
2

Methods to detect CNVs in the human genome

Emmelien Aten, Stefan J. White, Margot E. Kalf, Rolf H.A.M. Vossen, Helene H. Thygesen, Claudia A. Ruivenkamp, Marjolein Kriek, Martijn H.B. Breuning, Johan T. den Dunnen

Cytogenet Genome Res. 2008;123(1-4):313-21.

Abstract

The detection of quantitative changes in genomic DNA, i.e. deletions and duplications or Copy Number Variants (CNV), recently gained a considerable interest. First, detailed analysis of the human genome showed a surprising amount of CNV, involving thousands of genes. Second, it was realised that the detection of CNVs as a cause of genetic disease was often neglected, but should be an essential part of a complete screening strategy. In both cases new efficient CNV screening methods, covering the entire range from specific loci to genome-wide, were behind these developments. This paper will briefly review the methods that are available to detect CNVs, discuss their strong and weak points, show some new developments and look ahead. Methods covered include microscopy, fluorescence in situ hybridisation (incl. fiber-FISH), Southern blotting, PCR-based methods (incl. MLPA), array technology and massive parallel sequencing. In addition we will show some new developments, incl. a 1400-plex CNV bead assay, fast-MLPA (from DNA to result in ~ 6h) and a simple Melting Curve Analysis assay to confirm potential CNVs. Using the 1400-plex CNV bead assay, targeting selected chromosomal regions only, we detected confirmed rearrangements in 9% of 320 mental retardation patients studied.

Introduction - Methods to detect CNVs

Microscopy

The first report describing a CNV in the human genome was that by Leujeune et al. (1959); using a simple microscope he discovered that children with Down's syndrome have an extra copy of chromosome 21 ("trisomy 21"). Later specific methods were developed to generate chromosome-banding patterns. This simplified the discrimination of individual chromosomes as well as resolution inside a chromosome and facilitated the identification of inter and intra-chromosomal rearrangements. Together these tools were used to identify many other rearrangements in relation to genetic disease and cancer, especially numeric changes (aneuploidies), Structural Variants (SVs; translocations, inversions, insertions, transpositions) and large CNVs (>5-10 Mb deletions and duplications).

Resolution of microscopic chromosome analysis was further improved using in-situ hybridisation, initially using radioactive (ISH) and later fluorescent labelled probes (FISH, Landegent et al. (1985)). Nowadays FISH, especially using multiple probes labelled in different colours, is widely used in clinical diagnostics as a screening technology to confirm the presence of CNVs and SVs in either patients or carriers. Using probes close to the telomere of human chromosomes, Flint and co-workers were able to detect the presence

of CNVs in a significant fraction of patients with mental retardation and an apparently normal karyotype (Knight et al. (2000)).

Depending on the question, metaphase chromosome spreads are most commonly used. However, when a higher resolution is required (closely spaced probes) or when larger numbers of cells need to be counted (cancer), FISH can also be performed on interphase nuclei. Ultimate resolution is obtained by fiber-FISH, where probes are visualised on stretched single DNA fibers (Florijn et al. (1995)). Fiber-FISH is currently the method preferred to precisely determine the genomic structure of complex CNVs (Perry et al. (2007)).

Southern blotting

Although Southern blotting is a versatile tool for the detection of CNV, it has not been widely used. The main problems associated are the workload involved (DNA digestion, electrophoresis, blotting, hybridisation and exposure) and the fact that quantitative analysis is not simple; it demands technical experience and high-quality results. The exception has been the analysis of male samples for X-linked diseases where CNVs are frequent, e.g. in Duchenne and Becker muscular dystrophy (D/BMD) where hundreds of deletions have been revealed using Southern blotting (White and Den Dunnen (2006)). Interestingly, although duplications were described to be present in 5-10% of D/BMD patients early on (Den Dunnen et al. (1989)), only recently, using newly developed techniques, do we see that all screening studies report duplications (White and Den Dunnen (2006)). Similarly for autosomal diseases, e.g. breast cancer (Petrij-Bosch et al. (1997)) and Alzheimer (Rovelet-Lecrux et al. (2006)), it often took many years after identification of the disease gene before CNVs were first reported. These examples indicate that, although Southern blotting is a good tool for CNV detection, it is probably not the simplest tool to apply.

In specific cases, Southern blotting has been applied more widely. This includes diseases where specific CNVs are recurrent, facilitating the design of an assay targeting the unique breakpoint fragments, e.g. in alpha- and beta-thalassemia (Craig et al. (1994)). Another example is Southern blotting in combination with Pulsed-Field Gel-Electrophoresis (PFGE), allowing CNV detection at great distance from a specific probe (Den Dunnen et al. (1987)) or to determine the number of repeat sequences inside a large DNA fragment like in facioscapulohumeral muscular dystrophy (FSHD, Van Der Maarel et al. (1999)).

PCR-based methods

Because of its ease of use, PCR has been the most widely applied method to screen for CNVs. Since quantitative PCR is not that simple, PCR methods initially were used to confirm CNVs indirectly, i.e. by amplifying across unique deletion or duplication breakpoints, e.g. in delta/beta-thalassemia and hereditary persistence of fetal hemoglobin (Craig et al. (1994)).

Since the introduction of real-time PCR, facilitating simplified quantification, it has been applied as quantitative-PCR (qPCR) to directly screen for CNVs, discriminating 1 versus 2 (deletion) or 2 versus 3 copies (duplication). The main disadvantage of qPCR approaches is that only one fragment can be analysed and this fragment should be inside the CNV to get a positive result. Since CNVs in disease genes are rarely identical but can cover any part of the gene, several qPCRs will be required to perform a complete CNV screen. To bypass this problem, multiplex-qPCR approaches have been designed, for each locus tested yielding a fragment with either a unique length and/or colour. Except for again X-linked diseases to detect patients or carrier females (Beggs et al. (1990), Yau et al. (1996)), these assays have not found widespread application.

