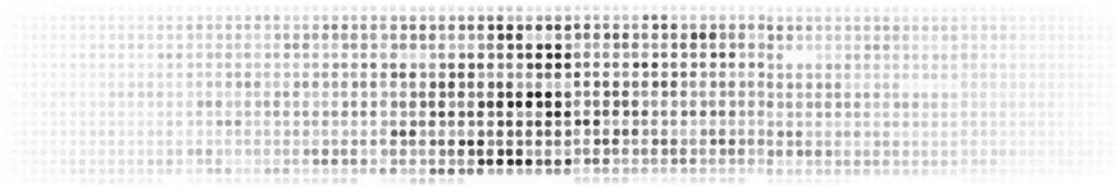
Advanced genome-wide screening in human genomic disorders

Knijnenburg, J.

Chapter 1

Introduction



1. Genetics

Genetics, the science of genes, heredity and variation of organisms, is one of the pillars of modern biology. Pioneering experiments of inheritance with pea plants were performed by Gregor Mendel between 1856 and 1863. He discovered that certain traits, like the color of seeds and the shape of pods, could be inherited in a dominant heterozygous and recessive fashion. Initially, Mendel's work received little attention; only after the rediscovery of the Mendelian inheritance, by Hugo de Vries and Carl Correns in 1900, his laws of inheritance became a basis for the understanding of genetics. Subsequently, the central role of DNA in inheritance was established in 1944, when DNA was shown to be the biochemical material responsible for transfer of genetic information in organisms.

2. Cytogenetics

2.1 Chromosomes

Karyotyping of human cells was developed in the 1950's, stimulated by the publication of Tjio and Levan in 1956 [1], who showed convincingly that human cells contain 46 chromosomes. The first constitutional genetic abnormality was found by Lejeune in 1959 [2]. He saw that the nuclei of patients with Down syndrome contained an extra small chromosome, which he linked to this syndrome. From that moment onwards, cytogenetic testing was boosted and became an important screening tool in clinical genetics. However, with no banding techniques yet available and no consensus on nomenclature, it was a difficult task. Based on morphological data, in the 1960's researchers could only classify the chromosomes into seven groups. It became clear that in congenital and prenatal disorders as well as in malignancies numerical and structural chromosome alterations are often present. Next to numerical chromosome alterations, called aneusomy, unbalanced translocations were detected causing an abnormal phenotype [3] and also deletions of chromosomes were found in patients with constitutional abnormalities [4,5]. This led to the conclusion that numerical chromosomal alterations and structural aberrations could be responsible for clinical phenotypes and encouraged the hunt for other alterations, a process that is still ongoing.

2.2 Chromosome banding

In 1968 Caspersson et al. [6] described banding of plant chromosomes with quinacrine dihydrochloride, quickly followed by the first banded human karyotype by the same group in 1970 [7]. After the development of this so-called fluorescent Q banding, several other banding techniques were developed independently of each other, including C-banding and G-banding [8]. They are based on staining with Giemsa dye and have the advantage of giving a non-fluorescent permanent staining of the chromosomes. G-banding, especially trypsin mediated G-banding, is nowadays one of the most widely used techniques for karyotyping. The technique does not require a fluorescence microscope and slides can be stored after staining for reevaluation.
Chromosome banding facilitated the recognition of individual chromosomes, as they could be identified more simply and more precisely. This led to the conclusion that the classification of chromosomes would have to change completely from the seven chromosome groups into a much more detailed system of naming chromosomes and their individual bands.
Leading researchers in human cytogenetics agreed upon a system of nomenclature, which was documented in the report of the Paris Conference (1971) [9]. This provided a basis for the description of structural and numerical rearrangements and variants. In time, a growing number of researchers and cytogeneticists joined the field of human genetics and the increasing amount of data generated resulted in updates of the nomenclature report, correcting errors and solving problems raised by increase of resolution using prometaphase chromosomes. Eventually this resulted in the publication of “An International System for Human Cytogenetic Nomenclature” (ISCN) in 1985 [10]. Continuous refining work, new findings and insights into for instance tumor cytogenetics were followed up with further updates in 1995 and 2005. The current version of the book also gives guidelines in nomenclature dealing with molecular cytogenetic techniques.
Discovery of chromosome banding has been the most important tool in cytogenetics for decades and allowed a serious leap forward in the diagnosis of genetic alterations. It is being used in the field of constitutional genetics in patients with mental retardation or developmental delay, growth retardation, congenital abnormalities and dysmorphisms, infertility and recurrent miscarriages [11]. The recognition of distinct chromosomal bands allowed scientists to discriminate between chromosomes and to make firm conclusions about trisomies, monosomies, deletions, duplications, inversions and translocations. The cause of syndromes could be pinpointed or better defined [12] and inherited traits resulting from balanced carriers were found.
To facilitate karyotyping, automated systems have been developed [13]. Complete systems combine automated metaphase finding and ranking of recorded images with segmentation and karyotyping of the chromosomes, after

which the user can check and, if needed, correct the resulting karyogram. Simpler systems lack the possibility of metaphase finding. In routine diagnostic laboratories where substantial numbers of cases are screened, these systems have become standard equipment during the last two decades.

2.3 Limitations of chromosome banding

Banding technologies also have their downsides. They heavily depend on the accumulated experience and the subjective interpretation of the cytogeneticist. Moreover, it is a time consuming technique and success relies heavily on the quality of the cytogenetic preparations, mainly concerning the compactness and spreading of the metaphase chromosomes, as this influences the resolution. In general the resolution limit is estimated to be around 5 to 10 megabases (Mb). Another important issue is the fact that there is a limit in resolving complex karyotypes, leaving cryptic alterations undetected and markers unidentified in constitutional cytogenetics. To overcome a part of these limitations, novel genetic screening approaches have been developed.

