4	EXT1	c.963-?_3287+?del	EX2_EX11del	Multiple exons
5	EXT2	c.-30-10_441del	EX2del	Long range PCR and sequencing

*The cDNA reference sequences used are NM_000127.1 for EXT1 and NM_000401.1 for EXT2. Nucleotide numbering uses the A of the ATG-translation initiation codon as nucleotide +1.

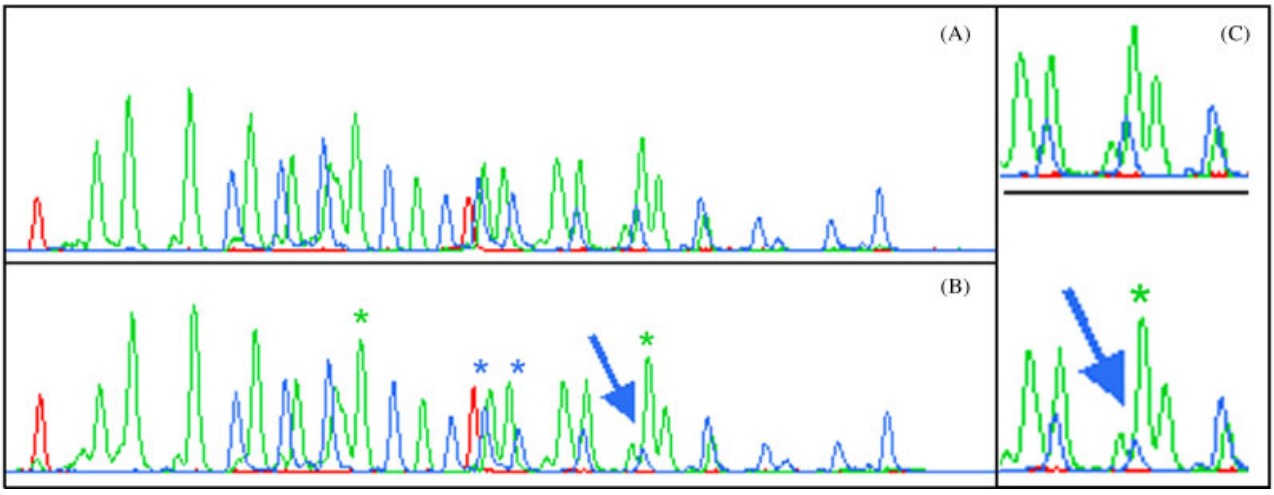

FIGURE 1. **Traces showing the peaks from the two probe sets for EXT1 exons (blue) and EXT2 exons (green). Each set contains two control probes for normalization purposes (marked with *). A:** A normal trace. **B:** An ~50% reduction in the height of the peak corresponding to exon 1 of EXT1 (indicated by arrow). The red peaks are size standard peaks (from left to right; 75 bp and 100 bp). **C:** An enlargement around the EXT1 exon 1 peak.

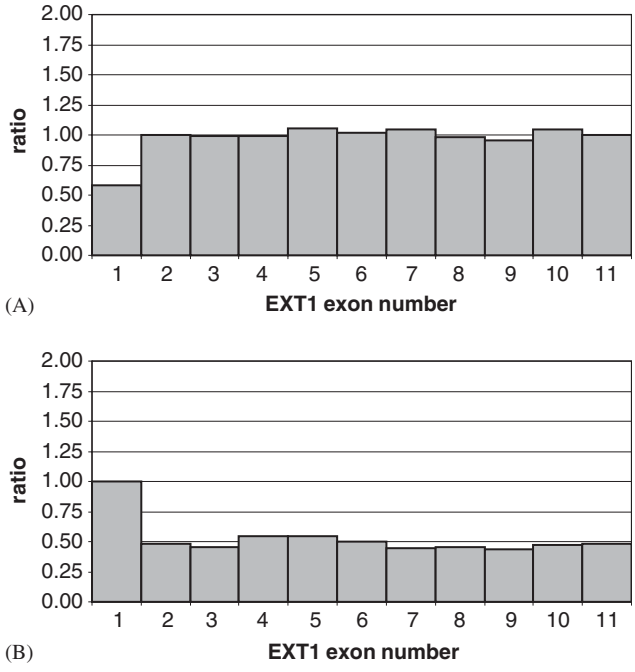

FIGURE 2. **Graphs showing two of the mutations found. A:** EXT1 exon 1 deletion. **B:** EXT1 exons 2–11 deletion.

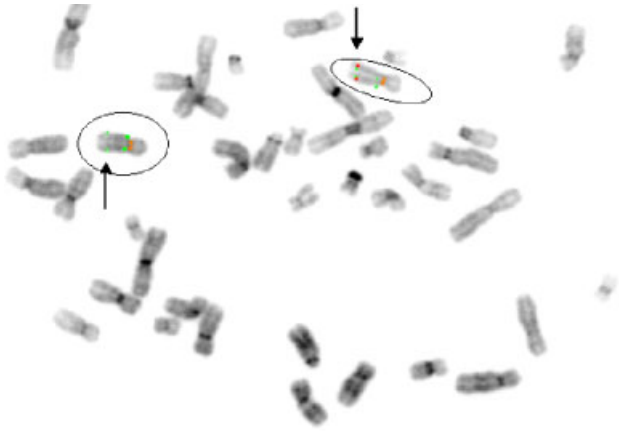

FIGURE 3. **FISH analysis showing a deletion of EXT1 exon 1. The following probes have been hybridized: D822 (orange), 90D8 (red), and 46F10 (green). D822 is the reference probe for chromosome 8, 90D8 matches exon 1 and the 5′ upstream region of EXT1, and 46F10 covers exons 6–11 of EXT1 [Bernard et al., 2001]. The two copies of chromosome 8 are circled, and the EXT1 region is indicated with an arrow. There is no red signal on one of the chromosomes, indicating a deletion of the region corresponding to 90D8.**

oligonucleotides allows rapid and cheap probe development. Furthermore, as each test is performed with only 6 fmol of each oligonucleotide, a synthesis yield of 6 nmol would be sufficient for 1 million reactions.

A size range of 80–125 bp was used for the different probes, with up to 15 probes being combined within a single probe mix. The use of synthetic oligonucleotides limits the length of the probes that can be used. We partially compensated for this by combining two probe sets, each labeled with a different fluorophore. Using two colors effectively doubles the number of probes that can be used in this size range, and in this report 28 probes were used. The size range available is dependent on the maximum length of oligonucleotide synthesis that can be achieved. Our observations with other probe sets are that individual oligonucleotides of up to 75 nt in length can be effectively used, meaning that products of up to 150 bp can be generated. Work is in progress regarding the possibility of using a third primer pair, labeled with a different fluorophore. Together, these factors could allow up to 75 probes to be combined in a single reaction.

We observed that, as previously reported [Schouten et al., 2002], the reliability and reproducibility of the technique is primarily dependent on the quality of the

genomic DNA. We noticed that comparisons made between DNA samples from different sources lead to larger standard deviations than when the same data were normalized only within samples from one source. This was presumably due to different methods of DNA isolation. This observation may have implications when analyzing a series of samples from different laboratories.

The ability to multiplex allows much greater flexibility with regard to future applications. We previously described the use of MAPH as an alternative to FISH with regards to confirming the presence or absence of a rearrangement [White et al., 2003]. Although this multicolor approach should be equally applicable to MAPH, MLPA is perhaps more attractive. The ligation step means that it is not necessary to immobilize the genomic DNA on a filter, and consequently the washing steps can be omitted.

We detected exonic deletions in 5 out of 18 HME samples (28%), four in EXT1 and one in EXT2. The mutations found in EXT1 all have one of the breakpoints within intron 1 of the gene. Notably, this intron makes up ~85% of the total size of the gene. There has recently been a report of a familial translocation, also within intron 1 [Pramparo et al., 2003]. Further work needs to be performed on these samples to characterize the breakpoints and to see if there is a common mechanism involved. Additionally, haplotype analysis could be performed on the patients with the EXT1 exon 1 deletion to see if a common ancestor might be involved. This, however, is unlikely, as the three DNA samples are from three different countries (Spain, the Netherlands, and the United States). In addition, the patient from the Netherlands has no previous family history of HME.

As point mutations are found in EXT1 and EXT2 in 70 to 80% of HME patients, our findings suggest that deletions of one or more exons occur in 5 to 8% of all cases. There are several possible reasons mutations were not found in the remaining samples. The methods applied so far will not detect positional changes (i.e., translocations, inversions, insertions, or transpositions) that affect the structure of the gene without changing the sequence or dosage of any of the exons. This kind of rearrangement will not usually be detected by either MLPA or sequencing. To detect such mutations, analysis at the RNA level may be appropriate [Gardner et al., 1995; Beroud et al., 2004]. Another possibility is that the causative mutation lies not in EXT1 or EXT2, but in another gene. The existence of a third gene (EXT3) on 19p has been postulated [Le Merrer et al., 1994], but to date no specific gene has been identified. Both EXT1 and EXT2 belong to the EXT gene family [Duncan et al., 2001], whose other members also show glycosyltransferase activity. No mutations, however, have been reported in any of the genes (EXTL1, EXTL2, or EXTL3) in HME individuals [Wuyts and Van Hul, 2000]. These genes are potential targets for future copy number analysis.

In summary, we show that MLPA is compatible with the use of synthetic oligonucleotides and a two-color analysis. This combination should facilitate quick and inexpensive probe set development, allowing any gene or region of interest to be rapidly scanned for changes in copy number. In total, design, testing, and application should be feasible within two weeks, with most of the time taken up by oligonucleotide ordering and delivery.

ACKNOWLEDGMENTS

We would like to thank Patrick van Bunderen for expert technical assistance, and the referring physicians and the patients for their cooperation.

REFERENCES

Ahn J, Ludecke HJ, Lindow S, Horton WA, Lee B, Wagner MJ, Horsthemke B, Wells DE. 1995. Cloning of the putative tumour suppressor gene for hereditary multiple exostoses EXT1. Nat Genet 11:137–143.

Armour JA, Barton DE, Cockburn DJ, Taylor GR. 2002. The detection of large deletions or duplications in genomic DNA. Hum Mutat 20:325–337.

Bartsch O, Wuyts W, Van Hul W, Hecht JT, Meinecke P, Hogue D, Werner W, Zabel B, Hinkel GK, Powell CM, Shaffer LG, Willems PJ. 1996. Delineation of a contiguous gene syndrome with multiple exostoses, enlarged parietal foramina, craniofacial dysostosis, and mental retardation, caused by deletions in the short arm of chromosome 11. Am J Hum Genet 58:734–742.

Bernard MA, Hall CE, Hogue DA, Cole WG, Scott A, Snuggs MB, Clines GA, Ludecke HJ, Lovett M, Van Winkle WB, Hecht JT. 2001. Diminished levels of the putative tumor suppressor proteins EXT1 and EXT2 in exostosis chondrocytes. Cell Motil Cytoskeleton 48:149–162.

Beroud C, Carrie A, Beldjord C, Deburgrave N, Llense S, Carelle N, Peccate C, Cuisset JM, Pandit F, Carre-Pigeon F, Mayer M, Bellance R, Recan D, Chelly J, Kaplan JC, Leturcq F. 2004. Dystrophinopathy caused by mid-intronic substitutions activating cryptic exons in the DMD gene. Neuromuscul Disord 14:10–18.

Den Dunnen JT, Grootscholten PM, Bakker E, Blonden LAJ, Ginjaar HB, Wapenaar MC, Van Paassen HMB, Van Broeckhoven C, Pearson PL, Van Ommen GJB. 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.

Duncan G, McCormick C, Tufaro F. 2001. The link between heparan sulfate and hereditary bone disease: finding a function for the EXT family of putative tumor suppressor proteins. J Clin Invest 108:511–516.

Gardner RJ, Bobrow M, Roberts RG. 1995. The identification of point mutations in Duchenne muscular dystrophy patients using reverse transcript PCR and the protein truncation test. Am J Hum Genet 57:311–320.

Gille JJ, Hogervorst FB, Pals G, Wijnen JT, van Schooten RJ, Dommering CJ, Meijer GA, Craanen ME, Nederlof PM, de Jong D, McElgunn CJ, Schouten JP, Menko FH. 2002. Genomic deletions of MSH2 and MLH1 in colorectal cancer families detected by a novel mutation detection approach. Br J Cancer 87:892–897.