Only recently more powerful PCR-based methods have been developed, including Multiplex Amplifiable Probe Hybridization (MAPH, Armour et al. (2000)), Multiplex Ligation-dependent Probe Amplification (MLPA, Schouten et al. (2002)) and Multiplex Amplicon Quantification (MAQ, Suls et al. (2006)). These methods largely circumvent the inherent problems of multiplex-PCR, i.e. each fragment having different amplification characteristics with overall smaller fragments give higher yields. These methods achieve this by combining the advantages of an optimised primer design (incl. universal amplification primers), carefully determined primer concentrations and specific primer-mixes to achieve uniform amplification yields. Of these methods, MLPA has been by far the most successful, at the time of writing (May 2008) giving over 260 hits in PubMed. The winning features undoubtedly include its ease of use, the fact that standard laboratory equipment can be used (PCR and capillary electrophoresis) and that many disease-specific kits are readily available through commercial suppliers. More, recently array technology was applied to increase the performance of MLPA, e.g. allowing the accurate analysis of the entire *DMD* gene using 128 probes spotted in duplicate (Zeng et al. (2008)).

Array technology

The main advantage of array-based methods is their multiplexability, i.e. the number of loci that can be screened simultaneously, ultimately covering the entire human genome. Array-CGH (Comparative Genomic Hybridisation) was first developed, using arrays that contained ~ 3,000 probes (100-200 Kb genomic insert PAC/BAC clones), covering the genome at a 1 Mb resolution (Pinkel et al. (1998)). Soon these arrays contained >30,000 clones, covering the entire genome (Ishkanian et al. (2004)). In addition, arrays for genome-wide linkage and association studies determining alleles based on Single Nucleotide Polymorphism (SNP), were used early on for CNV detection (e.g. Zhou et al. (2004)). Since these arrays were not designed for this purpose, specific software had to be developed to allow quantitative analysis of the data. While the first arrays contained only 10,000 SNPs

(Zhou et al. (2004)), the latest arrays contain up to 1 million SNPs. Next to the SNP probes, these latter arrays contain yet another 1 million non-SNP probes, filling in the gaps and yielding the best genome coverage possible. These genome-wide SNP-arrays revealed an unexpectedly large and complex variability in the human genome (Redon et al. (2006)), a variability we still have not completely mapped (Kidd et al. (2008)). In addition these arrays have been instrumental in the recent discovery of many new genes and gene regions involved in genetic disease (Beckmann et al. (2007)).

New methods

Although many powerful methods are available, there is always room to improve. Array-based approaches are very powerful but overall, although prices steadily drop, they are relatively costly and not ideal when thousands of samples need to be screened. Furthermore, for diagnostic applications genome-wide screens can often not be used. Using a genome-wide tool to determine whether there is a pathogenic CNV in the *DMD* gene of a suspected carrier female for Duchenne muscular dystrophy might reveal CNVs elsewhere in the genome, giving the diagnostic lab unwanted dilemmas. Similarly, CNVs might be found for which the phenotypic consequences are unclear, yielding an inconclusive diagnosis. With these considerations in mind, we set out to develop a new assay to bridge the gap between genome-wide (array-based) and locus-specific methods. It should facilitate the cost-effective screening of 1000 or more loci, with a flexible choice of loci to include (custom design) and the possibility of automated analysis of many samples.

1400-plex CNV bead assay

The assay developed was based on Illumina's GoldenGate assay (Fan et al. (2003)) with two important changes; targeting non-SNP DNA sequences and adapted to a single colour assay. The assay was designed to screen patients with mental retardation (MR) of unknown aetiology. It should detect all trisomies, telomere-end rearrangements (incl. unbalanced translocations), known micro-deletion/-duplication syndromes and perform a rough whole genome CNV scan (see M&M probe design). Initial assessment of the array performance was made by analysing 44 samples containing known rearrangements (see M&M). All known rearrangements were detected and all CNV breakpoints matched those known.

Detailed analysis of the first 80 MR-patient samples detected one or more CNVs in 69 cases (Fig.1), in total 103 losses and 255 gains. Ten cases carried likely pathogenic autosomal CNVs. Five of these were selected and all were confirmed using a second technique (MLPA, FISH or SNP array - Fig.1B). Three cases could be proven to contain *de novo* aberrations. 59 of the 80 patients showed CNVs in regions known to be polymorphic (Toronto Database

of Genomic Variants - http://projects.tcag.ca/variation), incl the *CDKN1c, TERT, SMN1* and *NSF* genes. In addition we detected polymorphic CNVs in the *FOXD3, SOX12, TBX4, HOXD1* and *TBX21* genes, not reported in the Toronto database. The results from 240 additional patients are currently under study. So far, likely pathogenic autosomal CNVs could be confirmed in 29 cases (Table 1); 1 trisomy, 7 unbalanced translocations, 7 telomeric deletions (incl.1 ring chromosome), 10 deletions and 4 duplications (incl. a partial tetrasomy). Overall the assay thus detected a pathogenic CNV in 9% of the cases analysed.

Table 1: 1400-plex CNV bead assay. In total 320 mental retardation patients were screened using the 1400-plex CNV bead assay. Data from the first set of 80 patients has been analysed extensively, that of the second set of 240 additional patients is still in progress. Shown is an overview of all confirmed CNVs and where known the diagnosis of the phenotype in addition to mental retardation. DGS = DiGeorge syndrome, MCA = multiple congenital anomalies, dup22q11 = 22q11 microduplication syndrome, MR = mental retardation, MWS = Mowat Wilson syndrome, PLS = Potocki-Lupski syndrome, SMS = Smith Magenis syndrome, WBS = Williams Beuren syndrome.