3. Molecular Cytogenetics

3.1 Fluorescence *in situ* Hybridization

In the late 1960s methods for specific detection of RNA-DNA [14] and DNA-DNA hybrids [15] became available. These *in situ* hybridization techniques were at first performed using radioactive labeled probes. A series of technical improvements in labeling chemistry, microscopy and imaging lead in the 1980s to the development of fluorescence *in situ* hybridization (FISH) using fluorescently labeled probes [16,17]. FISH is a tool that enables the microscopic detection of specific genetic regions within a morphologically intact cell. The technique requires labeling of a DNA molecule with a fluorochrome or a hapten to which fluorescently labeled antibodies are developed. For this purpose several labeling techniques are available nowadays. In general they can be divided into enzymatic or chemical labeling strategies. Initially, enzymatic incorporation of biotinylated [18] or digoxigenin-bound [19] nucleotides was the favoured method. These haptens are subsequently visualized with fluorescently labeled avidin or antibodies. Nowadays, fluorescent nucleotides are used directly in the enzymatic labeling procedure, eliminating the need for secondary detection reagents [20].
Both interphase nuclei and metaphase spreads can serve as targets for a variety of fluorescently labeled probes. With the development of this technique, together with the diversity of specific DNA probes, FISH became an important additional tool for the cytogenetic diagnostic field. Structural alterations could be better mapped and sub-microscopic resolution alterations could be detected [21].

3.2 Labeling strategies

Early chemical labeling approaches used DNA binding molecules such as acetylaminofluorene and mercury. Mercurated nucleic acids were subsequently detected with sulphydryl containing ligand carrying a hapten for indirect detection or a fluorochrome for direct detection [17]. Acetylaminofluorene binds to DNA covalently and acts directly as a hapten for immunocytochemical detection [22]. These techniques are not used anymore. Nowadays, chemical labeling is achieved using *cis*-platin labeling, also known as Universal Linkage System (ULS). In an aqueous solution at high temperatures the *cis*-platin molecule, to which a fluorochrome or hapten is coupled, predominantly binds coordinatively to the guanine-bases of DNA molecules [23]. A very diverse range of molecules can be coupled to these *cis*-platin molecules, for instance haptens, like biotin, dinitrophenol or digoxygenin, enzymes such as horseradish peroxidase as well as fluorescent dyes covering the spectrum of light.
Enzymatic labeling of DNA fragments has become the method of choice in most labs, for instance nick translation. Here the probe DNA is nicked with DNase I, and the resulting gaps are repaired using DNA polymerase I in the presence of a mix of labeled and unlabeled nucleotides [24]. Another enzymatic labeling reaction is random primed labeling, in which the probe DNA is denatured and random hexamers or octamers are allowed to anneal to the DNA. Using the Klenow fragment, a part of the DNA polymerase I enzyme, the random primers are extended in an isothermal reaction with a mix of labeled and unlabeled nucleotides [25]. Another enzymatic labeling strategy is to perform locus-specific polymerase chain reaction (PCR) or degenerate oligonucleotide primed (DOP)-PCR in the presence of a hapten or fluorochrome labeled nucleotide mix.

3.3 Types of probes

Broadly speaking, a chromosome consists of a centromere, telomeres and other repeat regions, with in between specific sequence. The total chromosomal DNA or parts of it can be used as a probe for FISH. A whole chromosome paint (WCP) or a partial chromosome paint (PCP) can for instance confirm a translocation found by G-banding. These probes are usually generated by flow sorting or microdissection of chromosomes that are subsequently randomly amplified using a polymerase chain reaction with random oligonucleotides (DOP-PCR) [26].
Usually, for chromosome ploidy determination alpha-satellite DNA specific for the centromeric region of chromosomes is used as a probe. These regions contain repetitive DNA which in most cases is unique for the chromosome to be investigated. Only the chromosomes 1/5/19, 13/21 and 14/22 cannot be distinguished using these probes, since these chromosome groups have highly homologous centromeric repeats. Instead, chromosome-specific sequences close to the centromere are selected, to minimize the chance of not detecting small

marker chromosomes. For the hybridization of these euchromatic probes, unlike for centromeric probes, cohybridization of unlabeled C_0t-1 is generally needed to suppress hybridization of repetitive sequences that are randomly present in all euchromatic DNA.
For diagnostic purposes an increasing number of commercially available, smaller sized, targeted probes have become available. Usually they consist of bacterial artificial chromosome (BAC) or cosmid vectors, containing a large insert (35-50 kb for cosmids and 100-200 kb for BACs) of specific human genomic DNA. These types of clones are used for locus-specific FISH and for detection and characterization of different chromosomal aberrations such as microdeletions, duplications and inter and intrachromosomal rearrangements such as inversions, insertions and translocations.

3.4 Types of FISH

Interphase FISH is a good screening tool for prenatal testing for the most common aneusomies like trisomy 13, 18 or 21 and aneusomies of the sex chromosomes, such as these aneusomies together form the vast majority of cases in clinical cytogenetics [27]. These whole chromosome imbalances are in general a result of nondisjunction in meiosis in one of the parents and can result in moderate to severe congenital abnormalities and delayed mental development. The development of the interphase FISH technique was a major improvement in prenatal screening of aneusomies, since it reduces the time from a laborious 7 to 14 days test with cell culture and G-banding to a rapid 24 hour test [28].
Interphase FISH can also be used for pre-implantation genetic diagnosis approaches prior to *in vitro* fertilization, both for aneusomy screening and for inheritance screening of the unbalanced results of balanced translocation carriers [29,30]. But because of the technical challenge of handling one or two cells, ethical issues and ongoing discussion with respect to governmental regulations, this procedure is not common practice in cytogenetic laboratories.
FISH on metaphase spreads is a powerful addition to conventional karyotyping. It can provide locus-specific information about alterations such as translocations, inversions, deletions and duplications. It is important to note that prior information is needed about the involved chromosomes to be able to perform these detections, except for multicolor FISH approaches that are discussed below. Metaphase FISH using whole chromosome paints can confirm translocations found with G-banding.
Locus-specific large insert clones can be used as a probe in metaphase and sometimes interphase FISH to genetically confirm suspected diagnoses in constitutional genetics. An example of this is the detection of a microdeletion syndrome like the Williams-Beuren syndrome [31], in which a part of 7q11.23 is deleted. Another application is the detection of cryptic imbalances of chromosome ends using probes specific for the gene-rich subtelomeric regions.