Hogervorst FB, Nederlof PM, Gille JJ, McElgunn CJ, Grippeling M, Pruntel R, Regnerus R, van Welsem T, van Spaendonk R, Menko FH, Kluijt I, Dommering C, Verhoef S, Schouten JP, van't Veer LJ, Pals G. 2003. Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. Cancer Res 63:1449–1453.

Kent WJ. 2002. BLAT–the BLAST-like alignment tool. Genome Res 12:656–664.

Le Merrer M, Legeai-Mallet L, Jeannin PM, Horsthemke B, Schinzel A, Plauchu H, Toutain A, Achard F, Munnich A, Maroteaux P. 1994. A gene for hereditary multiple exostoses maps to chromosome 19p. Hum Mol Genet 3:717–722.

Ludecke HJ, Wagner MJ, Nardmann J, La Pillo B, Parrish JE, Willems PJ, Haan EA, Frydman M, Hamers GJ, Wells DE. 1995. Molecular dissection of a contiguous gene syndrome: localization of the genes involved in the Langer-Giedion syndrome. Hum Mol Genet 4:31–36.

Petrij F, Dauwerse HG, Blough RI, Giles RH, van der Smagt JJ, Wallerstein R, Maaswinkel-Mooy PD, van Karnebeek CD, van Ommen GJ, Van Haeringen A, Rubinstein JH, Saal HM, Hennekam RC, Peters DJ, Breuning MH. 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.

Pramparo T, Gregato G, De Gregori M, Friso A, Clementi M, Ardenghi P, Rocchi M, Zuffardi O, Tenconi R. 2003. Reciprocal translocation associated with multiple exostoses in seven members of a three generation family and discovered through an infertile male. Am J Med Genet 123A:79–83.

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30:e57.

Stickens D, Clines G, Burbee D, Ramos P, Thomas S, Hogue D, Hecht JT, Lovett M, Evans GA. 1996. The EXT2 multiple exostoses gene defines a family of putative tumour suppressor genes. Nat Genet 14:25–32.

Taylor CF, Charlton RS, Burn J, Sheridan E, Taylor GR. 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.

White S, Kalf M, Liu Q, Villerius M, Engelsma D, Kriek M, Vollebregt E, Bakker B, van Ommen GJ, Breuning MH, Den Dunnen JT. 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.

White SJ, Sterrenburg E, van Ommen GJ, Den Dunnen JT, Breuning MH. 2003. An alternative to FISH: detecting deletion and duplication carriers within 24 hours. J Med Genet 40:e113.

Wuyts W, Van Hul W, De Boulle K, Hendrickx J, Bakker E, Vanhoenacker F, Mollica F, Ludecke HJ, Sayli BS, Pazzaglia UE, Mortier G, Hamel B, Conrad EU, Matsushita M, Raskind WH, Willems PJ. 1998. Mutations in the EXT1 and EXT2 genes in hereditary multiple exostoses. Am J Hum Genet 62:346–354.

Wuyts W, Van Hul W. 2000. Molecular basis of multiple exostoses: mutations in the EXT1 and EXT2 genes. Hum Mutat 15: 220–227.

Wuyts W, Van Hul W, Bartsch O, Wilkie AO, Meinecke P. 2001. Burning down DEFECT11. Am J Med Genet 100:331–332.

Yau SC, Bobrow M, Mathew CG, Abbs SJ. 1996. Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis. J Med Genet 33:550–558.

Chapter 5.2

Roelfsema J.H., White S.J., Arıyürek Y., Lacombe D., Bartholdi D., Niedrist D., Papadia F., den Dunnen J.T., van Ommen G.J., Breuning M.H., Hennekam R.C., Peters D.J.M. Genetic Heterogeneity in Rubinstein-Taybi syndrome: mutations in the *CBP* and *EP300* gene are both disease causing. Submitted for publication.

(Colour images from this chapter can be seen in the appendix)

Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in the *CBP* and *EP300* gene are both disease causing.

Jeroen H Roelfsema[1], Stefan J White[1], Yavuz Arıyürek[1], Deborah Bartholdi[3], Dunja Niedrist[3], Francesco Papadia[4], Carlos Bacino[5], Johan T den Dunnen[1], Gert-Jan B van Ommen[1], Martijn H Breuning[1], Raoul C Hennekam[5] and Dorien JM Peters[1]

[1]Center for Human and Clinical Genetics, LUMC, Sylvius Laboratory, Wassenaarseweg 72, 2333AL, Leiden, The Netherlands.
[2] Institute of Medical Genetics, University of Zurich, Switzerland.
[3] Department of Metabolic Diseases and Clinical Genetics, Pediatric Hospital Giovanni XXIII, Bari, Italy.
[4] Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands.
[5] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, United States of America.

Abstract

CREB Binding Protein and p300 function as transcriptional coactivators in the regulation of gene expression through various signal transduction pathways. Both are potent Histone Acetyl Transferases. The level of CREB Binding Protein is essential for normal development, as inactivation of one allele causes Rubinstein-Taybi syndrome. There is a direct link between loss of acetyl transferase activity and Rubinstein-Taybi syndrome, which indicates that the disorder is caused by aberrant chromatin regulation. We screened the entire CBP gene for mutations in Rubinstein-Taybi syndrome patients using methods to find point mutations and larger rearrangements. In 92 patients we were able to identify a total of 36 mutations in the CBP gene. Using Multiple Ligation-dependent Probe Amplification we not only found several deletions but also the first duplication in a Rubinstein-Taybi syndrome patient. We extended the search for mutations to the *EP300* gene and showed that mutations in *EP300* also cause this disorder. These are the first mutations identified in *EP300* in a congenital disorder.

Introduction

Rubinstein-Taybi syndrome (RSTS) is a congenital disorder characterized by mental and growth retardation and a wide range of typical dysmorphic features. Facial dysmorphology includes down slanted palpebral fissures, broad nasal bridge, a beaked nose and micrognathia. Particularly noticeable are the broad thumbs and broad big toes. In addition, RSTS patients have an increased risk for tumor formation. Although various types of tumors have been described, there is an excess of tumors arising from developmental defects and tumors of brain or neural crest cell

derived tissue (Miller and Rubinstein 1995). Mutations in the gene coding for the CREB binding protein (CREBBP, also known as CBP), located on chromosome 16p13.3, were found to be responsible for causing the disorder (Petrij et al. 1995).

CBP serves as a transcriptional coactivator (Kwok et al. 1994). It has a transactivation domain but does not specifically bind to DNA. The name of the protein is based on the interaction with the CRE binding protein (CREB); however, CBP interacts with a large number of transcription factors. It is thought that CBP acts as an integrator of the signals from various pathways (Goodman and Smolik 2000). Transcription factors downstream from these pathways need to compete with each other for the limited amount of CBP available in the nucleus. The protein forms a physical bridge between the DNA binding transcription factors and the RNA polymerase II complex. In addition, CBP has intrinsic histone acetyl transferase (HAT) activity (Bannister and Kouzarides 1996). By acetylating histones it opens the chromatin structure at the locus that needs to be expressed, a process essential for gene expression. CBP is also capable of acetylating a large number of other proteins, for example the transcription factor p53 (Gu and Roeder 1997).

RSTS is considered to be an autosomal dominant disorder, however patients very rarely have children. Almost all mutations, therefore, occur *de novo*. The mutations found in patients range from relatively large microdeletions, removing the gene entirely, to point mutations. In addition, five translocations and two inversions disrupting the gene have been reported (Petrij et al. 2000). The microdeletions that remove the entire gene indicate that haploinsufficiency is the ultimate cause of the syndrome. Presumably, at critical moments during development the amount of CBP drops below a certain threshold because of the loss of one allele. How this loss of one allele actually causes the particular symptoms of RSTS, however, is unclear. Nevertheless, we know from patients with missense mutations and splice site mutations affecting only the HAT domain of CBP, that loss of HAT activity is sufficient to cause the syndrome (Murata et al. 2001; Kalkhoven et al. 2003).

In order to elucidate the complete spectrum of mutations we screened 92 RSTS patients for point mutations, small deletions or insertions and for larger deletions and duplications. Because we could not find mutations in the *CBP* gene in the majority of our patients we assumed that the remaining patients have mutations in other genes.

CBP shares homology with another protein, p300, encoded by the *EP300* gene on chromosome 22q13.2 (Lundblad et al. 1995). Both proteins are particularly homologous at their binding sites for transcription factors and p300 also has a HAT domain. Like CBP it serves as a transcriptional coactivator. A likely candidate to screen, therefore, is *EP300* and, indeed, we found three

mutations. These are the first mutations described in *EP300* in a congenital disorder and they also prove that RSTS is a genetically heterogeneous disorder.

Material and Methods

The majority of the DNA samples described in this study were sent to us by clinicians in the Netherlands and many other countries as soluble genomic DNA from patients with a clinical diagnosis of RSTS. DNA from the rest of the patients was isolated from peripheral blood in our laboratory using standard protocols.

DGGE

DGGE was performed with a GC-clamp on either the forward or the reverse primer. Primers were selected to anneal to the flanking intron sequences in order to screen the splice sites and the branch sites, and were chosen using either WINMELT (Biorad) or MELT-INGENY (Ingeny B.V.) software. All oligonucleotides were synthesized by Sigma-Aldrich. Amplified fragments were analyzed on 9% polyacrylamide gels (37.5:1) with various linear denaturing gradients, optimized for each fragment, on the DCode system from Biorad. Gels were run at 90V at a constant temperature of 60ºC. An acrylamide mixture with 40% formamide and 7M urea was defined as 100% denaturant and acrylamide without these denaturing agents was defined as 0% denaturant.

SSCP

Electrophoresis was performed at room temperature using two types of gels. The first type was a polyacrylamide gel (49:1) with 1*TBE without glycerol and the second type was 0.5*MDE (National Diagnostics, Atlanta, Georgia) with 0.6*TBE and 10% glycerol. During amplification the fragments for SSCP analysis were radioactively labeled either by incorporation of α^{32}P-dCTP or by using primers that were kinated using γ^{32}P-dATP (Amersham). Visualization of the fragments was done using the PhosphorImager (Molecular Dynamics).

MLPA

Probes were designed for 20 exons of the *CBP* and *EP300* genes. MLPA was performed as described in (White et al. 2004). All samples were tested at least twice.

Sequencing and restriction digestions.

Sequencing was performed on the ABI 3700 from Applied Biosystems using the manufacturers standard protocol and reagents. Restriction digestions were performed according to the instructions of the manufacturer. Digestions or second sequencing reactions to confirm the first result were done on PCR fragments generated in an independent reaction. The deletion of 8

nucleotides in patient 256-1 was confirmed by PCR with an allele-specific primer, *tcctccatctactagtagtg*, that skips the deleted part and anneals with 2 nucleotides after the deletion. The reverse primer has the sequence *gtcctaaccaaatcaaacag*.

Results

Point mutations and small deletions or insertions in the CBP gene

We screened the entire *CBP* gene for point mutations and small deletions or insertions using primarily DGGE, with target sequences that were not suited for DGGE being screened by SSCP analysis. The complete coding sequence and splice sites of the *CBP* gene required a total of 49 fragments of which 40 were screened using DGGE, approximately 83% of the coding sequence. Direct sequencing was used to identify the mutation after aberrant bands were found on DGGE or SSCP gels. All mutations were confirmed either by digestion with restriction enzymes when a restriction enzyme site was altered or by a second sequence analysis.

In 92 patients we found a total of 27 mutations (see table 1). The majority is predicted to lead to a premature translation stop but we also detected 5 putative missense mutations. Base substitutions leading to a premature stop codon or deletions and insertions leading to frame shifts can be clearly identified as disease causing mutations. A change of amino acids is much less clear, however, RSTS patients as a rule have *de novo* mutations. Since we were able to confirm the mutation as *de novo* for three of the mutations we consider them most likely to be disease causing. We do not have parental DNA of patients 228-1 and 260-1. All putative missense mutations are at the highly conserved HAT domain of CBP and the amino acids that are changed have residues that are conserved in both the mouse and the fruit fly (see fig.1).

Unless we have an RNA sample from a patient we cannot check whether a splice site mutation actually leads to aberrant splicing. These mutations however, should also comply with the rule that mutations in RSTS patients occurred *de novo*. Except for the mutation in patient 39-1 for which parental DNA was not available, we could confirm the mutations that way. The mutation, a G to A, in the splice donor site flanking exon 24 in patient 39-1, however, is at the first position, which should in all splice donor sites, without exception, be a guanine. The splice site mutation of patient 211-1 could be analyzed on RNA isolated from a cell line. Subsequent sequence analysis proved that the mutation in the splice acceptor site flanking exon 22 leads to a deletion of exon 22 in the processed mRNA (Kalkhoven et al. 2003).

