Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis

Lisenka E L M Vissers,1 Bert B A de Vries,1,2 Joris A Veltman1

ABSTRACT
Structural chromosomal rearrangements can lead to a wide variety of serious clinical manifestations, including mental retardation (MR) and congenital malformations. Over the last few years, rearrangements below the detection level of conventional karyotyping have been proved to contribute significantly to the cause of MR. These so-called copy number variations are now routinely being detected using various high-resolution microarray platforms targeting the entire human genome. In addition to their clinical diagnostic use, the introduction of these high resolution platforms has facilitated identification of novel microdeletion and microduplication syndromes as well as disease genes. The aims of this review are to address several aspects of this revolutionising technology including its application in the diagnostics of MR, the identification of novel microdeletion and microduplication syndromes, and the finding of causative genes for known syndromes. In addition, a future prospect is provided for the detection of disease causing mutations and structural variants by next generation sequencing technologies.

INCREASING THE RESOLUTION TO STUDY THE HUMAN GENOME
Structural chromosomal rearrangements can lead to a wide variety of serious clinical manifestations, including mental retardation (MR) and congenital malformations. Chromosomal rearrangements larger than 5–10 Mb in size can be detected by conventional karyotyping. A considerable number of clinical disorders, however, is caused by submicroscopic chromosomal rearrangements smaller than 5–10 Mb in size. Depending on the clinical diagnosis, specific (Q)PCR or fluorescent in situ hybridisation (FISH) probe(s) can be used to analyse a specific chromosomal region and confirm a clinical diagnosis. However, an efficient and robust technology was needed to routinely detect rearrangements beyond the level of karyotyping in an unbiased and genome wide fashion.

Genome profiling technologies, such as array based comparative genomic hybridisation (CGH), have dramatically changed the nature of human genome analysis by combining the targeted high resolution approach of the FISH technology and the whole genome approach of the karyotyping technology. Initially, genomic microarrays were developed in academia and contained mostly genomic fragments obtained from large insert genomic clones, mainly bacterial artificial chromosomes (BACs).1,2 Different clone sets have been used, the most popular ones containing one clone per 1 Mb or later on using a tiling resolution clone set of approximately 30 000 clones, covering the genome with one clone per 100 kb.3–10 In the last few years genomic microarray production has been taken over by private enterprises and many companies are now offering microarrays for genome wide copy number profiling. With the increasing resolution of the different array platforms, detection of smaller and smaller genomic copy number variations (CNVs) has become possible (see table 1 for an overview of most popular genomic microarray platforms).

In 2007, a novel method was developed to estimate the ability of a microarray to reliably detect genomic CNVs of different sizes and types all over the genome.11 The method is based on the following variables: (1) the genomic coverage of the platform; (2) an estimate of the noise in the microarray experiment (the standard deviation of the test-over-reference ratio of the autosomal targets); (3) an estimate of a single copy number loss (ratio of the chromosome X unique regions of sex mismatch experiments); and (4) the desired statistical power. Four widely used high density genomic microarray platforms for CNV detection were tested for their performance, including 32k BAC arrays, 100 and 250k single nucleotide polymorphism (SNP) microarrays and 385k oligonucleotide arrays. By doing so, it was found that the high density oligonucleotide platforms are superior to the BAC platform for the genome wide detection of CNVs smaller than 1 Mb. The capacity to reliably detect single CNVs below 100 kb, however, at that time appeared to be limited for all platforms tested. These analyses provided a first objective insight into the true capacities and limitations of different genomic microarrays to detect and define CNVs. Moreover, the study showed that, depending on the microarray platform being used and the pre-processing steps being performed before CNV detection, 3–18 adjacently located targets were required for reliable detection of single copy number losses or gains. In addition, the analysis revealed an unexpected platform dependent difference in sensitivity to detect a single copy number loss and a single copy number gain. Single copy number gains are more difficult to detect due to the fact that the intensity ratios of single copy number gains (in theory a test-over-reference ratio of 1.5, or 0.58 in log2 scale) are sometimes close to the experimental noise level of a microarray platform. In conclusion, this study showed that genomic microarray platforms vary in their capacity to...
reliably detect CNVs of different sizes and different types. This should be taken into account for estimating the practical resolution of a platform to detect genomic CNVs. At the time of this analysis (2007) many of the platforms still contained considerable gaps in coverage, mostly due to the fact that these regions contained no unique sequences and were often excluded from the SNP microarray design because of Mendelian inconsistencies.