These chromosome imbalances are estimated to account for ~5-10% of all unexplained mental retardation [32].
Metaphase chromosomes can also be used for reverse chromosome painting. With this technique, an unidentified marker chromosome is isolated using either flow sorting [33] or microdissection [34]. The isolated DNA material is amplified, for instance by DOP-PCR, after which the amplified material is fluorescently labeled and hybridized to normal metaphase chromosomes. Marker chromosomes of unknown origin can be typed quickly this way, but there are some disadvantages using this approach. Both isolation techniques are highly specialized, and need specific knowledge and specialized equipment. When reverse painting on metaphase spreads is used to map breakpoints, the resolution is determined by the compactness of the chromosomes as in conventional karyotyping. Also, if a marker chromosome consists of heterochromatic DNA sequence or of the satellite DNA of acrocentric chromosomes, this technique will not give a conclusive answer about the origin of the marker chromosome [35].
For certain complex rearrangements, in which G-banding analysis is insufficient to determine the karyotypes, multicolor FISH approaches have been developed. Combinations of probes and fluorochromes were used to allow more detection possibilities with the available fluorescent dyes and filters to separate the fluorescence signal. This so-called combinatorial labeling allows the distinction of 2^n-1 targets with n fluorescent dyes [36].
Whole genome screening techniques require the differential labeling of all human chromosome paints with combinations of different fluorochromes. The main approaches are multicolor FISH, spectral karyotyping (SKY), and “combinatorial binary ratio labeling” (COBRA), resulting in 24 color FISH [37-39]. Subsequently the multiplicity of FISH was even further increased by introducing an extra fluorochrome to distinguish between the p and q arms of an individual chromosome [40]. In pre- and postnatal screening these multicolor approaches can facilitate the characterization of for instance unidentified or complex marker chromosomes [41,42].
A disadvantage of genome wide multicolor FISH and FISH of WCPs is that these techniques are incapable of detecting small intrachromosomal deletions, duplications, and paracentric inversions [43]. A second disadvantage is that, similar to G-banding, a limited spatial resolution of maximally about 3-5 Mb is achieved. While multicolor FISH, multicolor telomeric FISH [44] and multicolor centromeric FISH [45] can be used as whole genome scanning tools, implementation is difficult and not wide-spread because specialized equipment and analysis skills are needed and the throughput is low.

3.5 Comparative Genomic Hybridization (CGH)

In the early 1990s a genome wide analysis method was developed based on the quantitative *in situ* comparison of the genome content of a test sample to a reference sample [46]. Using this technique, whole genomic DNA of a test sample and a normal control sample is differentially labeled with haptens or with fluorochrome labeled nucleotides. The two labeled samples are cohybridized to metaphase chromosomes of a normal human control, in the presence of C_0t-1 DNA. The amounts of hybridized products relate linearly to the concentrations originally present in the samples, when hybridization equilibrium conditions are reached. Images of the metaphases are then analyzed and karyotyped using 4',6-diamidino-2-phenylindole (DAPI) counterstaining, followed by the computational measurement of the ratio of the fluorescence intensities along the whole chromosomes. A fluorescence ratio of 2:2 (=1) between the test sample and the reference sample reflects a normal copy number and is seen as a mixed color. The amount of 2 in the ratio calculation represents the copy number of a given locus or chromosome in a somatic cell. A gain or loss of a particular (part of a) chromosome will result in a 3:2 or a 1:2 readily detectable color ratio, compared to regions with equal copy number [47].
A main advantage of this genome wide screening technique is that there is no requirement for mitotic cells, which makes the technique very valuable for obtaining genotype information of certain tumors that are too complex to karyotype or too difficult to grow *in vitro* [48,49], and from paraffin-embedded tissue sections [50,51]. Furthermore a priori genetic information is not needed. The resolution of this technique depends mainly on the length of the normal metaphase chromosomes used to hybridize the samples to, and consequently resembles the resolution of G-banding studies at a limit of around 5 to 10 Mb for low copy gains or losses.
In constitutional cytogenetics, the value of this technique was appreciated as an addition to conventional karyotyping. Prenatal and postnatal cases with unsolved chromosomal aberrations as unbalanced translocations, insertions or markers of unknown origin could be correctly identified using a single CGH experiment [52,53]. It is noteworthy, however, that balanced rearrangements can not be detected using CGH.

3.6 Other molecular tools

In the field of clinical genetics obviously more molecular tools than discussed so far have been used in the past and are still being used in part. Examples are PCR and sequencing for mutation analysis and Southern blotting for copy number analysis, for instance in the fragile X syndrome [54]. Since these tools are mostly applied in region specific analysis, they will not be discussed further here.

4. Diagnosis of patients with mental retardation

On average, about 1-3% of the human population has a developmental delay or mental retardation, with or without dysmorphic features [55]. Mental retardation is a neurological disorder which is noted by for instance developmental delay of motor function and speech in infants, slow learning skills, decreased skills of interaction with other people and a low intellectual capability. The cause of mental retardation is highly variable, it can be congenital or acquired through malnutrition, poisoning, infectious disease or trauma. Congenital forms of mental retardation can have a genetic, metabolic or environmental cause and can present with or without congenital abnormalities or dysmorphic features.
A phenotypic investigation of patients with idiopathic congenital mental retardation by pediatricians and dysmorphologists is often the first step in the elucidation of the cause. Generally, in about 40-60% of the investigated patients an etiological diagnosis could be made [55], depending on study-inclusion selection criteria. The importance of detailed clinical examination is illustrated by Van Karnebeek et al. [56] and Shevell et al. [57], who state that in around 60% of cases physical examination, including dysmorphological and neurological examination, was essential for achieving the diagnosis of the patient.
The most common form of inherited mental retardation is called fragile X syndrome [58]. The syndrome has a broad spectrum of developmental delay including forms of autism and is reported with and without congenital abnormalities. Because of the high variance and relative high frequency, patients with an idiopathic mental retardation are usually screened for fragile X syndrome. Prior to the finding of the molecular cause of the fragile X syndrome phenotype it was noted that most patients showed a fragile site on chromosome Xq27, when cells were cultured under folic acid stress conditions [59]. This type of culture was used as a diagnostic tool but proved to be unreliable [59]. In 1991 genetic instability of a trinucleotide repeat was detected at the site of the fragile X breakpoint [60]. It was noted that the amount of repeats played a critical role in the expression of the downstream gene. The common number of this repeat is between 6 to 44 copies, but when the number of repeats exceeds 200 in an individual, the downstream *FMR1* is transcriptionally silenced causing fragile X syndrome. In general, patients with fragile X syndrome have inherited the expansion of repeats from the mother with an instable intermediate status of repeats called premutation. For unknown reasons, inheritance from males with a premutation generally results in offspring with lower copy number of this repeat, whereas inheritance from females often results in expansion of the trinucleotide repeat [61].
Recurrent reports of patients with phenotypes comparable to each other has led to the identification of several distinct syndromes such as Down syndrome [62], Cri-du-Chat syndrome [63], Edwards syndrome [5], Cornelia de Lange syndrome [64] and Sotos Syndrome [65]. Since the first identification of chromosomal