Mutations in the *CBP* gene

Individual	Exon	Mutation	
		Nonsense mutations	
7-1	Exon 2	c.304 C>T	Q102X
177-1	Exon 5	c.1237 C>T	R413X
212-1 *	Exon 28	c.4669 C>T	Q1558X
27-1	Exon 29	c.4879 A>T	K1627X
2-1	Exon 31	c.6010 C>T	R2004X
16-1	Exon 31	c.6133 C>T	Q2045X
178-3	Exon 31	c.6283 C>T	Q2095X
		Missense mutations	
209-1 *	Exon 21	c.3823 G>A	E1278K
201-1	Exon 26	c.4340 C>T	T1447I
260-1	Exon 26	c.4348 T>C	Y1450H
228-1	Exon 27	c.4409 A>G	H1470R
2644 *	Exon 30	c.4991 G>A	R1664H
		Deletions & Insertions	
153-1	Exon 2	c.235 del G	G79fsX86
199-3	Exon 3	c.904_905 del AG	S302fsX348
205-1	Exon 6	c.1381_1388 del 8	G461fsX469
239-1	Exon 6	c.1481 dup A	N494fsX527
203-1	Exon 8	c.1735 dup A	A581fsX586
57-3	Exon 18	c.3396_3400 del 6	P1132fsX1166
10-1	Exon 18	c.3432_3433 del AG	T1144fsX1168
232-1	Exon 21	c.3824 dup T	F1275fsX1282
231-1 *	Exon 25	c.4256_4258 del CT	S1419fsX1419
34-3	Exon 27	c.4399 del G	V1467fsX1467
213-1 *	Exon 29	c.4837 del G	V1613fsX1634
		Splice site mutations	
198-3 *	Exon 20	c.3779 +5 G > C	
211-1 *	Exon 22	c.3837 -2 A > T	
47-3	Exon 23	c.3915 -1 G > A	
39-1 *	Exon 24	c.4133 +1 G > A	
		Rearrangements found by MLPA	
267-1	Del Exon 1	c.-198-?_85+? del	
36-3	Del Exon 1_2	c.-198-?_798+? del	
74-1	Del Exon 1_19	c.-198-?_3698+? del	
15-1	Del Exon 1_31	c.-198-?_+1150+? del	
41-3	Del Exon 1_31	c.-198-?_+1150+? del	
127-2 *	Del Exon 2	c.86-?_798+? del	
252-1	Del Exon 12	c.2159-?_2283+? del	
253-1	Del Exon 31	c.5173-?_+1150+? del	
162-1	Dup Exon 1	c.-198-?_85+? dup	

Mutations in the *EP300* gene

Individual	Exon	Mutation	
254-1	Exon 10	c.1942 C>T	R648X
256-1	Exon 15	c.2877_2884 del 8	S959fsX966
149-1	Del Exon 1	c.-1200-?_94+? del	

Table 1: List of all mutations found in the *CBP* gene, described in relation to GenBank file NM_004380, and *EP300*, GenBank file NM_001429.1, counting the A of the ATG start codon as nucleotide +1. The mutations are denoted according to the nomenclature as published by (den Dunnen and Antonarakis 2001). Del and dup means deletion and duplication; question marks indicate the breakpoints are unknown. The changes on DNA level have been confirmed by restriction digests or by second sequencing reactions. The changes on protein level listed here are predictions. All patients marked with an asterisk have mutations that have been published before in Kalkhoven et al. (2003) except for 127-2, which has been described as a deletion in mRNA by Petrij et al.(2000).

Large deletions and duplications at the CBP gene

Previous research suggests that approximately 10% of the mutations of RSTS patients are microdeletions affecting the *CBP* gene (Blough et al. 2000; Petrij et al. 2000). We performed Fluorescent *in situ* Hybridization (FISH) using five cosmids spanning the entire gene to detect such deletions when metaphase chromosome spreads of patients are available (Petrij et al., 2000). The recently developed technique of Multiple Ligation-dependent Probe Amplification (MLPA) can also be used to detect microdeletions on soluble genomic DNA (Schouten et al. 2002). Because that is the type of material available to us for the majority of our patients we set up MLPA on the *CBP* gene.

	E1278K		T1447I
Hs:	KKKNDTLDPEPFVDCKECG	Hs:	HFFRPRCLRTAVYHEILIG
Mm:	KKKNDTLDPEPFVDCKECG	Mm:	HFFRPRCLRTAVYHEILIG
Dm:	EKKNDHLELEPFVNCQECG	Dm:	HFFRPRQYRTAVYHEILLG
209-1:	KKKNDTLDPKPFVDCKECG	201-1:	HFFRPRCLRIAVYHEILIG
	Y1450H		H1470R
Hs	RPRCLRTAVYHEILIGIFH	Hs	VKKLGYVTGHIWACPPSEG
Mm:	RPRCLRTAVYHEILIGIFH	Mm:	VKKLGYVTGHIWACPPSEG
Dm:	RPRQYRTAVYHEILLGYMD	Dm:	VKQLGYTMAHIWACPPSEG
260-1:	RPRCLRTAVHHEILIGIFH	228-1:	VKKLGYVTGRIWACPPSEG
	R1664H		
Hs	LLSCDLMDGRDAFLTLARD		
Mm:	LLSCDLMDGRDAFLTLARD		
Dm:	LLSCDLMDGRDAFLTLARD		
2644:	LLSCDLMDGHDAFLTLARD		

Figure 1: Conservation of amino acids predicted to change by missense mutations. All five mutations that are predicted to change the amino acid residue that we have found are situated in the highly conserved HAT domain. The changed residues are conserved in man (*Homo sapiens*), mouse (*Mus musculus*) and the fruit fly (*Drosophila melanogaster*).

The resolution of MLPA is related to the number of probes one uses. We made a set of 20 MLPA probe pairs covering most of the *CBP* gene. This allows us to screen for deletions that cannot be detected by FISH. Southern blotting could have been an alternative but is in our case impractical, if not impossible, because it requires too much DNA.

The quality of DNA is slightly more critical in MLPA than in a normal PCR, therefore, we could not screen all patients with MLPA that have been screened with DGGE and SSCP. In total we screened 53 patients and as controls we used material from 3 patients with known microdeletions already detected using FISH, including one with a deletion of the entire gene. Our MLPA analysis detected those positive controls flawlessly and we found a number of previously undetected mutations. In total we found 9 new deletions, ranging from single exon deletions to the entire gene. One deletion, of exon 2, has been described previously on RNA level (Petrij et al. 2000). At the time Southern blots did not reveal a deletion in the genomic DNA, therefore, it was not clear whether this was a genomic deletion or a splicing aberration. This mutation has been found in family 127, which consists of an affected mother and child, one of the very few cases of inherited RSTS.

Next to the nine deletions we have detected we also found a duplication in one individual. Patient 162-1 has a duplication of the first exon of the *CBP* gene. How this leads to the inactivation of this allele is not clear but a disease causing duplication of first exon has been described before in Opitz syndrome (Winter et al. 2003).

The exon 1 deletions and duplication were confirmed using extra probe pairs, one at the promoter region and three probe pairs in intron 1.

Mutations in the EP300 gene

Point mutation screening and MLPA analysis of *CBP* yielded a total of 36 mutations in 92 patients, suggesting that other genes could be involved in RSTS as well. The most likely candidate is the *EP300* gene, coding for p300, on chromosome 22q13.1. That gene was screened as well, using the same approach. We used 37 DGGE fragments, covering approximately 79% of the coding sequence of *EP300*, with the remaining part was covered by 10 SSCP fragments. MLPA was performed with a set of 20 exon specific probe pairs.

Indeed, 3 inactivating mutations were detected in the *EP300* gene (see fig.2). Two mutations were found using DGGE: one mutation, in exon 10, is a transition (c.1942 C>T) that converts the triplet coding for the arginine at position 648 into a stop codon. The other mutation, in exon 15, is a deletion of 8 nucleotides that predicts a frameshift from codon 959 with a stop codon after 7 amino acids. The exact location of the 8 bp deletion (c.2877_2884) was confirmed with an allele specific PCR. We analyzed DNA from the healthy parents of both patients with DGGE and sequencing and confirmed that the mutations occurred *de novo*. The biological parentage was confirmed by genotyping with 17 independent markers (data not shown). Both mutations lead to predicted proteins less than half their normal size, that do not contain the HAT domain. The third

mutation, a deletion of the first exon, was found using MLPA. Four probes revealed this deletion, two probes upstream of exon 1, one in exon 1 and the fourth in intron 1, close to the first exon. They all showed decreased signal whereas a probe in exon 2 showed a normal dosage (see fig.2c). It is probable that this deletion will lead to no expression from the affected allele.

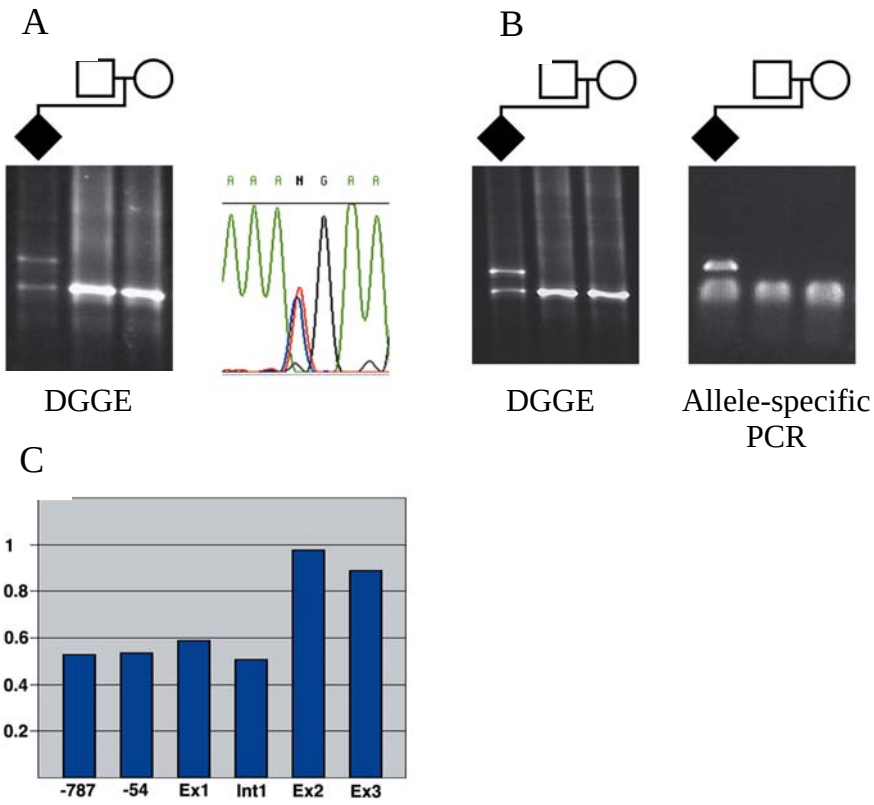

Figure 2: Mutations in *EP300* in RSTS patients. **(a)** Patient 254-1. DGGE of patient 254-1 and the healthy parents shows that only the affected child has the mutation. Subsequent sequence analysis revealed a transition c.1942 C>T that predicts p.Arg648X in this patient. **(b)** Patient 256-1. DGGE of family 256 with patient 256-1 shows a *de novo* mutation. The allele specific PCR confirms the exact location of the deletion seen by sequence analysis. The patient has an 8 bp deletion (c.2877_2884 del) in the following sequence: gcc*tcctccatctactagtag*CACAGAAG*tg*aat. The deleted region is indicated with capitals. The allele specific forward primer, consisting of the underscored nucleotides, skips the deleted part and anneals with 2 nucleotides after the deletion. Only the PCR on DNA from the patient shows a band of 168 bp, in the lanes with the PCR on DNA from the healthy parents only the prominently visible primer dimers can be seen. **(c)** Bar diagram of MLPA results for patient 149-1. MLPA reveals a deletion at the first exon of the *EP300* gene. The bars indicate the dosage of the various probes used. The probes upstream of the first exon are at positions 787 to 716 and 54 to 5 bp before the transcription start site. In the figure they are indicated as –787 and –54 respectively. Ex refers to exon and Int to intron. The Y-axis represents the dosage of DNA: a dosage of 1 indicates the presence of the normal amount of DNA, that is, both alleles are present, whereas bars reaching approximately 0.5 typically indicate a deletion of one allele. In the figure it can be clearly seen that the deletion runs from the upstream region of exon 1 into intron 1 and that exons 2 and 3 are present for both alleles. The exact size of the deletion is unknown.