CNV	Number of patients	Phenotype diagnosis
Trisomy (21q)	1	Down's syndrome
Unbalanced translocations (telomeric deletion + duplication)	7	MCA
Ring chromosome (11) (11q telomeric deletions)	1	MCA
Deletion, telomeric	6	2 alpha thalessemia, 4 MCA
9p tetrasomy (9p triplication)	1	MCA
Micro-deletion	10	6 DGS, MWS, SMS, WBS
Micro-duplication	2	dup22q11, PLS
1 probe duplication	1	-

Fast-MLPA

Most CNV detection methods typically take one to several days to get from DNA to result. Although time is usually not a critical aspect, there are diagnostic cases where the presence of a CNV needs to be confirmed as soon as possible. For instance newborns with multiple

congenital malformations and complex heart defects may require extensive surgery and intensive care to survive. In cases where trisomies for chromosomes 13 or 18 are present, prognosis is particularly unfavourable and one would prefer to refrain from operating to minimize the child's suffering. Therefore we developed a test that could be performed within 1 day (<8 h).

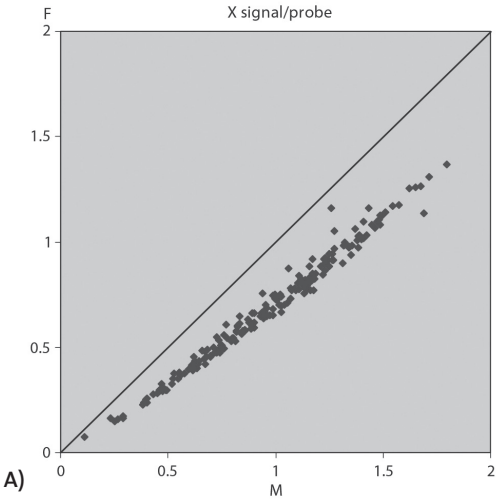
In the ultimate set up, a column-based DNA isolation method was used yielding purified DNA from a blood sample in 30 min. The time required for the MLPA could be reduced to ~ 4.5h. This was achieved by reducing the over-night hybridisation step to 2.5h and by redesigning the probe set, selecting for those probes that allowed short MLPA reaction times. Time for MLPA readout, usually acquired through capillary electrophoresis taking several hours, was reduced to 20 min. using flow-through micro-array technology (Wu et al. (2004); Zeng et al. (2008)). Overall, starting from a blood sample, the time to get to a CNV result was ~ 6 h.

The assay was designed to detect aneusomies for chromosomes 13, 18, 21 and X. First the assay was tested using samples from 4 individuals with a known aberrant karyotype, incl. trisomy-13, -18, -21. The results obtained confirmed previous findings (Fig.2). Next the assay was tested using 23 blinded DNA samples of which a number had an abnormal karyotype. All sexes and all rearrangements were correctly scored. Two samples initially gave a discrepant result, one due to a mislabelling and the other because the individual involved had received treatment for sex-reversal, with the gender reported being her new sex.

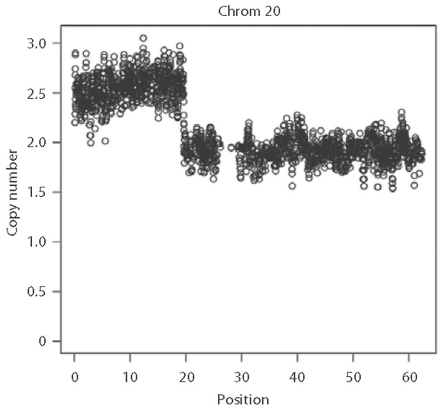
As a proof-of-principle and to check the time required we started with freshly taken blood samples of four healthy individuals. DNA was isolated, MLPA performed, array signals measured and scored. Based on the signal of the Y-chromosome probe and the altered ratios of the X-probes compared to those of the other chromosomes, we were able to correctly score the sex for all samples in about 6.5h. Thus far the assay has been applied once in a clinical setting, confirming a suspected trisomy-18 case directly after birth.

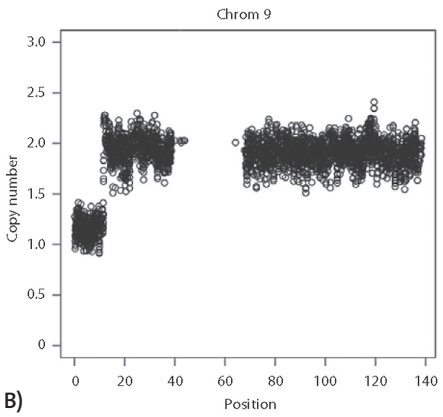
hrMCA-CNV

Now that genome-wide tools are applied in diagnostic labs worldwide, hundreds of new CNVs are being detected and there is an urgent need for a quick, reliable and cheap method to confirm the initial findings. Most CNV methodologies can not be easily used in a custom design setting or are labour and/or time intensive to develop and thus costly. While screening disease genes for sequence variants using high-resolution Melting Curve Analysis (BRCA - van der Stoep et al, DMD - Al-Momani et al, submitted) we were impressed by its resolution and sensitivity. We therefore tested its performance to confirm CNVs, simply by comparing the melting curves of different pre-defined mixes of the opposite alleles and the test sample.