involvement in Down syndrome 50 years ago [2], many different genetic alterations have become known to play a role in embryonic and mental development. This knowledge sheds light on the spectrum of mental retardation and one may now conclude that the underlying genetics of mental retardation is extremely heterogeneous. With the emerging cytogenetic studies of patients with mental retardation and congenital abnormalities, genotype/phenotype correlations could be made for many syndromes, thereby proving the genetic cause of the syndrome.
As an example, Cri-du-Chat syndrome was found to have a partial deletion of chromosome 5p [63] and the Edwards syndrome was caused by a trisomy of chromosome 18 [5]. The correlation in the Sotos syndrome was harder to detect. In 2002 a Japanese group isolated the causal gene from the breakpoint of patients with Sotos syndrome and chromosomal translocations [66]. In the Cornelia de Lange syndrome (CDLS) it took even longer to find the genetic cause. In several reports of cytogenetic investigations of CDLS, chromosomal translocations and imbalances were found [67,68]. But heterogeneity in the phenotype caused confusion whether the described alterations were the main cause for CDLS [69]. Based on balanced translocations in CDLS patients Tonkin et al. [70] found the involvement of a gene called NIPBL and demonstrated mutations in this gene in more than half of their patients. Involvements of other genes in the same protein complex are now recognized to cause an X-linked [71] and a milder form of the CDL syndrome [72].

5. Development of new screening tools

5.1 Array Comparative Genomic Hybridization (Array-CGH)

In the last decade of the previous century researchers realized that improvement of the resolution of CGH could be established by using thousands or millions of DNA fragments each specific to a unique location of the genome immobilized on a glass surface as a target, instead of metaphase chromosomes. This technique was initially called matrix comparative genomic hybridization [73], nowadays the name microarray based or array comparative genomic hybridization (array-CGH) is generally used [74].
The basis of this technique was partly the Human Genome Project, a collective effort to sequence the whole human genome. As a spin-off large BAC and PAC clone libraries became available that contained well characterized and mapped human genomic DNA fragments of about 100 to 150 kb, covering nearly the whole genome. Selected clones were used for spotting on microscope slides. The hybridization principle of these printed array slides was comparable to classical chromosome based CGH hybridizations (Figure 1).
In the first protocols, the target DNA to be printed was isolated from bacterial cultures and after cleanup under the right conditions directly spotted on the

array slides [73,74]. To generate enough material to comfortably print these arrays, linker adapter PCR [75] or degenerate oligonucleotide primed PCR (DOP-PCR) protocols were used [26,76] and were optimized [77]. In the latter paper Fiegler et al. suggested an underestimation of the fluorescence ratio of single copy changes that was seen in previous array-CGH publications was not mainly due to the incomplete suppression of repeat sequences, but was for a large part caused by the contamination of bacterial genomic sequence in the spotted clones. To overcome this problem, they designed degenerate oligonucleotide primers different from the commonly used 6MW DOP primer [26], which the help of the knowledge of the sequence from the human genome project.