Discussion

We undertook a rigorous screening for point mutations, small deletions or insertions as well as larger deletions and duplications in the coding region of the *CBP* gene on genomic DNA of a large set of RSTS patients. There is no predominant type of mutation, nor is there a clear indication for clustering of mutations within the *CBP* gene. If we take a look, however, at missense mutations we see that they are all situated in the HAT domain of CBP. We have published some of these mutations previously and have shown that they affect the HAT activity of CBP. In addition, two papers each reported a *de novo* missense mutation that is within the HAT domain, clearly underpinning the importance of this domain in relation to the disorder (Murata et al. 2001; Bartsch et al. 2002). A study by Coupry et al. reported 4 putative missense mutations, of which only one was located in the HAT domain (Coupry et al. 2002). The sequence variations were not found in the other patients, and the affected residues were conserved in mouse.

We have found mutations in less than half of the patients, approximately 40%, which is comparable with the outcome of the study by Coupry et al. DGGE and SSCP analysis are, together with detection of nucleotide substitutions, only capable of identifying relatively small deletions and insertions. To detect larger deletions we chose to set up MLPA for the *CBP* gene and for *EP300* as well. We have shown that MLPA is capable of detecting deletions in the *CBP* gene that were previously identified by FISH. Because we have probe pairs corresponding to the majority of exons in both *CBP* and *EP300* our MLPA screening also negates the need for Southern blotting. The use of MLPA has increased the detection power for mutations, allowing us to find smaller deletions than could be detected with FISH.

The combined analysis of our samples with both MLPA and DGGE or SSCP nevertheless resulted in mutations being found in less than half of the patients. Although some of the patients we screened could possibly have a different syndrome that resembles RSTS we think the majority should be considered as true RSTS patients. Diagnosis of the syndrome has been performed by many clinicians but we do not see that some have a significantly better record in number of mutations found than others. Either the *CBP* gene is mutated at parts where we did not screen, such as the promoter or other regulatory elements, or the mutations are in other genes. The unscreened parts of the *CBP* gene may harbor some mutations but it is highly unlikely they will contain the majority of the missing 60%. Indeed, RSTS is genetically heterogeneous since we identified mutations in the *EP300* gene.

A striking finding in our study is that the number of RSTS patients with *EP300* mutations, now 3, is small compared to the number of RSTS patients, 36, with *CBP* mutations. Possibly, this ratio of 1 to 12 represents the different chances of mutations occurring in these two genes. Alternatively, the *EP300* gene could have an equal mutation rate as the *CBP* gene but the carriers may not be diagnosed with RSTS. In view of this latter explanation it is interesting that we found many more polymorphisms in the *EP300* gene including some that lead to amino acid changes (data not shown). Nevertheless, the majority of point mutations found in the *CBP* gene are likely to lead to truncated proteins and two mutations in the *EP300* gene are also predicted to truncate the protein so it is difficult to explain the skewed ratio with a different genotype/phenotype relationship. We therefore think that there is a different mutation rate between the two loci.

The CBP gene contains an unstable region around exon 2. This region was designated as unstable because all translocation and inversion breakpoints in RSTS patients, except for one, could be found there, as well as all leukemia breakpoints where CBP functions as a fusion partner. In addition, this same genomic piece of DNA proved very difficult to clone when the positional cloning of the RSTS syndrome gene took place (Giles et al. 1997). The deletion of exon 2 and the deletions and duplication of exon 1 may be caused by this unstable region. The instability in this region, however, cannot explain the majority of deletions found at the *CBP* locus as most of these deletions have their breakpoints elsewhere (Petrij et al. 2000).

Although CBP and p300 are probably redundant to a large extent, there are subtle but clear differences between the two proteins. During embryogenesis the two genes have similar but not completely overlapping expression patterns (Partanen et al. 1999). In addition, experiments with F9 teratocarcinoma cell lines showed that retinoic acid signaling is p300 dependent and does not require CBP, whereas cAMP signaling depends on CBP and not p300 (Kawasaki et al. 1998; Ugai et al. 1999). Recent work with transgenic mice indicated the importance of the acetyl transferase function of p300 in myogenesis, but the acetyl transferase function of Cbp does not seem to be necessary for this process (Roth et al. 2003). The skeletal abnormalities found in heterozygous *Cbp* knockout mice have not been reported for heterozygous *Ep300* knockout mice (Tanaka et al. 1997). We, however, do not see clear phenotypical differences in RSTS patients with mutations in the *EP300* gene instead of the *CBP* gene. Patient 256-1 has a very short metatarsal bone (figure 3), not often seen in RSTS patients, however, similar dysmorphology has been found in a patient with a deleted *CBP* gene (Petrij et al. 1995). Double heterozygous knockout mice for the *Cbp* and *Ep300* genes resemble the homozygous knockout mice for either gene, in that all three types of mice die *in utero*, which led to the idea that the combined levels of

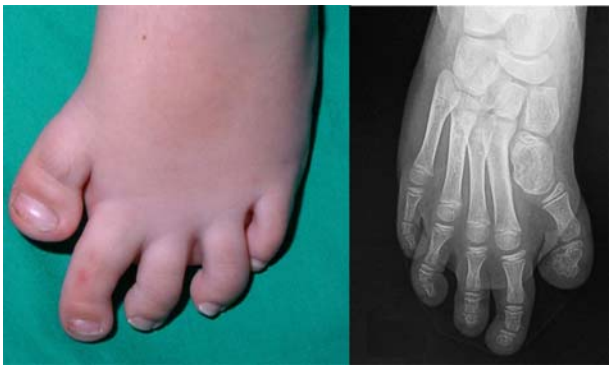

Figure 3: Patient 256-1 overall has the typical appearance of an RSTS patient with the exception of the feet. These feet have an abnormally short metatarsal I bone, as can clearly be seen in the X-ray photograph. Although it is not a typical feature, it does appear in some other RSTS patients as well with mutations in the *CBP* gene. The photograph of the foot was taken when the patient was 6 years old whereas the X-ray was taken when the patient was 9 years old. Photographs courtesy of the patient's parents.

CBP and p300 are critical during development (Yao et al. 1998). Our finding supports this hypothesis and reveals that even a relatively small decrease of either protein has significant developmental consequences. It is, however, unclear how a decrease of either protein leads to the specific features of RSTS. Perhaps the partial loss of p300 is compensated for by recruitment of CBP and subsequent depletion of CBP than leads to RSTS. Alternatively, both proteins could be involved in a common function and, therefore, the total dosage is required to prevent a syndrome like RSTS. If so, then this common function has a relationship with the HAT activity of the proteins because loss of only the HAT activity of CBP causes RSTS.

Interestingly, there is a direct link between HAT activity and long-term memory. Heterozygous *Cbp* knockout mice have diminished mental capabilities. Experiments on these knockout mice revealed that inhibiting histone deacetyltransferase could ameliorate the problems the mice have with their long-term memory (Alarcon et al. 2004). Transgenic mice with a dominant negative CBP gene, where only the HAT activity was ablated, also showed the long-term memory problems. Again, this could be reversed by a histone deacetylase inhibitor (Korzus et al. 2004). In view of these data it could be possible that other proteins with HAT activity, or with a function coupled to HAT activity, may also be involved in RSTS. After all, the three mutations we have found in the *EP300* gene together with the *CBP* gene mutations still leaves us with more than half of the RSTS patients to be accounted for.

ACKNOWLEDGEMENTS:

We thank the patients and their parents for their cooperation. We thank S. Trompet for technical assistance and Dr.P. de Knijff for genotyping the families. This work was funded by grants from

the Dutch Cancer Society and the Center of Medical System Biology (CSMB) established by the Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research (NGI/NWO).

References

Alarcon JM, Malleret G, Touzani K, Vronskaya S, Ishii S, Kandel ER, Barco A (2004) Chromatin acetylation, memory, and LTP are impaired in CBP+/- mice: a model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration. Neuron 42:947-59

Bannister AJ, Kouzarides T (1996) The CBP co-activator is a histone acetyltransferase. Nature 384:641-3

Bartsch O, Locher K, Meinecke P, Kress W, Seemanova E, Wagner A, Ostermann K, Rodel G (2002) Molecular studies in 10 cases of Rubinstein-Taybi syndrome, including a mild variant showing a missense mutation in codon 1175 of CREBBP. J Med Genet 39:496-501.

Blough RI, Petrij F, Dauwerse JG, Milatovich-Cherry A, Weiss L, Saal HM, Rubinstein JH (2000) Variation in microdeletions of the cyclic AMP-responsive element-binding protein gene at chromosome band 16p13.3 in the Rubinstein-Taybi syndrome. Am J Med Genet 90:29-34

Coupry I, Roudaut C, Stef M, Delrue MA, Marche M, Burgelin I, Taine L, Cruaud C, Lacombe D, Arveiler B (2002) Molecular analysis of the CBP gene in 60 patients with Rubinstein-Taybi syndrome. J Med Genet 39:415-21.

den Dunnen JT, Antonarakis SE (2001) Nomenclature for the description of human sequence variations. Hum Genet 109:121-4.
Giles RH, Petrij F, Dauwerse HG, den Hollander AI, Lushnikova T, van Ommen GJ, Goodman RH, Deaven LL, Doggett NA, Peters DJ, Breuning MH (1997) Construction of a 1.2-Mb contig surrounding, and molecular analysis of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome 16p13.3. Genomics 42:96-114

Goodman RH, Smolik S (2000) CBP/p300 in cell growth, transformation, and development. Genes Dev 14:1553-77.

Gu W, Roeder RG (1997) Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell 90:595-606

Kalkhoven E, Roelfsema JH, Teunissen H, den Boer A, Ariyurek Y, Zantema A, Breuning MH, Hennekam RC, Peters DJ (2003) Loss of CBP acetyltransferase activity by PHD finger mutations in Rubinstein-Taybi syndrome. Hum Mol Genet 12:441-50

Kawasaki H, Eckner R, Yao TP, Taira K, Chiu R, Livingston DM, Yokoyama KK (1998) Distinct roles of the co-activators p300 and CBP in retinoic-acid-induced F9-cell differentiation. Nature 393:284-9

Korzus E, Rosenfeld MG, Mayford M (2004) CBP histone acetyltransferase activity is a critical component of memory consolidation. Neuron 42:961-72

Kwok RP, Lundblad JR, Chrivia JC, Richards JP, Bachinger HP, Brennan RG, Roberts SG, Green MR, Goodman RH (1994) Nuclear protein CBP is a coactivator for the transcription factor CREB. Nature 370:223-6

Lundblad JR, Kwok RP, Laurance ME, Harter ML, Goodman RH (1995) Adenoviral E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP. Nature 374:85-8

Miller RW, Rubinstein JH (1995) Tumors in Rubinstein-Taybi syndrome. Am J Med Genet 56:112-5.

Murata T, Kurokawa R, Krones A, Tatsumi K, Ishii M, Taki T, Masuno M, Ohashi H, Yanagisawa M, Rosenfeld MG, Glass CK, Hayashi Y (2001) Defect of histone acetyltransferase activity of the nuclear transcriptional coactivator CBP in Rubinstein-Taybi syndrome. Hum Mol Genet 10:1071-6

Partanen A, Motoyama J, Hui CC (1999) Developmentally regulated expression of the transcriptional cofactors/histone acetyltransferases CBP and p300 during mouse embryogenesis. Int J Dev Biol 43:487-94

Petrij F, Dauwerse HG, Blough RI, Giles RH, van der Smagt JJ, Wallerstein R, Maaswinkel-Mooy PD, van Karnebeek CD, van Ommen GJ, van Haeringen A, Rubinstein JH, Saal HM, Hennekam RC, Peters DJ, Breuning MH (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 37:168-76.

Petrij F, Giles RH, Dauwerse HG, Saris JJ, Hennekam RC, Masuno M, Tommerup N, van Ommen GJ, Goodman RH, Peters DJ, et al. (1995) Rubinstein-Taybi syndrome caused by mutations in the transcriptional co- activator CBP. Nature 376:348-51.