Most of the above mentioned problems have now been solved and companies have released microarrays containing more than two million oligonucleotides targeting random sequences, SNPs, or combinations thereof (table 1). These oligonucleotides have been more evenly spaced across the genome, and optimised protocols are now available for the quantitative detection of CNVs. With this, CNV detection can now reliably be performed at the kilobase level, resulting in the detection of hundreds of CNVs per individual. 

These advances have made genomic profiling technology an excellent tool for clinical genetic diagnostic applications as well as for fundamental genome research.

APPLICATION OF GENOMIC PROFILING IN THE DIAGNOSTICS OF MR

Genomic microarrays have been extensively used in studying the genetic causes of MR and this disorder can therefore be considered a model disease to study the clinical consequences of CNVs. MR occurs in 2–3% of newborns in the general population, but, in most cases, its cause has remained elusive. 

Establishing the cause in a mentally retarded individual improves clinical management and facilitates genetic counselling of the family. Chromosome abnormalities are detectable by microscopic analysis of chromosomes isolated from peripheral blood lymphocytes in ~5% of patients with unexplained MR. Molecular cytogenetic techniques, such as FISH and multiplex ligation dependent probe amplification (MLPA), have shown that causative submicroscopic rearrangements of the subtelomeric regions can be found in ~5% of patients with human malformations and MR. These results for the subtelomeric regions indicated early on that submicroscopic rearrangements such as CNVs may be a more common cause of MR than microscopically visible rearrangements.

From its introduction using genome wide 1 Mb BAC arrays, array CGH has proven useful in the diagnostics of MR with the detection of causative microdeletions and/or duplications in ~10% of individuals with MR with or without additional congenital anomalies. Additional studies have provided insight into the quality and reproducibility of the procedure, the need for validation of the microarray data by independent technologies such as FISH or MLPA, as well as the way to translate these data into clinical practise.

The clinical usefulness of molecular karyotyping was further substantiated in larger, less selected, cohorts of individuals with MR using 1 Mb resolution BAC arrays, 4, 6 tiling resolution BAC arrays, 10 or 100k SNP arrays. In the latter tiling resolution BAC array study, reproducible DNA copy number changes were detected in 97% of patients. The majority of these alterations appeared to be inherited from phenotypically normal parents, which reflected normal CNVs in the human population. In 10% of patients rare de novo alterations considered to be clinically relevant were found: seven deletions and three duplications, varying in size from 540 kb to 12 Mb and occurring scattered throughout the genome.

Many similar studies have been published since (reviewed by Koolen et al, Veltman, and Knight and Regan). When taking all studies together, two main conclusions can be drawn: (1) in addition to submicroscopic subtelomeric chromosome imbalances, rare, de novo, submicroscopic interstitial chromosome imbalances or CNVs are responsible for a considerable proportion of cases with MR varying between ~5–20%, depending on the clinical pre-selection of the individuals; and (2) these rare de novo CNVs occur all over the genome. When comparing these results to standard GTG banded karyotyping, the diagnostic yield of array CGH in the general population of patients with MR is at least twice as high.

Next to the apparent causative alterations, a large number of inherited submicroscopic CNVs without evident clinical consequences have been detected by array based methods, in patients as well as in control populations. 

CNV is now considered as a common form of structural genomic variation with ultrahigh resolution microarrays and sequencing approaches identifying >1000 CNVs in a single individual. Current clinical interpretation therefore involves an analysis of the frequency of CNVs in unaffected control cohorts as well as parental analysis (see Lee et al as well as Koolen et al for a review of literature on the application of genomic microarray to MR and a practical workflow for diagnostic applications). The identification of a (1) relatively large, (2) rare, and (3) de novo CNV in such a patient is a strong indicator of clinical significance, as this combination is rare in the normal population. These indicators are specifically helpful if the potential causative CNV can readily be detected. An additional level of complexity in interpreting array data is the presence of mosaicism, which can make the initial identification of such potential causative CNV may be hampered. In a comprehensive study of 638 neonates with various birth defects, array analyses detected 12 (1.9%) mosaic pathogenic variants. The notion that MR CNVs may occur postzygotically underscores the importance for automated detection of mosaicism.