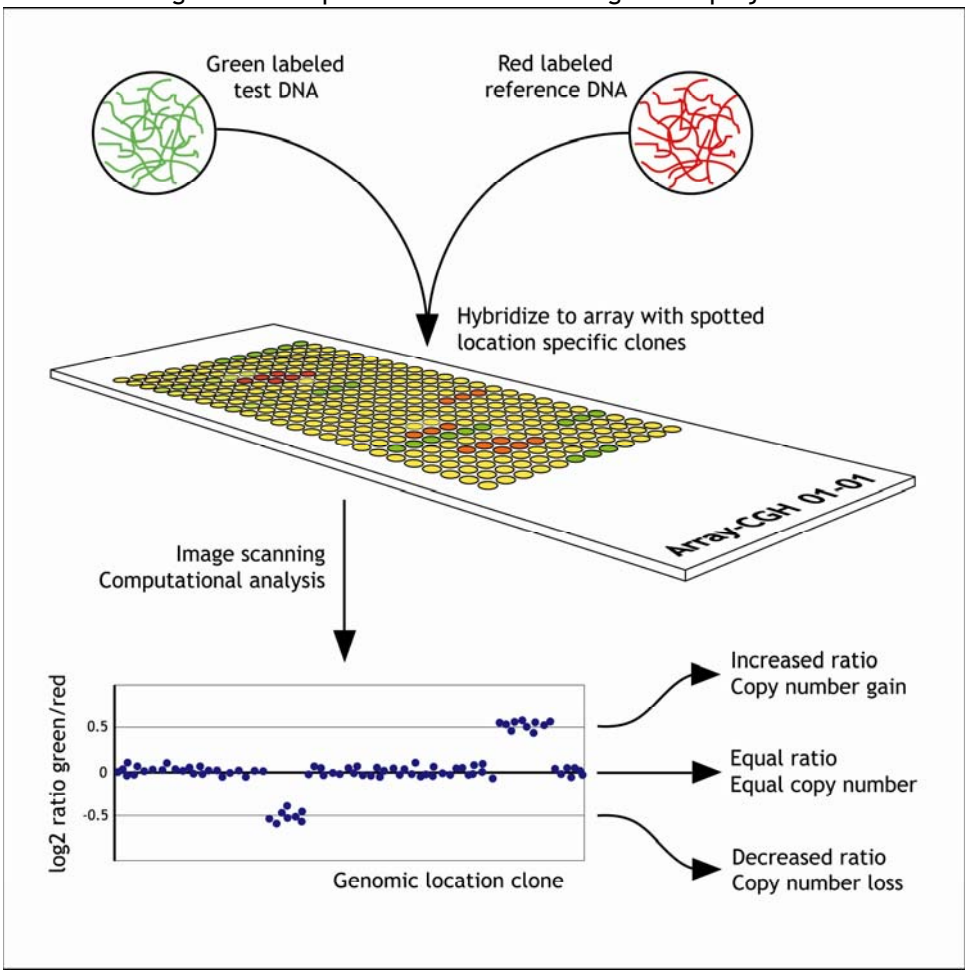


Figure 1. The principle of array-CGH

These DOP-PCR primers preferentially amplify human genomic DNA over *Escherichia coli (E. coli)* DNA because the primers have an averaged frequency of

matches per kilobase of 0.58 based on the human genomic reference sequence, compared to an average of 0.03 in *E. coli* DNA. As a comparison, the original 6MW primer has an averaged frequency of matches per kilobase of 0.65 and 0.40 for human and *E. coli* DNA, respectively. Using this approach the DOP-PCR generates less background and produces a more evenly represented product.
The first reports on array-CGH described a targeted array as a proof of principle [73,74]. Later, whole genome arrays were developed that covered the genome with a resolution of about one clone per megabase of euchromatin DNA, resulting in about 2400-3500 different clones per array [75,77], often printed in duplicate or triplicate. This development led to an increase of resolution of about 10 times compared to the conventional CGH technique, entering sub-microscopic detection levels of copy number changes in a genome wide fashion.
Imaging and analysis of the arrays was at first done with a CCD camera and ratios were manually calculated and presented in spreadsheet programs. Nowadays, scanning and analysis of arrays is almost fully automated. Hybridized arrays are mostly scanned using a laser scanner and the resulting images are then processed using commercial software. These software packages perform background corrections, remove outlier spots, normalize the data, calculate ratio values and display the data in comprehensive graphs. In general the ratio data is displayed as a 2log value, to better distinguish deletions from normal values.
In theory the resolution of array-CGH is unlimited, depending on the spacing of the used clones and their size. When the array-CGH technique was proven to work satisfactorily, several institutes developed large clone insert arrays tiling the whole euchromatic part of the genome [78-80]. Due to the performance of needle or inkjet printing techniques available, all these clones could be printed on one glass slide, but not in duplicate or triplicate. Consequently a second hybridization experiment is often needed to corroborate these results [79,80].
Spotted or immobilized oligonucleotides provide some advantages for the detection of copy number changes. They avoid the need for bacterial cultures and isolation of DNA and/or PCR amplification of large insert clones. Large oligonucleotide libraries are spotted and hybridized in a way comparable to other types of array-CGH [81], yielding similar results.
Commercial platforms for copy number analysis are now available, using oligonucleotides, photo-lithographically synthesized on the chips [82,83]. One type is the SNP-based genotype array platform that can be used for allelotyping and copy number analysis in the same experiment. An advantage of the system is the ability to detect copy number neutral alterations as uniparental disomy or copy number neutral homozygosity. A disadvantage of this platform is the lack of an internal quality control as a cohybridized reference sample, because these genotyping arrays are single color experiments. Other oligonucleotide platforms are hybridized using a similar test/reference principle as CGH and array-CGH. Results produced with these platforms need somewhat more statistical analysis compared to data produced with large insert clone arrays, such as binning of

data points and a moving average to reduce false positive calls. With ongoing technical improvement and advances in statistics, oligonucleotide platforms are increasingly robust.
Targeted arrays have been specially designed for specific regions in the genome. For instance, in constitutional genetics several institutes that investigate the role of the X chromosome in mental retardation developed targeted tiling path arrays for chromosome X [84-86]. This X-linked mental retardation is of interest to researchers, since there is a clear overrepresentation of males in mental retardation of 30 to 40% suggesting involvement of the X chromosome in many cases.

5.2 Paired-end mapping

In 2005 a DNA sequencing method was introduced [87], which allowed large-scale whole genome analysis of genomic DNA samples to screen for copy number alterations. The method is called paired-end mapping, in which genomic DNA was fragmented into pieces of about 3 kb. From these fragments the two ends were sequenced and mapped back on the reference sequence [88]. If the distance between the matches on the reference sequence was within certain cutoff values, no alteration was detected. If the distance between the matches in the reference genome was bigger than the high cutoff value, the analyzed sample had a deletion between the sequenced ends. When the distance was smaller than the low cutoff value, an insertion was detected in the tested genome. This approach allows detection of inversions as well, if the two matches mapped in different relative orientations on the reference genome. This feature offers a big advantage over array based whole genome screening, since these platforms cannot detect inversions. In the report of Korbel et al. the whole genome was covered 2.1 fold for one sample and 4.3 fold for a second sample, so on average every fragment was analyzed at least two times. Similar techniques are expected to play a role in whole genome analysis in the near future.

5.3 Target specific screening tools

In order to confirm the alterations found using whole genome screening tools as described before, and to investigate targeted regions for suspected specific copy number alterations as microdeletions and microduplications, several techniques have been used and developed. Although FISH has been the method of choice in many cytogenetic labs, different molecular techniques have been developed and standardized providing some advantages over FISH. Mostly, these techniques are based on quantitative amplification of DNA. Quantitative real-time polymerase chain reaction (qPCR) is based on the amplification of genomic DNA sequences with fluorescently labeled primers in a quantitative manner and the amplification is monitored in real-time. The number of cycles to reach a certain fluorescence