Roth JF, Shikama N, Henzen C, Desbaillets I, Lutz W, Marino S, Wittwer J, Schorle H, Gassmann M, Eckner R (2003) Differential role of p300 and CBP acetyltransferase during myogenesis: p300 acts upstream of MyoD and Myf5. Embo J 22:5186-96

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30:e57.

Tanaka Y, Naruse I, Maekawa T, Masuya H, Shiroishi T, Ishii S (1997) Abnormal skeletal patterning in embryos lacking a single Cbp allele: a partial similarity with Rubinstein-Taybi syndrome. Proc Natl Acad Sci U S A 94:10215-20

Ugai H, Uchida K, Kawasaki H, Yokoyama KK (1999) The coactivators p300 and CBP have different functions during the differentiation of F9 cells. J Mol Med 77:481-94

White SJ, Vink GR, Kriek M, Wuyts W, Schouten J, Bakker B, Breuning MH, den Dunnen JT (2004) Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses. Hum Mutat 24:86-92

Winter J, Lehmann T, Suckow V, Kijas Z, Kulozik A, Kalscheuer V, Hamel B, Devriendt K, Opitz J, Lenzner S, Ropers HH, Schweiger S (2003) Duplication of the MID1 first exon in a patient with Opitz G/BBB syndrome. Hum Genet 112:249-54

Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, Bronson RT, Li E, Livingston DM, Eckner R (1998) Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. Cell 93:361-72

Chapter 6

Fredman D.F., White S.J., Potter S., Eichler E.E., den Dunnen J.T., Brookes A.J. (2004). Complex SNP-related sequence variation within segmental genomic duplications. *Nat. Genet.* 36 (8):861-866.

(Colour images from this chapter can be seen in the appendix)

nature genetics

Complex SNP-related sequence variation in segmental genome duplications

David Fredman[1], Stefan J White[2], Susanna Potter[1], Evan E Eichler[3], Johan T Den Dunnen[2] & Anthony J Brookes[1,4]

There is uncertainty about the true nature of predicted single-nucleotide polymorphisms (SNPs) in segmental duplications (duplicons) and whether these markers genuinely exist at increased density as indicated in public databases. We explored these issues by genotyping 157 predicted SNPs in duplicons and control regions in normal diploid genomes and fully homozygous complete hydatidiform moles. Our data identified many true SNPs in duplicon regions and few paralogous sequence variants. Twenty-eight percent of the polymorphic duplicon sequences we tested involved multisite variation, a new type of polymorphism representing the sum of the signals from many individual duplicon copies that vary in sequence content due to duplication, deletion or gene conversion. Multisite variations can masquerade as normal SNPs when genotyped. Given that duplicons comprise at least 5% of the genome and many are yet to be annotated in the genome draft, effective strategies to identify multisite variation must be established and deployed.

Duplicons defined as being >1 kb with >90% similarity between copies comprise at least 5% of the human genome[1,2]. Their minimal extent has been defined[3], but the public human genome draft portrays duplicons neither accurately nor completely[4–6]. SNP databases report that SNPs are over-represented by a factor of ~2 in duplicon regions[3,7,8]. This is a minimum value, as SNP discovery efforts discard predicted variants from regions where densities are high or a duplicon is suspected[9,10]. Many or most duplicon SNPs may be nothing more than paralogous sequence variants (PSVs)[3,7,8]. Alternatively, gene conversion in duplicons may generate allelic diversity and SNP content[11,12]. Additionally, reduced selective pressure in duplicons may allow new mutations to increase in frequency more easily [13].

Initially, we undertook an *in silico* study of SNPs in duplicons to search for informative features. We noted an increased gene density in duplicons and observed that validated SNPs (65.2% of the dbSNP version used) were under-represented in duplicons compared with non-validated SNPs. Specifically, 3.7% (5.6% by two hit–two allele, 3.4% by cluster, 1.9% by frequency) of valid SNPs versus 13.1% of nonvalidated SNPs reside in the 4.5% of the genome comprised of duplicons. This could imply that duplicon SNPs are mostly PSVs, or it could reflect the difficulty of doing experiments with nonunique sequences.

We therefore devised an experiment to resolve PSVs from real SNPs. We used dynamic allele-specific hybridization (DASH)[14], which generates a DNA melting curve by heating an oligonucleotide probe duplexed with a PCR amplicon. Negative derivatives of these curves allow for direct comparisons of allele ratios in heterozygotes. Sample DNAs were from 16 normal Swedish females and 8 pathologically confirmed monospermic complete hydatidiform moles (CHMs)[15]. CHMs are fully homozygous genomes that allow distinction between true SNP alleles at a single genome locus (genotypes will always show single alleles) and PSV signals originating from multiple sites (genotypes will be 'heterozygote-like', including both alleles). The tested samples gave 98% power to detect alleles of 10% frequency[16]. We targeted 17 duplicons (**Table 1**) that fell into four broad classes according to their representation in the public genome assembly, their degree of sequence similarity and whether they seemed to be multicopy by analysis of whole-genome shotgun sequencing data (WSSD)[3]. We also included two genome regions known to be unique. For each tested region, we genotyped eight predicted SNPs that were outside known repeats as detected by RepeatMasker[17], as well as five other previously validated true SNPs of random location.

We knew that DASH would convert 90–95% of all true SNPs to useable assays[14], and we assumed that most copies of the duplicon targets would be amplified in the PCR (given the high sequence similarities of the tested duplicons). The derived results comprised various melting-curve patterns (**Fig. 1b**) that correspond to specific genetic structures (**Fig. 1a**). Overall, 107 markers were polymorphic and useable for our investigation, including 13 control markers that gave genotypes consistent with single-copy true SNPs (**Fig. 2a**). The 15 markers in duplicons that lacked WSSD support likewise produced signals consistent with true SNPs (**Fig. 2a**). This indicates that these unique genome regions were inappropriately assembled, leaving them as apparent duplicons in the public draft. It is estimated that >50% of duplicons represented in the genome draft are not real[3]. As illustrated by our data, SNP genotyping can provide an efficient means to identify these for targeted resolution.

[1]Center for Genomics and Bioinformatics, Karolinska Institute, Berzelius väg 35, S-171 77 Stockholm, Sweden. [2]Human and Clinical Genetics, Leiden University Medical Center, Wassenaarseweg 72, 2333 AL Leiden, the Netherlands. [3]Department of Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA. [4]Present address: Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK. Correspondence should be addressed to A.J.B. (anthony.brookes@cgb.ki.se).

Published online 11 July 2004; doi:10.1038/ng1401

Table 1 Target regions

Region	WSSD	NCBI	Chrom	ChromStart (bp)	ChromEnd (bp)	Size (bp)	Name	Dispersal
A	Dup	Unique	1	85,402,915	85,427,399	24,485	–	Unknown
B	Dup	Unique	2	89,796,158	89,812,623	16,466	–	Unknown
C	Dup	Unique	16	18,167,513	18,191,332	23,820	–	Unknown
D	Dup	Unique	16	69,832,810	69,854,823	22,013	–	Unknown
E	Dup	Dup <98%	7	75,865,780	75,891,118	25,339	–	Intra
F	Dup	Dup <98%	9	85,988,721	86,012,093	23,373	–	Inter
G	Dup	Dup <98%	10	46,657,428	46,672,624	15,197	–	Intra
H	Dup	Dup <98%	11	88,972,901	88,996,892	23,992	–	Intra
I	Dup	Dup <98%	16	32,022,851	32,039,556	16,706	–	Inter
J	Dup	Dup >98%	8	7,161,589	7,293,710	132,121	8p23	Intra
K	Dup	Dup >98%	15	20,852,650	20,890,966	38,316	HERC2	Intra
L	Dup	Dup >98%	15	30,161,462	30,293,362	131,900	CHRNA7	Intra
M	Dup	Dup >98%	16	16,603,367	16,682,029	78,662	LCR16a	Intra
N	Dup	Dup >98%	17	44,072,366	44,126,506	54,140	MS	Intra
O	Unique	Dup >98%	1	57,845,958	57,856,075	10,117	–	Intra
P	Unique	Dup >98%	11	133,555,034	133,578,684	23,650	–	Intra
Q	Unique	Dup >98%	12	51,307,117	51,382,529	75,412	–	Intra
R	Unique	Unique	16	21,560,883	21,636,826	75,943	–	Unique
S	Unique	Unique	22	20,825,861	20,875,861	50,000	–	Unique
T	Unique	Unique	Various	Random validated SNPs	–	Unique		

Coordinates are from the July 2003 NCBI assembly. These comprise 17 duplicons and additional controls, covering a total of 1 Mb, taken from 12 different chromosomes. The target regions were grouped into four broad classes: A–D, domains that are present uniquely in the NCBI assembly but that are indicated to be duplicons by WSSD; E–I, duplicated domains in the NCBI assembly having 90–98% sequence similarity and WSSD support; J–N, duplicated domains in the assembly with >98% similarity and WSSD support; O–Q, duplicated domains in the assembly with >98% similarity but no WSSD support. Regions R–T are unique control sequences.

Behavior of markers in WSSD-positive regions was substantially different from that of those in control regions (**Fig. 2a**,**b**). A full 91% (72 of 79) of duplicon assays gave apparent heterozygote signals in at least one CHM. To interpret the various genotype patterns, we established a classification schema (**Table 2**). Many duplicon markers behaved as real SNPs, residing either in unique sequence (7 of 79, 8.9%) or in one copy of a duplicon (32 of 79, 41%). This total (50%) equates to a SNP density that is equivalent to the genome average, as duplicons are enriched for predicted SNPs by a factor of 2 in public databases[3,7,8]. In addition, and contrary to previous evidence[3,7,8], only 23% (18 of 79) of duplicon markers behaved as PSVs. The remaining 28% (22 of 79) of predicted SNPs in duplicons were neither PSVs nor SNPs but gave complex genotyping patterns that have not been described before. We called this new form of polymorphism multisite variation (MSV).

When we assessed MSVs in CHMs, they generated either homozygous genotypes, indicative of SNPs, or apparently heterozygous signals, indicative of PSVs, (**Fig. 1b**). Two such signals are combined in diploid DNAs, and so MSVs gave genotypes in normal samples that

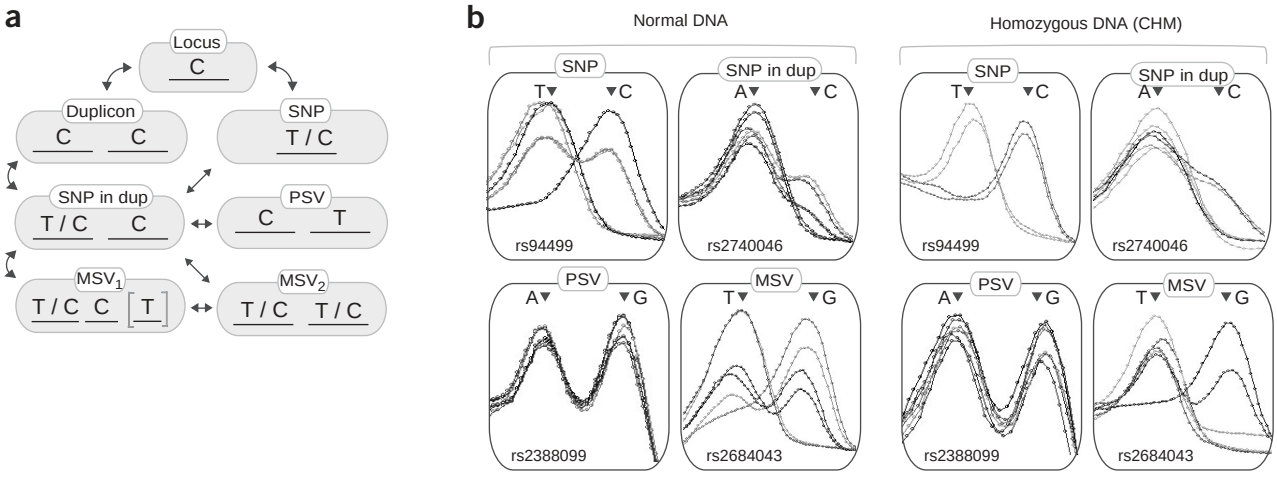

Figure 1 Genotyping patterns identifying evolutionary sequence states. (**a**) Evolutionary sequence changes from a monomorphic base to a polymorphic MSV. Arrows depict processes such as mutation, fixation, duplication, deletion and gene conversion. Most events are reversible. (**b**) Representative DASH genotyping patterns observed in normal and CHM samples for the corresponding structures in **a**. Each line shows the negative derivative of the melting curve of a probe-target duplex for one DNA sample. The temperature on the *x* axis ranges from 45 to 75 °C. Peaks marked by arrowheads indicate the presence of each particular allele as marked, with peak heights indicating the relative amount of each allele present in the tested DNA. Dup, duplicon.