IDENTIFICATION OF RECURRENT MICRODELETION AND MICRODUPlication SYndromes

The possibility to perform genome wide CNV studies in patients with MR has substantially increased the chance to identify novel ‘microdeletion/microduplication syndromes’. In general, the identification of novel syndromes is based on an accurate phenotype–genotype correlation. From a historical perspective, this correlation relied on a detailed and accurate phenotypic...
description of the patients after which overlapping chromosomal rearrangements were uncovered. Nowadays, obtaining a genotype has become much easier and, in addition, has never been more accurate. As a result, the identification of novel syndromes may start with the identification of overlapping genotypes—that is, a ‘genotype first’ approach, 46 or ‘reverse phenotypics’, in which patients are characterised by a similar genomic aberration before a common clinical presentation is defined. This approach has proven to be successful considering the growing list of microdeletion/microduplication syndromes (table 2).

The 17q21.31 microdeletion syndrome was the first microdeletion syndrome identified through this approach, simultaneously described by three groups. 48–50 Recurrent overlapping de novo microdeletions in 17q21.31 were identified in patients with MR using array CGH and MLPA. Clinical comparison of these patients revealed pronounced phenotypic similarities—that is MR, hypotonia and characteristic facial features, including a long hypotonic face with upslanting palpebral fissures, epicantic folds, ptosis, large prominent ears, a tubular or pear shaped nose with a bulbous nasal tip, long columella with hypoplastic alae nasi and a broad chin. 48 Other clinically important features include epilepsy, heart defects and kidney/urologic anomalies. The identification of more patients with the same aberration showed that the 17q21.31 microdeletion syndrome is a frequent cause of MR with an estimated prevalence of ~ 1 in 16 000, and allowed the detailed clinical and molecular delineation of this syndrome. 51, 52 Currently, it is still unknown whether the 17q21.31 microdeletion syndrome can be caused by a mutation in a single gene located within the deletion interval (MAPT, CRHR1, IMP5 and STH). Efforts in sequencing the coding regions of the MAPT gene in 122 patients resembling the 17q21.31 deletion phenotype but without the deletion have so far not revealed any mutations. 50–51

Another example of a clinically well recognisable microdeletion syndrome is the 15q24 microdeletion. Initially only four individuals with submicroscopic overlapping deletions of the 15q24 region were ascertained by screening a total of ~ 1200 individuals with idiopathic MR. 52 They shared several clinical features, including MR, growth retardation, microcephaly, digital abnormalities, genital abnormalities, hypospadias and loose connective tissue. In addition, similar facial dysmorphisms were noted, including high frontal hairline, broad medial eyebrows, downslanting palpebral fissures and a long philtrum, indicating that the 15q24 deletions represent a clinical syndrome. 52 A further 15q24 microdeletion case showed similar