level in a test locus and a reference locus in one sample is used to estimate a relative copy number [89].
Two other quantitative techniques called multiplex amplifiable probe hybridization (MAPH) and multiplex ligation dependent probe amplification (MLPA) are based on quantitative amplification of DNA fragments matching to genomic DNA. These fragments contain sequences 5' and 3' unrelated to human DNA, which are recognized with a universal primer pair. For MAPH, the fragments to be amplified are cloned parts of exons, for instance. They are hybridized to genomic DNA of a test sample which is cross-linked to a membrane. After a stringent wash only the hybridized fragments are available for quantitative PCR with the universal primer pair [90-92]. In MLPA the fragments to be amplified are generated using two probes adjacently annealed onto genomic test DNA of interest, followed by ligation to each other. Multiple probes recognizing different regions can be combined in one ligation reaction and only the ligation products can be used for amplification with universal primers [93]. Quantitative readout of the multiple fragments both for MAPH and MLPA is generally done with fluorochromes bound to the universal primers and a capillary sequencer.
Compared to FISH these techniques have the advantage that they can be performed using multiple probes to screen for multiple regions per sample (up to 60 regions with dual color MLPA [93]) on multiple samples per experiment.
Recent developments include efforts to analyze parts of the genome with high-throughput resequencing [87,94]. These sequencing methods are based on random sequencing short stretches of a pool of fragments representing specific parts of the genome and aligning those sequences back to the reference sequence. Different groups have applied this next generation sequencing to enriched DNA targets. Specifically designed oligonucleotide microarrays have been used to capture parts of the genome, such as genes and locus specific regions [95,96]. The yielded DNA was used for massive parallel sequencing and it proved to be feasible for mutation screening. For copy number analysis based on quantitative analysis of sequence reads per region, which in theory is possible, the current capture methods are not uniform enough [97]. Nevertheless, this type of detailed region-specific copy number analysis may play a role in the future.

6. Applications of Array-CGH in constitutional genetics

The development of the genome wide array-CGH technique opened new possibilities for genetic screening in constitutional genetics. In individuals with idiopathic mental retardation with or without congenital abnormalities often the causal genetic diagnosis remained unclear if the karyotype revealed no alterations in the G-banding resolution. Several feasibility studies showed the additional value of array-CGH to conventional cytogenetics [79,98-103], through gain of resolution and robustness. In a research setting this technique has a

proven detection rate of genetic alterations of 10 to 15%, depending on the selection criteria. Alterations found are mainly interstitial deletions or duplications not detected by classical cytogenetic screening [104]. The deletions and duplications appeared to be scattered over the genome with very little overlap, proving the heterogeneous nature of genetic causes for mental retardation. However, despite this heterogeneity the genetic cause for CHARGE syndrome could be unveiled because of detailed investigation of recurrent genetic alteration data in combination with a precise phenotypic description. Two unrelated patients with a similar phenotype described as CHARGE syndrome were found to have an overlapping microdeletion on chromosome 8q12. In subsequent investigations of patients with the same syndrome but showing no microdeletion, mutations in a gene called *CHD7*, located in the same genetic region, were found in almost 60% of cases [105].
To investigate the involvement of regions flanked by large segmental duplications in recurrent genomic disorders, Sharp et al. [106] designed an array containing 2007 BAC clones concentrated around 130 identified sites which have a unique sequence flanked by segmental duplications. It was shown previously that these regions may be involved in recurrent rearrangements because of large homologies between the flanking segmental duplications, causing non-allelic homologous recombination [107]. They found 16 pathogenic rearrangements of regions flanked by segmental duplication when applying this array to investigate 290 patients with mental retardation, including 4 patients with a similar deletion on 17q21.31. A similar deletion was also found by three other groups [108-110]. This microdeletion of about 600 kb in size causes a comparable phenotype with a moderate developmental delay, severe hypotonia, amiable behavior and characteristic facial features with a long face, a bulbous nasal tip and a broad chin. The genomic region of 17q21.31 contains at the site of the microdeletion a common 900 kb inversion polymorphism which is present in about 20% of the European, Icelandic and Middle Eastern population [111]. A direct consequence of this inversion polymorphism is the appearance of a directly oriented segmental duplication, causing this lineage to be subject to non-allelic homologous recombination. In the non-inverted lineage these repeats are reverse oriented which prevents recombination. As a result of the relative high frequency of inversion polymorphism carriers, this microdeletion is among the most common causes of a microdeletion syndrome with an estimated prevalence of around 1 in 13,000 to 1 in 20,000 [109] and an estimated frequency of around 1% in patients with mental retardation [112]. Finding the reciprocal duplication of this region is a relatively logical consequence of the presumed non-allelic homologous recombination mechanism [113].
Clinical studies of large numbers of patients using arrays specifically designed for certain genomic regions have been performed as well [114,115]. These specifically designed arrays contain probes for known microdeletion and subtelomeric regions, aiming to detect deletions and duplications which are

clinically relevant. In contrast, high resolution arrays covering the whole genome, as previously described, can also detect alterations of unknown clinical consequence. The concept of not including detection of regions of unknown clinical consequence gives a clearer diagnosis in testing of patients with idiopathic mental retardation and also in patients with an atypical appearance of a syndrome linked to a known genetic rearrangement. This approach is very useful to confirm suspected genetic alterations as well, but the obvious disadvantage is that unsuspected disease-causing genetic alterations remain undetected.

7. Genetic variation

The assumption that the genome of two unrelated individuals is 99.9% identical still leaves room for millions of different base pairs. Genetic variation can either be the cause of a disease, a predisposition to a disease, a defense against a disease or a normal variation. It may comprise variation at the nucleotide level, variation of repeat rich elements interspersed throughout the genome, copy number variations (CNVs) of for example segmental duplications or structural variations. Whether a detected variation is normal or predisposing to disease can be difficult to assess, since predispositions might be multifactorial, or they might have a low penetrance or a combination of both. Some variations of the genome which are not disease-causing at all for the individual itself like a Robertsonian translocation or most reciprocal translocations, can not be seen as normal variation, since there is a high chance that these balanced structural aberrations can have serious consequences for the next generation. In contrast, for instance the balanced inversion at chromosome 17q21.31 described earlier, from which about 20% of the European population is a carrier, is considered a normal variant. Although it predisposes to microdeletion or microduplication in the 17q21.31 region [109,110], the risk for the recombination event is in the same range as in other microdeletion/microduplication syndromes and is thus negligible.