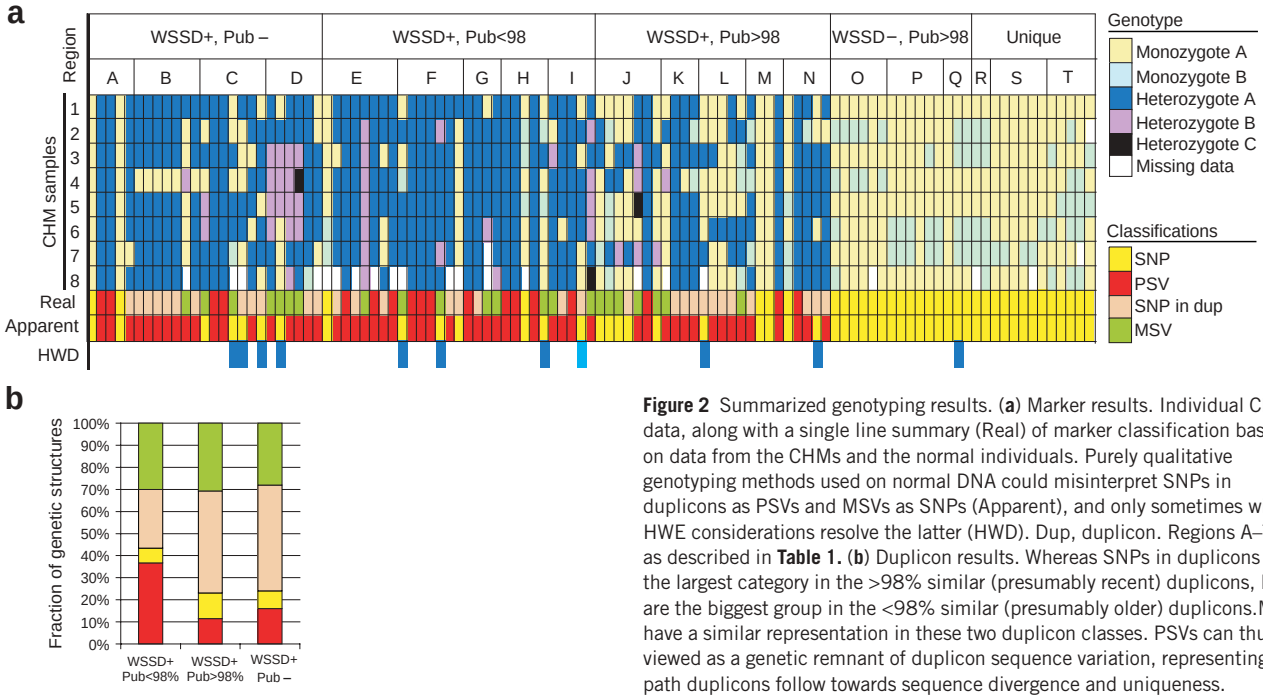

Figure 2 Summarized genotyping results. (**a**) Marker results. Individual CHM data, along with a single line summary (Real) of marker classification based on data from the CHMs and the normal individuals. Purely qualitative genotyping methods used on normal DNA could misinterpret SNPs in duplicons as PSVs and MSVs as SNPs (Apparent), and only sometimes will HWE considerations resolve the latter (HWD). Dup, duplicon. Regions A–T are as described in **Table 1.** (**b**) Duplicon results. Whereas SNPs in duplicons are the largest category in the >98% similar (presumably recent) duplicons, PSVs are the biggest group in the <98% similar (presumably older) duplicons.MSVs have a similar representation in these two duplicon classes. PSVs can thus be viewed as a genetic remnant of duplicon sequence variation, representing the path duplicons follow towards sequence divergence and uniqueness.

masqueraded as typical SNPs, but with variable allele ratios across individuals. These patterns may be explained as the sum of individual genotyping signals from various similar-sequence duplicon copies, with those duplicons themselves varying in the population. This variation may be due to (i) duplicon copy-number differences that lead to an increase, decrease or elimination of signals from different alleles that reside on the inserted or deleted duplicon copies (**Fig. 1a** ; MSV_1 pattern) or (ii) gene conversion events that lead to dispersion, mixing and perhaps homogenization of single-base alternatives across the various copies of a duplicon (**Fig. 1a**; MSV_2 pattern).

There is considerable evidence that gene conversion[18,19] and copy number variation[20,21] are active in subsets of duplicons. To evaluate the generality of these processes, we assessed sequences adjacent to 16 discovered MSVs (in nine duplicons) and two control SNPs for copy-number variation using multiplex ligation-dependent probe amplification (MLPA)[22,23]. We used another six control sequences for normalization. No CHM had more than about ten copies of any interrogated sequence (**Supplementary Fig. 1** online), and there was considerable evidence for copy-number variation in 50% (8 of 16) of cases (**Table 3**). Furthermore, sequences close to MSVs with a larger number of different allele ratios (as assessed by DASH) tended to report greater copy-number variability (**Supplementary Fig. 2** online). Thus, MSVs are a consequence (at least in part) of widespread duplicon copy-number variation. This interpretation is supported by Fosmid end-mapping data (E.E.E., unpublished results) and studies of copy-number differences related to disease[6,20,21,24]. Only some closely spaced markers showed correlated MLPA ratios (**Fig. 3**), however, indicating that there is substantial within-duplicon heterogeneity in this phenomenon.

Counting SNPs and MSVs together, at least two-thirds of predicted duplicon SNPs in public databases are polymorphic rather than PSVs. The one-third of these that are MSVs produce genotype patterns in diploid samples very similar to those of SNPs, other than having (sometimes subtle) allele ratio variability in heterozygotes. Genotyping technologies will need to detect this allele ratio variability to reliably identify MSVs. This raises a concern regarding whole-genome amplification procedures, which may distort these allele ratios. In pooled

Table 2 Identification of genomic structures by analysis of DASH genotypes for CHMs and normal DNA

Genetic structure	Material	Number of alleles	Genotypes	Het. allele ratios	Constraints
SNP	DNA	1 or 2	M, H, m	Fixed ratio	–
	CHM	1 or 2	M, m	–	–
SNP in duplication	DNA	1 or 2	M, H	2 different ratios	One DNA H ratio must match CHM ratio
	CHM	1 or 2	M, H	Fixed ratio	
PSV	DNA	2	H	Fixed ratio	Same H ratio in DNA and CHM
	CHM	2	H	Fixed ratio	
MSV	DNA	1 or 2	M, H, m	Variable ratio	–
	CHM	1 or 2	M, H, m	Variable ratio	–

Samples are either homozygous with respect to one allele (M or m) or apparently heterozygous (H). Single-locus SNPs produce consistent homozygous and heterozygous signals in normal individuals, and no heterozygotes in CHMs. For a true SNP present in one copy of a duplicon (SNP in duplicon), one of the alleles is additionally represented at the other duplicon version(s), generating a heterozygote signal in one or more CHM. In normal DNA, these completely lack one homozygote pattern and generate two distinctive heterozygote patterns with different allele ratios. PSVs render heterozygote signals of identical allele ratios in all tested samples. MSVs produce two or more heterozygote types in CHMs, three or more heterozygote types in normal DNA, or both homozygotes combined with at least one type of heterozygote in CHMs.

Table 3 MLPA analysis of 16 MSVs and two single-copy reference sequences

Nearest rs ID	Dup. region	Normalized MLPA ratios (triplicate means) CHM1	CHM2	CHM3	CHM4	CHM5	CHM6	CHM7	CHM8	s.d.	Copy-number variation
–	Unique	–	0.87	1.12	1.11	0.85	0.93	1.03	0.92	0.11	No
–	Unique	0.93	0.89	1.1	1.09	0.93	0.98	1.03	1.06	0.08	No
394595	B	1.16	1.05	0.97	0.63	0.91	1.01	–	1.04	0.18	Yes
2910545	C	1.13	1.01	1.01	1.00	0.94	0.93	0.93	1.04	0.07	No
1057729	D	1.28	1.22	0.85	0.85	0.77	0.86	1.17	1.02	0.2	Yes
2868008	D	1.28	1.17	0.83	0.92	0.73	0.89	–	0.96	0.19	Yes
2868007	D	1.35	1.18	0.89	0.78	0.74	0.93	1.00	1.14	0.21	Yes
2690641	E	1.04	0.94	1.09	1.16	0.88	0.91	–	0.82	0.12	No
505235	F	1.03	1.02	1.04	0.98	0.96	0.96	0.94	1.06	0.04	No
1836885	H	1.01	0.98	0.94	0.96	1.11	0.93	0.92	1.16	0.09	No
964055	I	1.05	1.18	0.95	1.01	1.01	1.18	0.72	–	0.16	Yes
2939843	I	1.04	1.05	0.92	1.11	1.07	0.94	0.85	1.03	0.09	No
2684043	J	1.15	1.1	1.02	1.16	0.79	0.97	0.92	0.89	0.13	No
2740736	J	1.17	1.1	1.21	1.24	0.7	0.82	1.03	0.74	0.22	Yes
2740083	J	1.03	1.1	1.01	1.11	0.91	0.89	0.97	0.98	0.08	No
746659	J	1.37	1.3	1.00	–	0.73	0.83	0.95	0.78	0.25	Yes
296349	K	0.99	1.00	1.12	1.06	0.89	1.02	0.81	1.1	0.1	No
380880	K	0.75	1.26	1.05	0.86	1.00	1.08	0.93	1.08	0.15	Yes

Half of the MSV sequences show substantial evidence of copy-number variation. The remainder, including the two reference sequences, either have a fixed number of sequence copies or have a relative difference below the threshold of detection (s.d. < 0.15 across the eight CHMs).

DNAs, because individual allele ratio information is lost, it will be impossible to identify MSVs. To detect MSVs in routine practice, CHMs or haploid genomes could be included in upstream assay validation routines. Mendelian inheritance tests might assist but will not be effective for MSVs involving intrachromosomal duplicons. Consideration of Hardy-Weinberg equilibrium (HWE) may help, but analysis will not be fool-proof if the 'single allele' and 'two allele' haploid signals for MSVs are consistent with HWE in the overall population. Beyond MSVs, SNPs residing in one copy of a duplicon may also be mis-scored, because the additional signal component from the non-polymorphic duplicon would make one of the two homozygotes appear to be a heterozygote.

How duplicon markers might be scored disregarding heterozygote allele ratio differences (which many methods tend to do) and without using CHMs is an important question. To explore this, we re-examined our total data set, ignoring these two pieces of evidence. This analysis incorrectly indicated an abundance of PSVs in duplicons (**Fig. 2a**; consistent with previous interpretations[3,7,8]), with only half of the apparent SNPs that were truly MSVs deviating from HWE (32 chromosomes; $P < 0.01$). Consistent with this, as of April 2004, four of the MSV markers we report are classified as experimentally validated SNPs with genotype data in dbSNP. Additionally, one PSV is described in current HapMap data, where it is listed as a monomorphic SNP.

In light of these considerations, we reviewed recent genotyping data from our production facility, which uses DASH. We considered almost 800 markers from different studies that used various SNP selection criteria, leaving 45 targets in duplicons. The initial validation (assessing 16–96 control individuals and considering HWE), identified 15 monomorphic single-allele signals and classified the remaining 30 markers as follows: 12 (40%) unique SNPs, 8 (27%) SNPs in one copy of a duplicon, 4 (13%) PSVs and 6 (20%) MSVs. Five of the unique SNPs had been used for production genotyping of 1,600–2,000 individuals, and only after observing several tens of heterozygote-like signals did it become clear that two of these were actually MSVs and another was a SNP in a duplicon. For the two MSVs, if samples that reported two alleles had been scored as heterozygotes (regardless of allele ratios), then the total genotype data were in complete HWE ($P = 0.115$ and 0.357).