<table>
<thead>
<tr>
<th>Name</th>
<th>Size (Mb)*</th>
<th>LCR</th>
<th>MIM</th>
<th>Clinical features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q21.1 microdeletion</td>
<td>1.1</td>
<td>+</td>
<td>612474</td>
<td>Mild-to-moderate MR, MC, cardiac abnormalities, cataracts, clear incomplete penetration</td>
</tr>
<tr>
<td>1q21.1 microduplication</td>
<td>1.1</td>
<td>+</td>
<td>612475</td>
<td>Autism or autistic behaviours, mild to moderate MR, microcephaly, mild FD</td>
</tr>
<tr>
<td>1q41q42 microdeletion</td>
<td>1.2</td>
<td>−</td>
<td>−</td>
<td>MR, seizures, various dysmorphisms, cleft palate, diaphragmatic hernia</td>
</tr>
<tr>
<td>2p15q16.1 microdeletion</td>
<td>3.9</td>
<td>−</td>
<td>−</td>
<td>MR, MC, receding forehead, ptosis, telecanthus, short palpebral fissures, downslanting palpebral fissures, broad/high nasal bridge, long/straight eyelashes, smooth and long philtrum, upper vermilion border, everted lower lip, high narrow palate, hydrencephalus, optic nerve hypoplasia</td>
</tr>
<tr>
<td>3q29 microdeletion</td>
<td>1.6</td>
<td>+</td>
<td>609425</td>
<td>MR, mild FD, including high nasal bridge and short philtrum</td>
</tr>
<tr>
<td>3q29 microduplication</td>
<td>1.6</td>
<td>+</td>
<td>611936</td>
<td>Mild/moderate MR, MC, obesity</td>
</tr>
<tr>
<td>7q11.23 microdeletion</td>
<td>1.5</td>
<td>−</td>
<td>609575</td>
<td>MR, speech and language delay, autism spectrum disorders, mild FD</td>
</tr>
<tr>
<td>9q22.3 microdeletion</td>
<td>6.5</td>
<td>−</td>
<td>−</td>
<td>MR, hyperactivity, overgrowth, trigonocephaly, macrocephaly, FD</td>
</tr>
<tr>
<td>12q14 microdeletion</td>
<td>3.4</td>
<td>−</td>
<td>−</td>
<td>Mild MR, failure to thrive, proportionate short stature and osteopeliosis</td>
</tr>
<tr>
<td>14q11.2 microdeletion</td>
<td>0.4</td>
<td>−</td>
<td>−</td>
<td>MR, widely spaced eyes, short nose with flat nasal bridge, long philtrum, Cupid’s bow of the upper lip, full lower lip, auricular anomalies</td>
</tr>
<tr>
<td>15q13.3 microdeletion</td>
<td>1.5</td>
<td>+</td>
<td>612001</td>
<td>MR, epilepsy, hypotonia, short stature, microcephaly and cardiac defects</td>
</tr>
<tr>
<td>15q24 microdeletion</td>
<td>1.7</td>
<td>+</td>
<td>−</td>
<td>MR, growth retardation, MC, digital abnormalities, genital abnormalities, hypospadias, loose connective tissue, high front hairline, broad medial eyebrows, downslanting palpebral fissures, long philtrum</td>
</tr>
<tr>
<td>16p11.2 microdeletion/duplication</td>
<td>0.6</td>
<td>+</td>
<td>611913</td>
<td>Association with MR, autism, schizophrenia</td>
</tr>
<tr>
<td>16p11.2p12.2 microdeletion</td>
<td>7.1</td>
<td>+</td>
<td>−</td>
<td>MR, flat facies, downslanting palpebral fissures, low-set and malformed ears, eye anomalies, orofacial clefting, heart defects, frequent ear infections, short stature, minor hand and foot anomalies, feeding difficulties, hypotonia</td>
</tr>
<tr>
<td>16p13.1 microduplication</td>
<td>1.6</td>
<td>+</td>
<td>−</td>
<td>MR, MC, epilepsy, short stature, phenotypic variability</td>
</tr>
<tr>
<td>17p11.2 microduplication †</td>
<td>3.7</td>
<td>+</td>
<td>610883</td>
<td>MR, infantile hypotonia, failure to thrive, autistic features, sleep apnoea, and structural cardiovascular anomalies</td>
</tr>
<tr>
<td>17q21.31 microdeletion</td>
<td>0.5</td>
<td>+</td>
<td>610443</td>
<td>MR, hypotonia, long hypotonic face with ptosis, large and low set ears, tubular or pear shaped nose with bulbous nasal tip, long columella with hypoplastic alae nasi, broad chin</td>
</tr>
<tr>
<td>19q13.11 microdeletion</td>
<td>0.7</td>
<td>−</td>
<td>−</td>
<td>MR, pre- and postnatal growth retardation, primary microcephaly, hypospadias, ectodermal dysplasia including scalp aplasia, dysplastic nails and dry skin</td>
</tr>
<tr>
<td>22q11.2 microduplication</td>
<td>3.7</td>
<td>+</td>
<td>608363</td>
<td>Highly variable. MR, FD, for example widely spaced eyes and downslanting palpebral fissures, velopharyngeal insufficiency, conotruncal heart disease</td>
</tr>
<tr>
<td>22q11.2 distal microdeletion</td>
<td>1.4–2.1</td>
<td>+</td>
<td>611867</td>
<td>MR, prematurity, prenatal/postnatal growth delay, mild skeletal abnormalities, arched eyebrows, deep set eyes, smooth philtrum, thin upper lip, hypoplastic alae nasi, small pointed chin</td>
</tr>
<tr>
<td>Xq28 microduplication</td>
<td>0.4–0.8</td>
<td>−</td>
<td>−</td>
<td>MR, severe hypotonia, progressive lower limb spasticity, absent or very limited speech</td>
</tr>
</tbody>
</table>

FD, facial dysmorphism; LCR, low copy repeat; MC, microcephaly; MR, mental retardation.