7.1 Heteromorphisms

Already prior to chromosome banding, investigators recognized chromosome variations between individuals, called heteromorphisms. At the time of the first conference on standardization in human cytogenetics in 1960 [116] the considerable variation in the length of the Y chromosome was already noted. Morphological variations present near the centromere of several chromosomes as well as in the short arms of D- and G-group chromosomes were found later. Through large studies on newborns it was concluded that these variations probably were normal heteromorphisms [117]. The first heteromorphism that was linked to a specific trait was a change in condensation of the part of the chromosome directly below the centromere of chromosome 1. It was noted in

three different families that the Duffy blood group type segregated with the chromosome 1 with relatively uncoiled chromatids in the paracentric region, leading to the conclusion that the gene for the Duffy blood group was probably located on this chromosome in the vicinity of this paracentric region [118].
In general, these microscopically visible heteromorphisms usually involve the heterochromatic regions and satellites of chromosomes. Less frequently, euchromatic variants are also reported [119]. These variants may have hidden phenotypic consequences which are not always directly evident.

7.2 Single Nucleotide Polymorphisms

At the beginning of this century, it was generally believed that the biggest part of genetic variation between two individuals was based on single nucleotide polymorphisms (SNPs) [120]. Only recently submicroscopic copy number variation was proven to be a major contributor to genetic variation. It was estimated that there are at least 10 million SNPs in the human population, an average of 1 SNP every 300 base pairs. In 2001 an international SNP map working group had mapped 1.42 million SNPs [121] and in 2005 approximately 1.3 million SNPs were truly genotyped [122] in the HapMap collection. This is a collection of genomic DNA samples from four different populations: 30 parent-offspring trios of the Yoruba population of Nigeria, 30 trios of European descent from Utah, 45 unrelated Han Chinese from Beijing, China, and 45 unrelated Japanese from Tokyo, Japan. The International HapMap study is an extension of the Human Genome Project. The latter project generated the human reference sequence, giving information for the non-variant part of the genome, while the International HapMap study aims for cataloguing the nucleotides that can vary between individuals. More recently the consortium released an updated version of the map in which 3.1 million SNPs were mapped [123].

7.3 Copy Number Variation (CNV)

Initially the frequency of larger genomic variations between individuals was expected to be low and generally disease-causing. Using novel emerging array-CGH techniques and SNP array platforms, submicroscopic variation of genomic fragments was detected [88,106,124-132]. The detected amount of variation causing no obvious phenotype was much larger than expected [133]. It was found that a considerable part of copy number variation was associated with the 5% of the genome that is present in segmental duplications in the genome. Not only were these CNV regions proposed to be involved in disease causing rearrangements through non-allelic homologous recombination [134], they were also associated with large scale normal variation [124,125] since many of these copy number variable regions contain genes with a variable expression that contributes to normal phenotypic variation [135,136].

These findings stressed the importance of extensively investigating the human genome in several different populations to assess if a reported CNV could be benign or disease causing. Several groups studied genetic variations in HapMap samples or normal healthy men and women, with different platforms and techniques. To profit most from this data and to share all this information on variation with the rest of the research community, all data was collected in the Database of Genomic Variants (http://projects.tcag.ca/variation) [125]. Still growing, it has become a valuable tool to evaluate if a found CNV is assumed to be harmless or could be possibly causing a disease. Nevertheless care must be taken when using this database to filter out benign CNVs from disease-causing, even if the CNV was inherited from a phenotypically normal parent. A CNV could theoretically be disease-causing in one individual and be benign in another, for instance because of reduced penetrance, variation of expression of the unaffected allele through a different genetic background or because of a recessive trait.

8. Scope of this thesis

Modern cytogenetics has experienced rapid technical changes and improvements since first determining the correct number of chromosomes in humans. This resulted in new methods for improvement of analysis of chromosomal alterations of which many have been described in the introduction of this thesis. Particularly, the last decade has had a big impact on the concept of cytogenetics with the introduction of whole genome molecular screening techniques. Chapter 2 of this thesis describes the implementation of array-CGH and its effect on the diagnosis of structural chromosome rearrangements. In chapter 3 the additional value of array-CGH to routine G-banding of patients with mental retardation and congenital abnormalities is described. A significant increase in the frequency of both inherited and *de novo* copy number variations in a group of 81 patients was found, in which no structural chromosome rearrangements were visible with conventional G-banding. This chapter and additional research combining array-CGH and MLPA, as described in chapter 4, have led to the insight that the current standard of routine whole genome screening with mainly G-banding needs to be reconsidered. What particular techniques will be adopted within diagnostics will depend on cost, quality and speed of the analysis procedure.
Chapter 5 of this thesis underlines the additional value of array-CGH to banding and FISH studies. It reveals the true complexity of a chromosome rearrangement that initially was thought to be a balanced three-way translocation. The mechanism of formation of a complex marker ring chromosome is explained in chapter 6. It was unraveled with array-CGH, combined with various other molecular techniques as FISH, MLPA and oligo array-CGH. With this investigation a novel translocation mechanism was shown to be involved in the formation of a ring chromosome.

In chapter 7, the molecular and clinical characterization of a new microduplication 3q29 syndrome is described, showing that copy number variation of this region may cause a heterogeneous phenotype. Extensive documentation of disease association with copy number variation will be important to accurately distinguish normal variation from variation causing disease, as illustrated in chapter 8. It describes a case with a homozygous deletion of a normal variation locus causing hearing loss. Both chapter 7 and 8 also show the importance of international collaboration on collecting trusted copy number variation data in databases that will be accessible to the entire community.
In chapter 9 the future impact of the molecular techniques in clinical cytogenetic research and diagnostics is discussed.