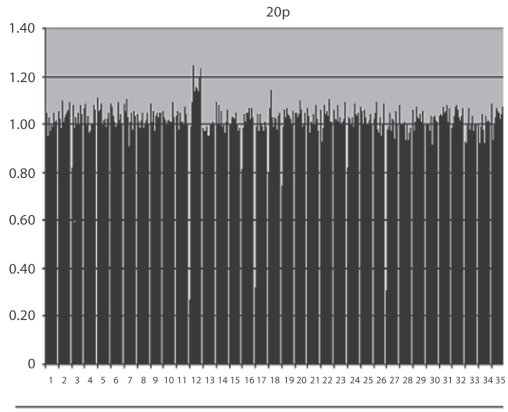


A)





B)



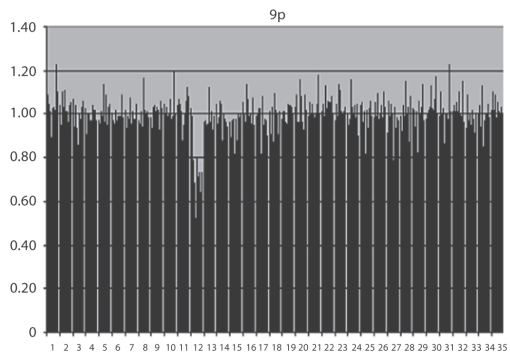


Figure 1: 1400-plex CNV bead assay. **A)** Strength of signal for probes on the X-chromosome obtained in males (M, X-axis) versus females (F, Y-axis). Signals are clearly different yet the difference in F:M signal obtained is not 1:0.5 as expected but 1:~0.7 (see also panel 1B bottom left and **1C**). **B)** Unbalanced translocation between chromosomes 20p (top, 1 extra copy) and 9p (bottom, 1 missing copy) detected in MR-patient 12 using the telomeric ruler probes (left panels). The rearrangement was confirmed using a SNP-array (right panels, Affymetrix 250K-StyI). Besides MR the patient had multiple congenital anomalies. **C)** Deletion spanning at least the COPS3 (top) and DRG2 (bottom) genes at chromosomes 17p11.2 detected in MR-patient 60 (confirmed to have Smith-Magenis syndrome). Note that in some cases only 1 of the probe pairs indicates a CNV (e.g. patient 18 for COPS3); such cases have not yet been considered.

C)

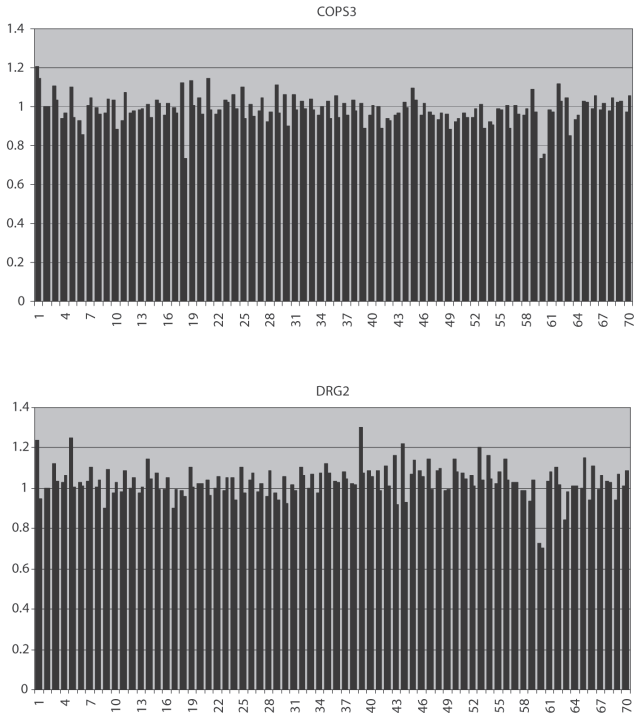


Figure 2: Fast-MLPA. Detection of a trisomy 21 (top left), Turner (45, Xo; top right) and 49, XXXXY case (bottom) using fast-MLPA.

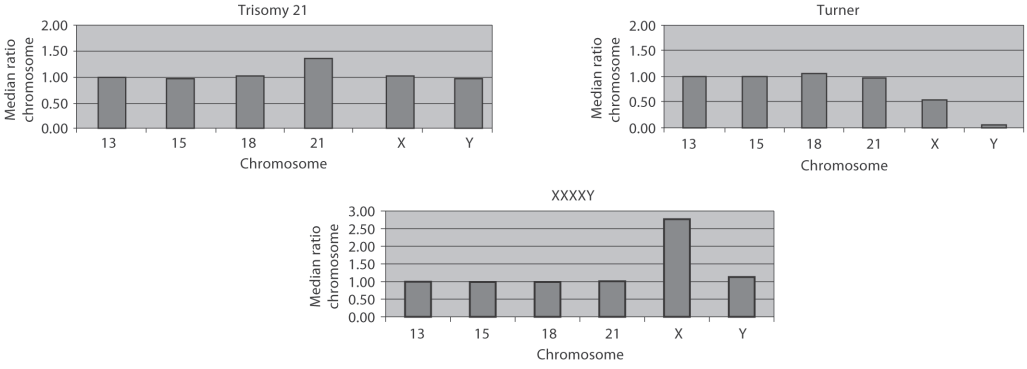


Table 2: Confirmation of CNVs using high-resolution Melting Curve Analysis (hrMCA)
Depending on the CNV to confirm the test sample, control samples (homozygote AA, AB and heterozygote AB) and specific mixes of test and control samples have to be prepared. Shared melting curves indicate the allelic composition of the test sample.

CNV	test sample	CNV confirmed when
Deletion	Ao?	Ao?/BB mix = 1:2 AA/BB control mix and differs from 1:1 AA/BB control mix and AB heterozygote
	Bo?	AA/Bo? mix = 2:1 AA/BB control mix and differs from 1:1 AA/BB control mix and AB heterozygote
Duplication	AAA?	AAA?/BB mix = 3:2 AA/BB control mix and differs from 1:1 AA/BB control mix and AB heterozygote
	AA?B	AA?B/BB mix = 2:3 AA/BB control mix and differs from 1:1 AA/BB control mix and AB heterozygote
	ABB?	AA/ABB? mix = 3:2 AA/BB control mix and differs from 1:1 AA/BB control mix and AB heterozygote
	BBB?	AA/BBB? mix = 2:3 AA/BB control mix and differs from 1:1 AA/BB control mix and AB heterozygote

The mix corresponding best with the melting curve of the test sample, resolves its allelic composition and confirms the CNV or not (Table 2).