*Common region.
†Potocki–Lipsi syndrome.
phenotypic features, although microcephaly and growth deficiency were absent. The deletions in the patients varied from 1.7–3.9 Mb in size. An additional four new deletion patients were identified after screening a cohort of 9000 diagnostic cases. Interestingly, this latter study presented two patients with a duplication of the same region. The phenotype of these patients partially overlapped with the microdeletion phenotype. The breakpoints were located in nearly identical segmental duplications, which turned non-allelic homologous recombination into the most likely underlying molecular mechanism of occurrence.54

Both the 17q21 and the 15q24 microdeletion syndromes are examples where the initial identification of the (overlapping) microdeletions led to a consistent and well recognisable clinical entity. However, an increasing number of genomic loci has been recently reported with variable inheritance and penetrance, challenging clinical interpretation. Examples of this have been reported for CNVs at 1q21.1,55 56 15q13.3,57 58 and 16p13.1.11,59 60 In more detail, the recurrent 1.35 Mb deletion within 1q21.1 was initially identified in 52 persons screened over 21 000 patients with unexplained MR, autism and/or congenital anomalies.55 56 The phenotype varied considerably and included mild-to-moderate MR, microcephaly, cardiac abnormalities and cataracts. Remarkably, several unaffected deletion carriers were noted, underscoring the clinical variability. Enrichment of 1q21.1 deletions in persons with schizophrenia was also reported, suggesting a role in psychiatric disorders as well.61 62 The reciprocal microduplication involving 1q21.1 was associated with autism or autistic behaviours, and common phenotypic features of the duplication carriers included mild to moderate MR, macrocephaly or relative microcephaly, and mild dysmorphic features.55 56 Also this microduplication was identified in apparently normal individuals.

The clinical variability for the 15q13.3 microdeletion is of a similar order to the 1q21.1 microdeletion. Initially, nine affected individuals were identified in a large cohort of individuals with MR of unknown aetiology. These patients included six probands: two with de novo 15q13.3 deletions, two who inherited the deletion from an affected parent, and two with an unknown mode of inheritance.57 The patients had MR, epilepsy and variable facial and digital dysmorphisms in common. The recurrent 1.5 Mb deletion encompasses six genes, including a candidate gene for epilepsy (CHRNA7). The clinical variability of the 15q13 microdeletion was underscored by other studies showing a clinical spectrum varying from non-pathogenic to a severe outcome with a highly variable intra- and inter-familial phenotype.58 In addition to cognitive impairment the phenotype might also include features of autism spectrum disorders and a variety of neuropsychiatric disorders.65 In order to further clinical interpretation, the continuous collection of (disease causing) CNVs and their associated phenotypes in databases such as ECARUCA (http://www.ecaruca.net) and DECIPHER (https://decipher.sanger.ac.uk/application/) is of major importance, not only for the confirmation of pathogenicity, but also for the proper counselling of patients and families.

RESOLVING THE GENETIC CAUSE OF KNOWN SYNDROMES

In addition to screening individuals with MR and defining new microdeletion and microduplication syndromes, high resolution genome profiling technologies may also facilitate the identification of disease genes underlying known syndromes for which the genetic cause has remained elusive. The specific phenotype observed in patients with such syndromes allows for a stringent pre-selection of patients whose DNA can subsequently be interrogated using such high resolution genome profiling techniques. The first syndrome for which the genetic basis was resolved by this approach was CHARGE syndrome.64