In conclusion, our study identifies MSVs as a new form of genome polymorphism. Careful laboratory practice should often recognize MSVs as aberrant markers, and MSVs may underlie the considerable fraction of markers that fail HWE. But some MSVs are probably being interpreted and used as unique SNPs, and HWE will not always identify these, even if large sample numbers are used. More generally, MSVs (or rather duplicon copy-number variation and duplicon gene

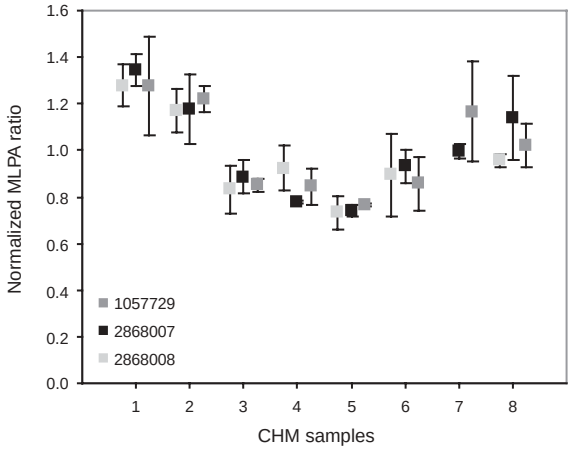

Figure 3 MLPA data for eight CHMs across three consecutive loci. These span 3.4 kb on chromosome 16 (**Table 1**). The graph shows mean ± 2 s.e.m. values across replicate experiments. For all three probes, CHMs 1 and 2 have ratios ~50% higher than those of CHMs 3–6 (a 3:2 relative copy-number difference). CHMs 7 and 8 are harder to classify because of a wider spread between replicates, but they seem to overlap mostly with CHMs 1 and 2. This result is in full agreement with observed genotyping data, in that the MLPA ratios correlate with the observed DASH heterozygote classes.

conversion processes) might underlie some common phenotypic differences between individuals. We therefore suggest that MSVs should be specifically targeted for evaluation in disease and pharmacogenomics research.

METHODS

***In silico* detection of SNP and duplicate region overlap.** Duplicon regions were as previously defined[3], derived from alignments of sequence fragments from the National Center for Biotechnology Information (NCBI) human genome assembly[2] combined with sequence read depth analysis of WSSD from the Celera human genome assembly[1]. We downloaded duplication sequence and June 2002 NCBI assembly locations from the human paralogy database. We used the most complete SNP list available with June 2002 NCBI assembly locations (dbSNP[25] build 112; 2,337,575 SNPs) and updated the annotation with data from dbSNP build 119. We downloaded gene lists from Ensembl[26]. We loaded the locations into a MySQL database and identified overlaps of chromosomal locations through SQL queries issued from a set of Perl scripts. Total counts were nonredundant so that each SNP was counted only once in our analysis, even if it mapped to multiple genome locations (duplicon paralogs).

We searched for any dbSNP annotations that might uniquely characterize duplicon SNPs. We tested the following factors: (i) validation (by cluster, 'SNP discovered by at least two different methods'; by two hit–two allele, 'SNP must be observed twice, in two different DNA samples which must have produced two alleles'; by frequency, 'allele frequency data available for SNP'); (ii) source (which discovery effort generated the SNP); and (iii) frequency of minor allele. Map weight was excluded from consideration, as these SNPs are, by definition, in repetitive sequence, and for any SNP in a duplicon with a map weight <2, the map weight is due to the difference in alignment methods and scoring thresholds between duplicon detection and SNP mapping.

DASH. We carried out DASH experiments, designed with DFold[27] software, using standard protocols as previously described[14]. Oligonucleotide sequences for all assays are available on request. We carried out PCR reactions in 20-μl volumes, containing 25–250 pg μl^{-1} of genomic DNA. We used DASH software (Thermo Hybaid) to visualize denaturation events by plotting the negative derivative of the fluorescence versus temperature profile. Genotypes were scored manually and blindly. We reviewed independent duplicate experiments for 25% of assays as a control for assay reproducibility and found scoring to be consistent across runs. We assessed deviation from HWE for individual markers using the χ^2 statistic ($P < 0.01$). We excluded 32% of assays across all regions from analysis; 3.2% (5 of 157) assays produced no PCR product, and 29% (13 of 45) of those in nonduplicon regions (control regions plus falsely predicted duplicons with support only from the public assembly) and 18% (20 of 112) of those in real duplicons gave no indication of polymorphism. These percentages were evenly distributed between different sources of SNPs (data not shown) and are consistent with what is generally found for public database SNPs[28]. Further, 4.4% (2 of 45) of assays in nonduplicon regions and 8.9% (10 of 112) of those in real duplicons were of low quality, and many gave three distinct allele signals. This is probably due to additional but uncharacterized sequence variants in the probe hybridization region at positions other than that being tested. This left 107 informative polymorphic assays covering all tested regions. Complete genotyping information is available on request.

The number of tested DNA samples affects the certainty of classification. Also, misclassifications may arise if a PCR does not amplify multiple duplicon copies with similar or equal efficiency. We cannot estimate the cumulative size of these biases, but both will tend to cause an overestimation of the number of PSVs at the expense of MSVs and suggest monomorphic sites over SNPs, SNPs over SNPs in duplicons and SNPs in duplicons over MSVs. Therefore, our PSV estimate must be considered a maximum, and our MSV estimate a minimum.

MLPA. We designed MLPA probes based on consensus sequences derived from global alignments of duplicated segments. Probes were localized in regions immediately flanking MSV variants identified by the DASH experiment. To avoid allelic discrimination and ensure specificity, no polymorphism or sequence differences between duplicon copies were allowed within 6 bp on either side of the ligation site (sequences available on request). The specific priming sequences in the 5′ ends of the half-probes allowed multiplex amplification with either the MLPA primers[23] or the MAPH primers[29]. Resulting PCR products had a minimal size difference of 2 bp, with the products ranging in size from 80 bp to 125 bp. The forward primer of each pair was fluorescently labeled (MLPAF-FAM or MAPHF-HEX), allowing probes to be distinguished also on the basis of color. Each color set included three control probes from known single-copy regions, for normalization purposes, and we added two other single-copy probes to one of the sets as controls for copy-number variation. All oligonucleotides were combined in a single mix at a final concentration of 4 fmol μl^{-1}.

We carried out the MLPA reaction essentially as described[23]. We heated 100 ng of DNA at 98 °C for 5 min. After cooling to 25 °C, we added 1.5 μl of probe mix and 1.5 μl of SALSA hybridization buffer to each sample, denatured them at 95 °C for 2 min and then hybridized them for 16 h at 60 °C. Ligation was done at 54 °C by adding 32 μl of ligation mix. After 10–15 min, we stopped the reaction by heat inactivation at 95 °C for 5 min. We carried out PCR amplification for 30 cycles in a final volume of 25 μl. In addition to the reagents described[23], we added MAPH-F and MAPH-R to each PCR reaction to a final concentration of 100 nM. From each PCR reaction, we mixed 1–2 μl of product with 10 μl (Hi Di) of formamide and 0.1 μl of ROX 500 size standard (Applied Biosystems) in a 96-well plate. We separated products by capillary electrophoresis on the ABI 3700 DNA sequencer (Applied Biosystems).

MLPA data analysis. We retrieved peak data using GeneScan (Applied Biosystems) and exported it to Excel (Microsoft) and SPSS 10 (SPSS) for further analysis. We obtained signals for 84% (16 of 19) of designed assays. We obtained a ratio for each of the working probes by dividing the height of the corresponding peak by the sum of the heights of three control peaks of the same color. We did three replicate experiments across all CHM samples, calculated the average value of the three ratios and discarded the results if the s.d. was >20%. This eliminated 6 of 144 measurements (4.2%). We then normalized the data for each probe around 1.0 by dividing by the average of the remaining values.

URLs. The Human Paralogy Server is available at http://humanparalogy.gene.cwru.edu/. The NCBI dbSNP is available at http://www.ncbi.nlm.nih.gov/SNP/. The International HapMap Project is available at http://www.hapmap.org/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank R.J. Fisher and M. Seckl for CHM DNA samples and R.A. Clark, S. Sawyer and C. Lagerberg for technical assistance. Funding was provided by Pfizer Corporation and Stiftelsen för Kompetens-och Kunskapsutveckling (to D.F. and A.J.B.) and by the US National Institutes of Health (to E.E.E.).

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Received 23 April; accepted 22 June 2004
Published online at http://www.nature.com/naturegenetics/

1. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
2. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Bailey, J.A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
4. Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J. & Eichler, E.E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).
5. Istrail, S. *et al.* Whole-genome shotgun assembly and comparison of human genome assemblies. *Proc. Natl. Acad. Sci. USA* **101**, 1916–1921 (2004).
6. Shaw, C.J. & Lupski, J.R. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum. Mol. Genet.* **13**, R57–R64 (2004).
7. Estivill, X. *et al.* Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome. *Hum. Mol. Genet.* **11**, 1987–1995 (2002).
8. Cheung, J. *et al.* Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol.* **4**, R25 (2003).

9. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
10. Tsui, C. *et al.* Single nucleotide polymorphisms (SNPs) that map to gaps in the human SNP map. *Nucleic Acids Res.* **31**, 4910–4916 (2003).
11. Hurles, M.E. Gene conversion homogenizes the CMT1A paralogous repeats. *BMC Genomics* **2**, 11 (2001).
12. Hurles, M. Are 100,000 "SNPs" useless? *Science* **298**, 1509 (2002).
13. Conant, G.C. & Wagner, A. Asymmetric sequence divergence of duplicate genes. *Genome Res.* **13**, 2052–2058 (2003).
14. Prince, J.A. *et al.* Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res.* **11**, 152–162. (2001).
15. Sebire, N.J., Fisher, R.A. & Rees, H.C. Histopathological diagnosis of partial and complete hydatidiform mole in the first trimester of pregnancy. *Pediatr. Dev. Pathol.* **6**, 69–77 (2003).
16. Kruglyak, L. & Nickerson, D.A. Variation is the spice of life. *Nat. Genet.* **27**, 234–236 (2001).
17. Smit, A.F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).
18. Jeffreys, A.J. & May, C.A. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nat. Genet.* **36**, 151–156 (2004).
19. Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
20. Hollox, E.J., Armour, J.A. & Barber, J.C. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. *Am. J. Hum. Genet.* **73**, 591–600 (2003).
21. Locke, D.P. *et al.* BAC microarray analysis of 15q11-q13 rearrangements and the impact of segmental duplications. *J. Med. Genet.* **41**, 175–182 (2004).
22. White, S.J. *et al.* Two-colour MLPA; detecting genomic rearrangements in hereditary multiple exostoses. *Hum. Mutat.* **24**, 86–92 (2004).
23. Schouten, J.P. *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30**, e57 (2002).
24. Lucito, R. *et al.* Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res.* **13**, 2291–2305 (2003).
25. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
26. Birney, E. *et al.* Ensembl 2004. *Nucleic Acids Res.* **32**, D468–D470 (2004).
27. Fredman, D., Jobs, M., Stromqvist, L. & Brookes, A.J. DFold: PCR design that minimizes secondary structure and optimizes downstream genotyping applications. *Hum. Mutat.* **24**, 1–8 (2004).
28. Carlson, C.S. *et al.* Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat. Genet.* **33**, 518–521 (2003).
29. White, S. *et al.* 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 (2002).

Summary

The work presented in this thesis describes the further development and application of two techniques for the detection of copy number changes in genomic DNA, namely Multiplex Amplifiable Probe Hybridization (MAPH) and Multiplex Ligation-dependent Probe Amplification (MLPA). Both techniques allow the quantitative analysis of up to 50 loci in a single reaction.

The initial work involved analysis of the *DMD* gene, mutations in which cause Duchenne and Becker Muscular Dystrophy (DMD/BMD). MAPH was used for the detection of deletions and duplications in Duchenne Muscular Dystrophy patients (Chapter 2.1). Using this approach several small duplications and deletions were identified. It is especially noteworthy that a duplication of exon 2 was described for the first time, yet is the single most common duplication found. A further investigation of duplications in the *DMD* gene, using both MAPH and MLPA, is described in chapter 2.2. Among several novel duplications found were several non-contiguous duplications and triplications, as well as a duplication that lead to a deletion at the mRNA level. Such cases emphasize the importance of screening all exons within the gene for rearrangements, and show that the reading-frame rule used for the prediction of disease progression has to be used with caution when duplications are involved.

Although detecting the causative mutation is obviously critical for the patient, determining the carrier status of relatives is also important. This is made considerably easier when the mutation to look for has already been identified in the patient. We took advantage of this in developing a rapid approach for testing female relatives of affected patients (Chapter 2.3).