An example of the sensitivity of the assay is shown in Fig.3. In the example alleles can be discriminated in steps of 12.5%, suggesting, besides confirming CNVs (~ 25% difference) it can be applied to type tetraploid and maybe octaploid organisms (plants). This is well within the sensitivity required to confirm CNVs where in mixed samples 3-5 alleles need to be discriminated. We have applied the hrMCA-CNV successfully to confirm both deletions and duplications detected using SNP-arrays as well as to determine the level of somatic mosaicism for CNVs in samples from monozygotic twins (Bruder et al. (2008)).

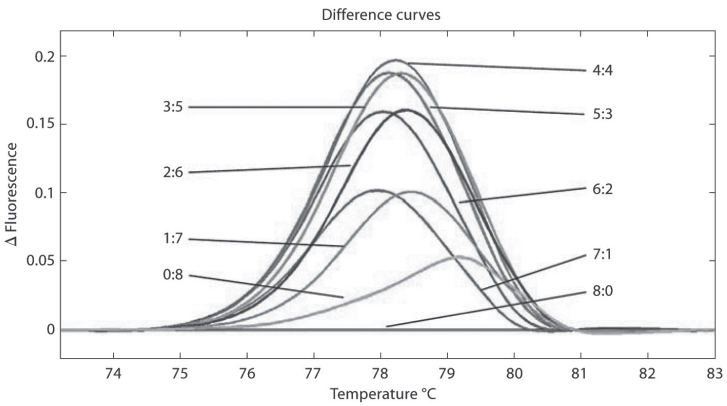
General considerations

There are many techniques and platforms available for detecting copy number changes in genomic DNA. The choice of the best method largely depends on project specific factors and the question to answer. Usually compromises have to be made with regard to the sample available, number of CNVs to analyse, resolution, cost and throughput. Not all CNV methods available can be performed with any sample. Microscopy/FISH methods require intact cells and/or nuclei. Southern blot analysis, especially when combined with Pulsed-

Field Gel-Electrophoresis, can only be performed when DNA of sufficient length is available (Van Der Maarel et al. (1999)). PCR-based methods are usually least demanding, although not all methods will work with low quality DNA (i.e. fragmented and/or contaminated (Kessler et al. (2004)). Overall methods that amplify short stretches of DNA perform better on low quality DNA compared to methods amplifying larger fragments. E.g. for whole genome array-based CNV studies, we obtained better results with SNP-arrays from Illumina then Affymetrix when low quality DNA samples were used. In rare cases the amount of DNA available might be limiting, making PCR-based methods most appropriate. However, even when only 1 or 2 cells are available, techniques like FISH are successfully applied to analyse CNVs in preimplantation diagnosis.

Multiplexability, i.e. the number of CNVs that can be studied simultaneously, often plays a decisive role. In most cases, sample throughput is inversely related with multiplexability; available methods do allow high-throughput analysis of a few loci but genome-wide CNV studies are difficult to perform for thousands of samples. In most cases the ease with which the assay can be automated determines the throughput that can be achieved. PCR-based methods are generally performed in single-locus mode and can be performed to analyse thousands of samples per week. MLPA, MAPH and MAQ facilitate the analysis of 20-50 loci using capillary electrophoresis and can be used to analyse hundreds of samples per week.

Figure 3: hrMCA CNV. High-resolution Melting Curve Analysis of two DNA samples being homozygous for the opposite SNP-alleles (rs213950:G>A) mixed in different proportions (from 8:0 to 0:8). The difference plots of all mixes can be easily discriminated, giving a sensitivity of at least 12.5%, suggesting that octaploid alleles could be typed.



To analyse more loci per sample (50-1000), either arrays (Zhou et al. (2004)) or bead-approaches (e.g. xMAP - Luminex, BeadArray - Illumina) can be applied. Such assays are ideal for diagnostic application, testing a range of targeted loci with known consequences only. The 1400-plex CNV bead assay described here is such an assay. The assay has the intrinsic possibility to score CNVs as well as specific SNPs or pathogenic mutations in one assay. Another design, which is also possible, is to use two different colours and either analyse two samples on one array (labelled in different colours, yielding improved data quality) or to double the number of loci analysed. When flexible custom-design is the most important aspect of the assay, the VeraCode technology (BeadXpress - Illumina) might have additional advantages, facilitating custom mixing of barcode-labelled probes.

Fast-MLPA tackles another, sometimes important aspect an assay, time-to-result. It allows a CNV-assay, from blood sample to result, to be performed within a working day. The PamChip technology used also facilitates real-time monitoring of hybridisation signals (Anthony et al. (2003)) and could be applied to analyse samples under very specific and highly stringent hybridisation conditions, e.g. to determine copy number of multi-copy sequences containing one or a few point mutations.

With the rapidly increasing popularity of whole genome CNV analysis using high-density SNP-arrays there is a great need for validation of the results obtained. While measuring 5-20 consecutive SNPs, Wagenstaller et al. (2007), analysing patients with mental retardation for CNV using 100K SNP-arrays, reported a false positive rate of 30%. Smaller aberrations, covering less then 5 probes, although true, are even more problematic. FISH, powerful yet laborious, is often not suitable for confirmation; cells/nuclei might not be available, probes difficult to get or spanning a region bigger then the CNV. MLPA is suitable, even for smaller aberrations, but developed to screen 20-40 loci simultaneously and taking 2-4 weeks to design (White et al. (2004)). The hrMCA-CNV method presented here is easy to perform and simple to design, requiring only SNPs from the region of interest. Assuming the CNVs were detected using a SNP-array, one often has a large and easy choice.

Future developments

Looking further ahead, the near future will undoubtedly see application of the new massive parallel sequencing technologies. Although simple and available for decades sequencing-based counting methods seem, with few exceptions (Bailey et al. (2002)), not to have been applied for CNV detection. The new massive parallel sequencing technologies seem to provide a new, very powerful tool for the detection of CNVs (e.g. using SAGE-like approaches); the resolution required determining the number of total sequences needed. In addition, applying paired-end sequencing, these new technologies will allow detection of all structural variation (Kidd et al. (2008)); CNVs as well as insertions, inversions and

translocations. In addition sequencing will be instrumental to determine the structure of the CNV as well as breakpoint sequences, the latter facilitating direct screening by breakpoint PCR.