CHARGE syndrome (MIM 214800) is an autosomal dominant disorder with a prevalence of one in 10 000.65 The acronym CHARGE was first proposed in 1981 based on the cardinal features identified when the association was clinically delineated: Coloboma, Heart malformation, choanal Atresia, Retardation of growth and/or development, Genital anomalies, and Ear anomalies.66 Most cases of CHARGE syndrome are sporadic, but several aspects of this condition, including the existence of rare familial cases and a high concordance rate in monozygotic twins, supported the involvement of a genetic factor. Rare de novo cytogenetic abnormalities have been described, but no specific locus had been identified before 2004. Also, systematic genome scans by conventional metaphase CGH and microsatellite analyses did not reveal a common genetic cause, nor did targeted sequencing of candidate genes such as PAX2 and PITX2. With the availability of microarray based approaches, new unbiased, genome wide screens were performed hypothesising that microdeletions and/or microduplications might be the underlying cause of CHARGE syndrome.64 Initial screening of two patients with CHARGE syndrome on a 1 Mb BAC array revealed a microdeletion of ~5 Mb in one of the patients at chromosome locus 8q12. Interestingly, an individual with CHARGE syndrome with an apparently balanced chromosome 8 translocation had been reported previously, with the breakpoint estimated within the 8q12 region.67 Array CGH analysis of this translocation unravelled two interspersed microdeletions, overlapping with the microdeletion of the first patient. Subsequent array analyses on a tiling resolution chromosome 8 BAC array of 17 additional CHARGE patients did not show any additional microdeletions. As such, it was reasoned that mutations could be present in these patients in one of the genes residing in the shortest region of deletion overlap. Sequence analysis of nine genes located within this region revealed causative mutations in CHD7, a novel member of the chromodomain helicase DNA binding gene family, in the majority of individuals with CHARGE syndrome without deletions. Based on these results, it was concluded that CHARGE syndrome is caused by haploinsufficiency of the CHD7 gene, either by microdeletions encompassing the CHD7 gene, or by mutations within this gene.64

A second well illustrated example of gene discovery through deletion and/or translocation mapping is the discovery of the euchromatic histone methyl transferase 1 (EHMT1) gene causing 9q subtelomeric deletion syndrome (MIM 610253). Submicroscopic subtelomeric deletions of chromosome 9q (9qSTSs) are associated with a recognisable MR syndrome, with clinical features including severe MR, hypotonia, brachy(micro)cephaly, epileptic seizures, flat face with hypertelorism, synophrys, everted lower lip, carp mouth with macroGLOSSIA, and heart defects.66–71 The identification of the molecular cause of 9qSTS started with the initial FISH screening of subtelomeric rearrangements in 12 patients narrowing down the commonly deletion region to ~1.2 Mb interval.70 Subsequently, this region was further reduced to an ~700 kb, still containing at least five genes and several expressed sequence tags (ESTs).71 The first evidence that 9qSTS was not a contiguous gene syndrome, but a single gene disorder, came from the characterisation of the breakpoints of a balanced translocation t(X;9)(p11.23;q34.3) in a patient presenting with typical features of 9qSTS. Molecular analyses revealed that the chromosome 9 breakpoint disrupted the EHMT1 gene in intron 9.72 Additional evidence for the causative
role of EHMT1 was provided by deletion screening and sequence analysis of the gene in 23 patients with a clinical presentation reminiscent of 9qSTDS.73 Of these 23 patients, three showed a deletion including the EHMT1 gene. However, more importantly, mutation analysis revealed de novo mutations in the EHMT1 gene in two additional patients. With this discovery, it was established that haploinsufficiency, of the EHMT1 gene, either by deletion or mutation, leads to 9qSTDS.69

Placed in a broader perspective, array based genomic profiling may be best suited for resolving the genetic cause of known syndromes that involve haploinsufficiency as the disease causing mechanism (table 3).74 Whether the latter is the case may be difficult to predict from the phenotype alone. Also, it is difficult to predict how many patients need to be included in the study to find a microdeletion. For instance, although both can be considered single gene disorders, for CHARGE syndrome CHD7 gene mutations are prevailing over gene deletions whereas deletions involving EHMT1 are more prevalent than mutations in 9qSTDS. This also shows that it is challenging to predict whether a syndrome is a single gene disorder, a contiguous gene syndrome, a genomic disorder, or a combination thereof.

**FUTURE PERSPECTIVE ON DISEASE GENE AND CNV IDENTIFICATION IN CLINICAL GENETICS**

The ultimate resolution to screen the human genome for disease causing mutations and CNVs is at the base pair level. Major advances in DNA sequencing technologies, collectively termed next generation sequencing (NGS) technologies, are now enabling the comprehensive analysis of whole genomes, transcriptionomes and interactomes.80 75–81 Currently, NGS comprises three main non-Sanger based sequencing methods: (1) pyrosequencing (Roche 454 technology); (2) sequencing with reversible terminators (Solexa technology); and (3) sequencing by ligation (SOLID technology).75 The main differences between the methods are read length, number of reads per run, and the costs involved.76 Although the method of choice is based upon the research/diagnostic question, all NGS methods are in principle capable of detecting both single base mutations and structural variation (figure 1).