There are many other forms of muscular dystrophy known. Among these are the recessive forms of the limb-girdle muscular dystrophies. In contrast to DMD/BMD, the mutation spectra of these diseases show that small mutations affecting only one or a few nucleotides are almost exclusively the cause of the disease. There are, however, situations when mutations can not be found at all, or only in one allele. In one such patient a single exon deletion was found in the sarcoglycan gamma (SGCG) gene. The same deletion had been described in two patients from another country, and breakpoint analysis showed the identical, complex rearrangement. The fact that this mutation segregated with the same haplotype in the three, unrelated individuals strengthened the argument for a common ancestor (Chapter 3).

Although MAPH is ideally suited for the analysis of exonic deletions and duplications, it can also be applied to the detection of larger rearrangements. Such changes, known to be involved in developmental delay, were investigated in a cohort of patients with mental retardation. Along

with the expected changes involving the subtelomeric regions, several interstitial rearrangements were found (Chapter 4). The discovery of affected regions in these patients will hopefully lead to the identification of the genes responsible for the development delay.

Given the relative simplicity of the assay, MLPA is becoming the technique of choice in diagnostic settings. A drawback of the technique, however, is the costly and time-consuming cloning required for generating probes. To circumvent this we developed a two color method using synthetic oligonucleotides only, and tested it by screening the *EXT1* and *EXT2* genes, involved in Hereditary Multiple Exostoses (HME). Using this we showed that partial gene deletions do occur in both genes, and account for 5-8% of all HME cases (Chapter 5.1). The same approach was also used to detect a deletion in the *EP300* gene in a patient with Rubinstein-Taybi Syndrome (RSTS), showing for the first time that mutations in this gene also cause the disease (Chapter 5.2). This simple manner of probe development should facilitate quantitative analysis of other genes in the future.

It is known that ~5% of the genome is composed of duplicons, which are highly similar sequences present in more than one copy. These are difficult to characterize by sequencing, and complicate Single Nucleotide Polymorphism (SNP) analysis. A number of apparently SNP-containing loci within these regions were analyzed by MLPA, which showed for several of the cases that differences seen between samples were the result of variations in copy number (Chapter 6). These multisite variants (MSVs) may appear as SNPs unless correctly analyzed, and application of techniques such as MLPA will be necessary to provide a clearer picture of the contribution of MSVs to genomic variation. Techniques such as MAPH and MLPA are therefore not only applicable to the search for intragenic deletions and duplications in disease, but also to the high resolution analysis of genomic variation.

Samenvatting

Het onderzoek dat in dit proefschrift uiteengezet wordt beschrijft de verdere ontwikkeling en toepassing van twee technieken voor het detecteren van veranderingen in kopie aantal in genomisch DNA, namelijk Multiplex Amplifiable Probe Hybridisation (MAPH) and Multiplex Ligation-dependent Probe Amplification (MLPA). Beide technieken maken het mogelijk tot 50 loci kwantitatief te analyseren in één enkele reactie.

Aanvankelijk richtte het onderzoek zich op de analyse van het *DMD* gen, waarin mutaties Duchenne en Becker spierdystrofie (DMD/BMD) veroorzaken. MAPH werd gebruikt voor de detectie van deleties en duplicaties in DMD patiënten (Hoofdstuk 2.1). Met deze aanpak werden meerdere kleine deleties en duplicaties gevonden. Het is vooral opmerkelijk dat voor het eerst een duplicatie van exon 2 werd gevonden terwijl dit de meest voorkomende duplicatie bleek te zijn. Een diepgaand onderzoek naar duplicaties in het *DMD* gen, gebruik makend van zowel MAPH als MLPA, is beschreven in hoofdstuk 2.2. Onder de verschillende nieuwe duplicaties die werden gevonden bevonden zich ook enkele niet-aaneengesloten duplicaties en triplicaties, als ook een duplicatie die leidt tot een deletie in het mRNA. Zulke gevallen benadrukken het belang van het screenen van alle exonen in het gen voor herrangschikkingen, en tonen aan dat de leesraam regel voor het voorspellen van ziekte verloop voorzichtig gebruikt moet worden wanneer het duplicaties betreft.

Hoewel het detecteren van de onderliggende mutatie uiteraard essentieel is voor de patiënt, is het bepalen van eventueel dragerschap van familieleden ook zeer belangrijk Dit is een stuk eenvoudiger indien de mutatie waarnaar gezocht moet worden in de patiënt reeds is gevonden. We hebben hiervan gebruik gemaakt voor het ontwikkelen van een snelle aanpak voor het testen van vrouwelijke familieleden van aangedane patiënten (Hoofdstuk 2.3).

Er zijn vele andere vormen van spierdystrofie bekend. Hieronder bevinden zich de recessieve vormen van limb-girdle spierdystrofieën. In tegenstelling tot DMD/BMD, tonen de mutatie spectra dat kleine mutaties van één of enkele nucleotiden vrijwel de enige oorzaken zijn van deze ziekten. Er zijn echter gevallen waarbij geen enkele mutatie gevonden kan worden, of in slechts één van de twee allelen. In één zo'n patiënt werd een deletie gevonden van een heel exon in het sarcoglycan gamma (*SGCG*) gen. Dezelfde deletie was eerder beschreven in twee patiënten uit een ander land, en breukpunt analyse toonde dezelfde, complexe herrangschikking. Het feit dat deze mutatie overerft met hetzelfde haplotype in deze drie niet-verwante individuen versterkt het argument voor het bestaan van een gezamenlijke voorouder (Hoofdstuk 3).

Hoewel MAPH bij uitstek geschikt is voor de analyse van exon deleties en duplicaties, kan het ook gebruikt worden voor het detecteren van grotere herrangschikkingen. Zulke veranderingen die een rol

spelen in vertraging van de ontwikkeling werden onderzocht in een group patiënten met mentale retardatie. Naast de verwachtte veranderingen van de subtelomere gebieden werden verschillende herrangschikkingen in de gebieden daartussen gevonden (Hoofdstuk 4). Het ontdekken van aangedane gebieden in deze patiënten leidt hopelijk tot de identificatie van genen die verantwoordelijk zijn voor de ontwikkelings achterstand. Dankzij de relatieve eenvoud van de analyse wordt MLPA de techniek van keuze in de diagnostiek. Een nadeel van de MLPA techniek is echter dat het kloneren van de benodigde probes duur en tijdrovend is. Om dit te omzeilen hebben we een twee kleuren methode ontwikkeld waarbij we alleen gebruik maken van synthetische oligonucleotides, deze methode hebben we uitgeprobeerd door de *EXT1* en *EXT2* genen te screenen die een rol spelen bij Erfelijke Multiple Exostosen (HME). Hiermee hebben we aangetoond dat gedeeltelijke gen deleties in beide genen voorkomen, en verantwoordelijk zijn voor 5-8 % van alle HME gevallen (Hoofdstuk 5.1). Dezelfde methode werd ook gebruikt om een deletie in het *EP300* gen in een patiënt met Rubinstein-Taybi Syndroom (RSTS) aan te tonen, hiermee voor het eerst aantonend dat mutaties in dit gen ook deze ziekte veroorzaken (Hoofdstuk 5.2). Deze eenvoudige methode van probes maken zou de kwantitatieve analyse van andere genen in de toekomst moeten vereenvoudigen. Het is bekend dat ~5 % van het genoom bestaat uit duplicons, dit zijn sterk overeenkomstige sequenties die in meer dan één kopie aanwezig zijn. Deze zijn moeilijk te analyseren door middel van sequencen en bemoeilijken Single Nucleotide Polymorphism (SNP) analyses. Een aantal schijnbaar SNP bevattende plaatsen in deze gebieden werden geanalyseerd met MLPA, waarmee werd aangetoond dat in meerdere gevallen de verschillende patronen tussen monsters te wijten waren aan verschillen in kopie aantal (Hoofdstuk 6). Deze ‘multisite variants’ (MSVs) kunnen zich voordoen als SNPs tenzij ze correct worden geanalyseerd, waarbij toepassing van een techniek zoals MLPA noodzakelijk is om een beter beeld te krijgen van de bijdrage van MSVs aan genoom variatie.

Technieken zoals MAPH en MLPA zijn daarom niet alleen geschikt voor het zoeken naar deleties en duplicaties in ziektegenen maar ook voor de hoge resolutie analyse van genoom variatie.

Publications

Articles

Wilson T., Wards B.J., White S.J., Skou B., de Lisle G.W., Collins D.M. (1998). Production of avirulent *Mycobacterium bovis* strains by illegitimate recombination with DNA fragments containing an interrupted *ahpC* gene. *Tubercle and Lung Disease* **78**, 229-235.

Cavaignac S.M., White S.J., de Lisle G.W., Collins D.M. (2000). Construction and screening of *Mycobacterium paratuberculosis* insertional libraries. *Archives of Microbiology* **173**, 229-231.

White S., Kalf M., Liu Q., Villerius M., Engelsma D., Kriek M., Vollebregt E., Bakker B., van Ommen G.J., Breuning M.H., den Dunnen J.T. (2002). Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. *American Journal of Human Genetics.* **71** (2):365-374.

White S.J., Sterrenburg E., van Ommen G.J., den Dunnen J.T., Breuning, M.H. (2003). An alternative to FISH: detecting deletion and duplication carriers within 24 hours. *Journal of Medical Genetics* **40** (10):e113.

Engels H., Ehrbrecht A., Zahn S., Bosse K., Vrolijk H., White S., Kalscheuer V., Hoovers J.M., Schwanitz G., Propping P., Tanke H.J., Wiegant J., and Raap A.K. (2003). Comprehensive analysis of human subtelomeres with combined binary ratio labelling fluorescence in situ hybridisation. *European Journal of Human Genetics* **11**, 643-651.

Kriek M., White S.J., Bouma M.C., Dauwerse H.G., Hansson K.B., Nijhuis J.V., Bakker B., van Ommen G.J., den Dunnen J.T., and Breuning M.H. (2004). Genomic imbalances in mental retardation. *Journal of Medical Genetics* **41**, 249-255.

White S.J., Vink G.R., Kriek M., Wuyts W., Schouten J., Bakker B., Breuning M.H., and den Dunnen J.T. (2004). Two-color multiplex ligation-dependent probe amplification: Detecting genomic rearrangements in hereditary multiple exostoses. *Human Mutation* **24**, 86-92.

Fredman D., White S.J., Potter S., Eichler E.E., den Dunnen J.T., and Brookes A.J. (2004). Complex SNP-related sequence variation in segmental genome duplications. *Nature Genetics* **36**, 861-866.

Vink G.R., White S.J., Gabelic S., Hogendoorn P.C.W., Breuning M.H., Bakker B. Mutation screening of EXT1 and EXT2 by direct sequence analysis and MLPA in patients with Multiple Osteochondromas: Splice site mutations and exonic deletions account for more than half of the mutations. *European Journal of Human Genetics;* in press.

White S.J., Uitte de Willige S., Verbove D., Politano L., Ginjaar H.B., Breuning M.H., den Dunnen, J.T. Sarcoglycanopathies and the risk of undetected deletion alleles in diagnosis. Submitted for publication.

Roelfsema J.H., White S.J., Arıyürek Y., Lacombe D., Bartholdi D., Niedrist D., Papadia F., den Dunnen J.T., van Ommen G.J., Breuning M.H., Hennekam R.C., Peters D.J.M. Genetic Heterogeneity in Rubinstein-Taybi syndrome: mutations in the *CBP* and *EP300* gene are both disease causing. Submitted for publication.

Book Chapter

White S.J., Breuning M.H., and den Dunnen J.T. Detecting copy number changes in genomic DNA - MAPH and MLPA. Cytometry 4ed, Volume 75 of the *Methods in Cell Biology*, in press.

Curriculum vitae

The author of this PhD thesis was born in Lower Hutt, New Zealand on July 25, 1973. He attended Upper Hutt College from 1986 until 1990, and started his university education at Victoria University of Wellington in 1991. He graduated with a BSc in Physiology and Biochemistry in 1993, and a BSc (Hons) in Biochemistry and Molecular Biology in 1994. Following two months touring Europe he started work at AgResearch in Upper Hutt in June 1995, working as technician on mycobacterial molecular biology until February 1999. The allure of Europe remained and he moved there in April 1999. In August of the same year he started work as a technician in the Center of Human and Clinical Genetics at Leiden University under the supervision of Professor Martijn Breuning and Dr. Johan den Dunnen, where the work described in this thesis was carried out. In January 2005 he started a post-doctoral postion in the same department working on the ZonMw funded project “Structural variation of the human genome in health and disease”.