One methodological hurdle remains to be solved; determining the exact copy number of multi-copy CNVs (>6-8 copies) in hundreds of samples and testing their possible association with specific diseases or phenotypic traits. For the latter to work, the assay should be exact and without errors, a demand where most current CNV methods already fail for copy numbers of 3-5, even when specific precautions are taken (Armour et al. (2007)). Massive deep sequencing technology might be applicable here but is still costly and not high-throughput. New innovative methods are under development, incl. automated fiber-FISH methods (BioNanomatrix) and NanoString's nCounter technology (Geiss et al. (2008)), but it is still too early to say whether they will be able to resolve this issue.

Materials and methods

Patient samples

All DNA samples were obtained from the department of Clinical Genetics (LUMC, Leiden). To check the performance of newly developed assays we used 44 control samples; 8 healthy individuals and 36 patients with known genomic rearrangements (incl. 23 DB/MD patients and carriers). In total 320 patients with mental retardation of unknown aetiology were analysed using the 1400-plex CNV bead assay, using gender information as an internal control. All subjects, or their representatives, gave informed consent for DNA studies.

Genomic DNA was isolated from blood samples using standard methods. DNA concentrations were measured using PicoGreen (Invitrogen-Molecular probes) and diluted to a concentration of 50ng/ul.

DNA isolation for fast-MLPA was performed using the Perfect gDNA Blood Mini kit (Qiagen), taking ~ 30 min; proteinase-K sample lysis (13 min), binding DNA to column (3 min), washing (4 min), elution (4min). MLPA required ~ 4.5 h; hybridization 2.5 h, ligation 20 min, PCR 1.5 h. PamChip®analysis ~ 20 min; hybridization 15 min, washing 3 min. Data analysis using ArrayPro; 10 min. Total assay time < 6 h.

1400-plex CNV bead assay

We wanted to develop a CNV assay that would facilitate the cost-effective screening of 1000 or more loci, with a flexible choice of loci to include (custom design) and the possibility of automated analysis of many samples. When we used an array-based readout (Zeng et al. (2008)) it clearly showed the advantage of using probes of equal length to obtain uniform

yields of all fragments amplified in a highly complex multiplex PCR. Illumina's Golden Gate assay uses a hybridisation-extension-ligation approach with similarities to MLPA (Fan et al. (2003)), allowing largely automated high-throughput screening of ~ 1600 loci using a bead-based micro-titer plate read-out format. Initial experiments confirmed the potential of this assay to detect CNVs, but also suggested that a non-SNP assay, i.e. using loci not covering SNPs, might have additional advantages (i.e. improved signal, simplified data analysis, doubling the number of loci that can be analysed).

For the 1400-plex CNV bead assay we designed in total 1324 non-SNP probes using standard design rules (Fan et al. (2003)). All probes for specific loci were designed in duplicate, i.e. separate for two closely spaced sequences, in unique sequences and where possible inside a gene (exonic). A telomeric ruler was created by designing probes at 0.5, 1.0, 1.5, 2.0, 3.0, and 4.0 Mb from the end of all chromosomes, except the p arm of the five acrocentric chromosomes (in total 482 probes). Regions known to be involved in micro-deletion / duplication syndromes were targeted with duplicate probes in at least one gene in every region selected. To determine the accuracy of the assay, control probes were designed for 19 exons in the *DMD* gene. 773 probes were designed for other loci of interest, incl. loci known to be copy number variable in a normal population (e.g. *CCL3L1, NSF*), containing other disease associated loci, potential regulatory regions and loci randomly spaced across the genome to provide a rough whole genome scan.

Per analysis we used 250 ng total genomic DNA following the manufacturer's recommendations. Labelled products were purified and hybridised to a Sentrix Array Matrix (SAM). After hybridisation the SAM was washed and imaged on the Illumina BeadArray Reader. The 44 control samples were used to evaluate the assay and probe performance. Both probes for one locus should give the same copy number; loci for which 85% or less of the samples gave a concordant outcome were omitted for further analysis. Probes showing an unexpected CNV in >5% of the controls were studied more carefully. These CNVs are either false positives (e.g. due to low probe quality) or true CNVs, i.e. located in hitherto unknown CNV polymorphic regions.

Intensity signals were extracted, imported to MS-Excel and analysed as described (White et al. (2004)). Ratios between 0.75 and 1.25 are regarded as a normal (i.e. two copies), below 0.75 as a CNV-loss and above 1.25 as a CNV-gain. Samples with poor DNA quality (defined as >10% of the control probes showing copy number variation within one sample) were excluded. The number of CNVs was calculated per locus, when three or more patient samples showed CNV for a locus it was regarded as polymorphic, when only one or two samples showed a CNV is was considered as a possible pathogenic variant.

Initial assessment of the array performance was made by analysing 44 samples containing known rearrangements. Overall signals were stable and the intensities of

the duplicate probes showed little variation. Comparing female and male samples for X chromosome probe signals (Fig 1) clearly separated both sexes, yet the difference in signal was not 1:0.5 as expected but only ~ 1:0.7 (Fig.1A), an as yet unexplained but known phenomenon (Pollack et al. (1999)). The possibility that the amount of hybridising material was saturating the available target was ruled out by performing the hybridisation with less material, which resulted in lower signal but no alteration in the derived ratios.

All known rearrangements were detected at exon resolution and all deletion / duplication breakpoints matched those known. The DMD deletions and duplications in 23 control samples were all detected, proving that CNVs involving a single probe could be ascertained. Although an increase in ratio is clearly seen when there are more then two copies present, it was not possible to distinguish between three and four copies of the *DMD* gene. This is not a significant problem however, as the primary purpose is to detect gains or losses per se. Where necessary to determine the precise copy number of a given locus, other methods can be applied.