With shotgun sequencing, the genome is shredded into smaller fragments of DNA which can be massively sequenced in parallel. Next, the sequenced fragments are assembled into contigs based on the overlap in the sequence reads (de novo assembly) or, alternatively, are aligned to a reference genome. In the latter situation, single base pair changes compared to a reference genome can be identified and as such may lead to disease gene identification. Proof-of-principle studies, using the autosomal dominant Freeman–Sheldon syndrome and X-linked MR, have already shown that causative point mutations can be identified using this approach.80 81 CNV can be identified by differences in read depth—for example, the number of reads mapping to a specific genomic locus also referred to as coverage (figure 1C). For instance, for heterozygous deletions half the number of reads should be expected compared to the surrounding regions where two copies are present, whereas for duplications 1.5 times the number of sequence reads should be present. Additional evidence for copy number variants is provided by so-called ‘split-reads’ in which one part of the sequence read maps to one side of the deleted or duplicated interval, whereas the remainder of the sequence read maps to the other side of the interval.

Currently, the most specific NGS application to identify CNV is paired-end mapping or mate-pair library sequencing, as this application directly provides detailed positional information.78 79 This application does not only identify unbalanced variants but also balanced rearrangements, such as translocations and inversions. For mate-pair runs, genomic DNA is randomly sheared and size selected. After several processing steps, shotgun reads are obtained by sequencing both ends of the size selected DNA library. This positional information determined by the size selection constrains the placement of paired reads within the reference genome. Deviations from this expected size distribution may point to structural variation (figure 1D). For example, fragments sequenced from 3 kb library are expected to map ~5 kb apart when mapped back onto the reference genome, whereas fragments mapping ~100 kb apart may point to a deletion in the DNA library tested. Additionally, mate-pairs with different strand location, orientation or mapping positions to different chromosomes may indicate inversions and translocations. Interestingly, paired-end mapping strategies have identified numerous structural variants currently not annotated in the reference genome, suggesting that the reference genome is still incomplete.79

**CONCLUSIONS**

In February 2001, the International Human Genome Sequencing Consortium and Celera Genomics reported the first draft sequence of the human genome.74 In the years that followed, this draft sequence has been instrumental for the systematic analysis of the human genome, including the identification/annotation of novel genes, the elucidation of regional differences in genome composition, and the identification of SNPs.
addition, new high-throughput approaches such as array CGH were developed that facilitated and notably accelerated the analysis of the human genome on a large scale, including the detection of an unprecedented level of CNV within it. Together, these approaches have contributed significantly to the rapid development of molecular karyotyping, which allows disease phenotypes to be directly linked to gene dosage alterations.

The concept of molecular karyotyping has significantly changed the field of clinical cytogenetics and clinical diagnostics in this decennium. The ability to obtain detailed quantitative copy number information has already led to a significant improvement in diagnostic yield in patients with MR and is likely to do so for other common diseases such as autism, epilepsy, and schizophrenia.\(^6\)\(^\text{--}\)\(^8\)\(^7\) The genetic basis of several clinical syndromes has been uncovered by this approach and novel microdeletion and microduplication syndromes have been identified from clinically heterogeneous cohorts. Without any doubt, the implementation of next generation sequencing technologies and medical resequencing strategies will continue to change clinical genetic research and diagnostics.\(^7\)\(^6\)\(^\text{--}\)\(^8\)\(^9\)

Eventually, up to 25% of all cases of MR may be explained by copy number dependent gene dosage variations, although not all of these variants will be fully penetrant, challenging clinical interpretation. In addition, high throughput (re-)sequencing may reveal disease associated variants in another 10–30% of cases. Clinical and biological interpretation of these variants will require large international and multidisciplinary collaborative efforts.

Acknowledgements This work was supported by grants from the Nethelands Organisation for Health Research and Development [ZonMW 916.86.016 to LEVM, ZonMW 917.86.319 to BBAv, ZonMW 917.68.363 to JAV], grants from the ArEduploidy project (LSHS-CT-2006-037627 to BBAv and JAV) supported by the European Commission under FP6. Funding Other funders: Netherlands Organisation for Health Research and Development and European Commission under FP6.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES


55. Hurst JA, Kiermer V. Primer: sequencing on October 6, 2023 by guest. Protected by copyright.http://jmg.bmj.com/ J Med Genet: first published as 10.1136/jmg.2009.072942 on 30 November 2009. Downloaded from