Independent confirmation of CNVs was performed using several methods; custom design MLPA (White et al. (2004)), Fluorescence in situ hybridization (Dauwerse et al. (1992)) and whole genome SNP arrays (Illumina-317K and Affymetrix-5.0, performed according to the manufacturer's protocols and analysed using Beadstudio, CNAG, dChip). PAC/BAC clones for FISH were obtained from the Welcome Trust Sanger Institute (Cambridge, UK).

Fast-MLPA

Fast-MLPA probe design was basically performed as described (White et al. (2004)). Several rounds of probe optimisation were performed, selecting probe combinations that allowed short MLPA hybridisation, ligation and amplification times. The final MLPA probe set consisted of six probes per chromosome (13, 18, 21 and X) and one probe for chromosomes 15 and Y. Of the MLPA probe pair, each left-hand oligonucleotide contained a 20-nucleotide zip-sequence, facilitating selective hybridisation to a PamChip (Wu et al. (2004)), PamGene, Den Bosch, Nederland). All zip detector probes were spotted in duplicate on the array.

MLPA was performed as described (White et al. (2004)) with the exception of the hybridisation step (2.5 hours instead of overnight). PCR was performed for 33 cycles with either a Cy5- (control) or Cy3- (patient) labelled forward primer.

Array experiments were performd in the PamStation-4 or -FD10 (Pamgene). Before hybridisation, arrays were washed with 20µl 1x PBS-Tween (1 cycle of 1 min.) and 20µl 5x SSPE (1 cycle). Pre-hybridisation, 10 min. at 55°C, was carried out using 2µl tRNA (10µg/µl), 5µl 20x SSPE and13µl H_2O. Hybridization was performed using a mix containing 6µl patient DNA sample (Cy3-labelled), 6µl control DNA sample (Cy5-labelled), 1µl MAPHF and 1µl MAPHR primer (2 µM/µl each), 5µl 20x SSPE, 1µl H_2O. Hybridisation was for 10 min. at

55°C, the array was washed three times (one cycle) with fresh 20µl 5x SSPE-buffer followed by image capture. For experiments performed in the PamGene-FD10 all volumes were doubled.

Images were analysed with ArrayPro (Media Cybernetics) and the intensity data exported to MS-Excel. The median ratio was taken for normalization (White et al. (2004)). The average ratio of the duplicate spots was calculated and then the median ratio per chromosome. This gives tight ratios close to 1.0 with the advantage that lower thresholds can be set, allowing e.g. detection of mosaic cases, where an aneusomy is present in only a percentage of the cells, as we could prove by mixing DNA from a normal and a trisomy 21 case in different proportions.

hrMCA-CNV

High-resolution Melting Curve Analysis (hrMCA) was performed using the LightScanner (Idaho Technologies) and LightCycler-480 (Roche Diagnostics). PCR amplicons were designed using Primer-3 and standard design parameters, targeting fragments of 100-200 bp covering the SNP of interest. To confirm a CNV, two SNPs in the CNV region are selected. The resolution of this assay depends on the shift in melting curves between a homozygous and heterozygous sample, with larger shifts giving higher sensitivity. Therefore, we selected those SNPs from the candidate CNV region that are expected to give a large shift (i.e. A>G changes). Next to the sample containing the potential CNV, samples carrying the two opposite alleles homozygously (AA and BB) are required (Table 2).

Depending on the CNV to confirm, a deletion or duplication, a set of samples mixes are made. For deletions a 1:1 and 1:2 mix of the AA and BB and a 1:1 mix of the test "A0?" and opposite homozygous BB sample. When for a deletion the mixed A0?/BB sample shares its melting curve with that of the 1:2 AA/BB mix ("ABB") and not with the 1:1 AA/BB mix ("AB") nor with that of the heterozygote AB control, the deletion is confirmed. For the samples and mixes needed to confirm duplications see Table 2.

Acknowledgements

We like to acknowledge the invaluable help and advise of Jian-Bing Fan, Marina Bibikova, Lixin Zhou, Jing Chen and Eliza Wickham-Garcia (Illumina Inc., San Diego, USA) for developing the 1400-plex MLPA and Ying Wu and Rini van Beuningen (PamGene, Den Bosch, Nederland) while developing the fast-MLPA assays.

2

Anthony RM, Schuitema AR, Chan AB, Boender PJ, Klatser PR, Oskam L: Effect of secondary structure on single nucleotide polymorphism detection with a porous microarray matrix; implications for probe selection. BioTechniques 34:1082-1089 (2003).

Armour JA, Palla R, Zeeuwen PL, den HM, Schalkwijk J, Hollox EJ: Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic Acids Res 35:e19.1-19.8 (2007).

Armour JA, Sismani C, Patsalis PC, Cross G: Measurement of locus copy number by hybridisation with amplifiable probes. Nucl Acids Res 28:605-609 (20-1-2000).

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 297:1003-1007 (2002).

Beckmann JS, Estivill X, Antonarakis SE: Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet 8:639-646 (2007).

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

Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, de Stahl TD, Menzel U, Sandgren J, von Tell D, Poplawski A, Crowley M, Crasto C, Partridge EC, Tiwari H, Allison DB, Komorowski J, Van Ommen GJB, Boomsma DI, Pedersen NL, Den Dunnen JT, Wirdefeldt K, Dumanski JP: Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am J Hum Genet 82:763-771 (2008).

Craig JE, Barnetson RA, Prior J, Raven JL, Thein SL: Rapid detection of deletions causing delta beta thalassemia and hereditary persistence of fetal hemoglobin by enzymatic amplification. Blood 83:1673-1682 (1994).

Dauwerse JG, Wiegant JCAG, Raap AK, Breuning MH, Van Ommen GJB: Multiple colors by fluorescence in situ hybridization using ratio-labelled DNA probes create a molecular karyotype. Hum Mol Genet 1:593-598 (1992).

Den Dunnen JT, Bakker E, Klein-Breteler EG, Pearson PL, Van Ommen GJB: Direct detection of more than 50% Duchenne muscular dystrophy mutations by field-inversion gels. Nature 329:640-642 (1987).

Den Dunnen JT, Grootscholten PM, Bakker E, Blonden LAJ, Ginjaar HB, Wapenaar MC, Van Paassen HMB, Van Broeckhoven C, Pearson PL, Van Ommen GJB: 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 (1989).

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS: Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 68:69-78 (2003).

Florijn RJ, Blonden LAJ, Vrolijk H, Wiegant J, Vaandrager JW, Baas F, Den Dunnen JT, Tanke HJ, Van Ommen GJB, Raap AK: High-resolution FISH for genomic DNA mapping and colour bar-coding of large genes. Hum Mol Genet 4:831-836 (1995).

Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, James JJ, Maysuria M, Mitton JD, Oliveri P, Osborn JL, Peng T, Ratcliffe AL, Webster PJ, Davidson EH, Hood L: Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26:317-325 (2008).

Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, Ling V, MacAulay C, Lam WL: A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet 36:299-303 (2004).

Kessler N, Ferraris O, Palmer K, Marsh W, Steel A: Use of the DNA flow-thru chip, a three-dimensional biochip, for typing and subtyping of influenza viruses. J Clin Microbiol 42:2173-2185 (2004).

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tuzun E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M, Saranga D, Brand A, Tao W, Gustafson E, McKernan K, Chen L, Malig M, Smith JD, Korn JM, McCarroll SA, Altshuler DA, Peiffer DA, Dorschner M, Stamatoyannopoulos J, Schwartz D, Nickerson DA, Mullikin JC, Wilson RK, Bruhn L, Olson MV, Kaul R, Smith DR, Eichler EE: Mapping and sequencing of structural variation from eight human genomes. Nature 453:56-64 (2008).

Knight SJ, Lese CM, Precht KS, Kuc J, Ning Y, Lucas S, Regan R, Brenan M, Nicod A, Lawrie NM, Cardy DL, Nguyen H, Hudson TJ, Riethman HC, Ledbetter DH, Flint J: An optimized set of human telomere clones for studying telomere integrity and architecture. Am J Hum Genet 67:320-332 (2000).

Landegent JE, Jansen in de Wal N, Van Ommen GJB, Baas F, De Vijlder JJM, Van Duijn P, Van der ploeg M: Chromosomal localization of a unique gene by non- autoradiographic in situ hybridization. Nature 317:175-177 (1985).

Leujeune J, Gautier M, Trupin R: Study of somatic chromosomes from 9 mongoloid children. C R Hebd Seances Acad Sci 248:1721-1722 (1959).

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC: Diet and the evolution of human amylase gene copy number variation. Nat Genet 39:1256-1260 (2007).

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: BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. Nat Genet 17:341-345 (1997).

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 20:207-211 (1998).

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO: Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 23:41-46 (1999).

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME: Global variation in copy number in the human genome. Nature 444:444-454 (2006).

Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A, Dumanchin C, Feuillette S, Brice A, Vercelletto M, Dubas F, Frebourg T, Campion D: APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 38:24-26 (2006).

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 30:e57 (2002).

Suls A, Claeys KG, Goossens D, Harding B, Van Luijk R, Scheers S, Deprez L, Audenaert D, Van Dyck T, Beeckmans S, Smouts I, Ceulemans B, Lagae L, Buyse G, Barisic N, Misson JP, Wauters J, Del Favero J, De Jonghe P, Claes LR: Microdeletions involving the SCN1A gene may be common in SCN1A-mutation-negative SMEI patients. Hum Mutat 27:914-920 (2006).

Van Der Maarel SM, Deidda G, Lemmers RJ, Bakker E, van der Wielen MJ, Sandkuijl L, Hewitt JE, Padberg GW, Frants RR: A new dosage test for subtelomeric 4;10 translocations improves conventional diagnosis of facioscapulohumeral muscular dystrophy (FSHD). J Med Genet 36:823-828 (1999).

Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D, Heye B, Glaser D, Liebscher V, Meitinger T, Strom TM: Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. Am J Hum Genet 81:768-779 (2007).

White SJ, Den Dunnen JT: Copy number variation in the genome; the human DMD gene as an example. Cytogenetic and Genome Research 115:240-246 (2006).

White SJ, Vink GR, Kriek M, Wuyts W, Schouten JP, Bakker E, Breuning MH, Den Dunnen JT: Two-colour MLPA; detecting genomic rearrangements in hereditary multiple exostoses. Hum Mutat 24:86-92 (2004).

Wu Y, De Kievit P, Vahlkamp L, Pijnenburg D, Smit M, Dankers M, Melchers D, Stax M, Boender PJ, Ingham C, Bastiaensen N, de Wijn R, van Alewijk D, Van Damme H, Raap AK, Chan AB, Van Beuningen R: Quantitative assessment of a novel flow-through porous microarray for the rapid analysis of gene expression profiles. Nucleic Acids Res 32:e123.1-e123.7 (2004).

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 33:550-558 (1996).

Zeng F, Ren ZR, Huang SZ, Kalf M, Mommersteeg M, Smit M, White S, Jin CL, Xu M, Zhou DW, Yan JB, Chen MJ, van BR, Huang SZ, Den Dunnen JT, Zeng YT, Wu Y: Array-MLPA: comprehensive detection of deletions and duplications and its application to DMD patients. Hum Mutat 29:190-197 (2008).

Zhou X, Mok SC, Chen Z, Li Y, Wong DT: Concurrent analysis of loss of heterozygosity (LOH) and copy number abnormality (CNA) for oral premalignancy progression using the Affymetrix 10K SNP mapping array. Hum Genet 115:327-330 (2004).