9. References

1. Tjio JH, Levan A. The Chromosome Number of Man. *Hereditas* 1956;**42**:U1-U6.
2. Lejeune J, Turpin R, Gautier M. Le Mongolisme Premier Exemple Daberration Autosomique Humaine. *Annales de Genetique* 1959;**1**:41-9.
3. Penrose LS, Ellis JR, Delhanty JD. Chromosomal translocations in mongolism and in normal relatives. *Lancet* 1960;**2**:409-10.
4. Wolf U, Reinwein H, Porsch R, Schroter R, Baitsch H. Deficiency on the short arms of a chromosome No. 4. *Humangenetik* 1965;**1**:397-413.
5. Edwards JH, Harnden DG, Cameron AH, Crosse VM, Wolff OH. A new trisomic syndrome. *Lancet* 1960;**1**:787-90.
6. Caspersson T, Farber S, Foley GE, Kudynowski J, Modest EJ, Simonsson E, Wagh U, Zech L. Chemical differentiation along metaphase chromosomes. *Exp Cell Res* 1968;**49**:219-22.
7. Caspersson T, Zech L, Johansson C, Modest EJ. Identification of human chromosomes by DNA-binding fluorescent agents. *Chromosoma* 1970;**30**:215-27.
8. Sumner AT, Evans HJ, Buckland RA. New technique for distinguishing between human chromosomes. *Nat New Biol* 1971;**232**:31-2.
9. Paris Conference (1971): standardization in human cytogenetics. *Birth Defects Orig Artic Ser* 1972;**8**:1-46.
10. An International System for Human Cytogenetic Nomenclature (1985) ISCN 1985. Report of the Standing Committee on Human Cytogenetic Nomenclature. *Birth Defects Orig Artic Ser* 1985;**21**:1-117.
11. Carpenter NJ. Molecular cytogenetics. *Semin Pediatr Neurol* 2001;**8**:135-46.
12. Yunis JJ. High resolution of human chromosomes. *Science* 1976;**191**:1268-70.
13. Lundsteen C, Martin AO. On the selection of systems for automated cytogenetic analysis. *Am J Med Genet* 1989;**32**:72-80.
14. Gall JG, Pardue ML. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci U S A* 1969;**63**:378-83.
15. Pardue ML, Gall JG. Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc Natl Acad Sci U S A* 1969;**64**:600-4.
16. Bauman JG, Wiegant J, Borst P, van Duijn P. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. *Exp Cell Res* 1980;**128**:485-90.
17. Hopman AH, Wiegant J, Raap AK, Landegent JE, van der Ploeg M, van Duijn P. Bi-color detection of two target DNAs by non-radioactive in situ hybridization. *Histochemistry* 1986;**85**:1-4.
18. Langer PR, Waldrop AA, Ward DC. Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. *Proc Natl Acad Sci U S A* 1981;**78**:6633-7.
19. Holtke HJ, Kessler C. Non-radioactive labeling of RNA transcripts in vitro with the hapten digoxigenin (DIG); hybridization and ELISA-based detection. *Nucleic Acids Res* 1990;**18**:5843-51.
20. Wiegant J, Ried T, Nederlof PM, van der ploeg M, Tanke HJ, Raap AK. *In situ* hybridization with fluoresceinated DNA. *Nucleic Acids Res* 1991;**19**:3237-41.
21. Beatty B, Mai S, Squire J, 2002. FISH, a practical approach. Squire J, Marrano P, Kolomietz E, editors. FISH in clinical cytogenetics. 183-203Oxford, Oxford University Press.
22. Landegent JE, Jansen in de Wal N, Baan RA, Hoeijmakers JH, van der Ploeg M. 2-Acetylaminofluorene-modified probes for the indirect hybridocytochemical detection of specific nucleic acid sequences. *Exp Cell Res* 1984;**153**:61-72.
23. Wiegant JC, van Gijlswijk RP, Heetebrij RJ, Bezrookove V, Raap AK, Tanke HJ. ULS: a versatile method of labeling nucleic acids for FISH based on a monofunctional reaction of cisplatin derivatives with guanine moieties. *Cytogenet Cell Genet* 1999;**87**:47-52.
24. Kelly RB, Cozzarelli NR, Deutscher MP, Lehman IR, Kornberg A. Enzymatic synthesis of deoxyribonucleic acid. XXXII. Replication of duplex deoxyribonucleic acid by polymerase at a single strand break. *J Biol Chem* 1970;**245**:39-45.
25. Feinberg AP, Vogelstein B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 1983;**132**:6-13.
26. Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 1992;**13**:718-25.
27. Rickman L, Fiegler H, Shaw-Smith C, Nash R, Cirigliano V, Voglino G, Ng BL, Scott C, Whittaker J, Adinolfi M, et al. Prenatal detection of unbalanced chromosomal rearrangements by array CGH. *J Med Genet* 2006;**43**:353-61.
28. Xu J, Chen Z. Advances in molecular cytogenetics for the evaluation of mental retardation. *Am J Med Genet C Semin Med Genet* 2003;**117**:15-24.
29. Munne S, Fung J, Cassel MJ, Marquez C, Weier HU. Preimplantation genetic analysis of translocations: case-specific probes for interphase cell analysis. *Hum Genet* 1998;**102**:663-74.

30. Munne S, Magli C, Bahce M, Fung J, Legator M, Morrison L, Cohert J, Gianaroli L. Preimplantation diagnosis of the aneuploidies most commonly found in spontaneous abortions and live births: XY, 13, 14, 15, 16, 18, 21, 22. *Prenat Diagn* 1998;**18**:1459-66.
31. Ewart AK, Morris CA, Atkinson D, Jin W, Sternes K, Spallone P, Stock AD, Leppert M, Keating MT. Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nat Genet* 1993;**5**:11-6.
32. Knight SJ, Lese CM, Precht KS, Kuc J, Ning Y, Lucas S, Regan R, Brenan M, Nicod A, Lawrie NM, et al. An optimized set of human telomere clones for studying telomere integrity and architecture. *Am J Hum Genet* 2000;**67**:320-32.
33. Carter NP, Ferguson-Smith MA, Perryman MT, Telenius H, Pelmear AH, Leversha MA, Glancy MT, Wood SL, Cook K, Dyson HM, et al. Reverse chromosome painting: a method for the rapid analysis of aberrant chromosomes in clinical cytogenetics. *J Med Genet* 1992;**29**:299-307.
34. Schermelleh L, Thalhammer S, Heckl W, Posl H, Cremer T, Schutze K, Cremer M. Laser microdissection and laser pressure catapulting for the generation of chromosome-specific paint probes. *Biotechniques* 1999;**27**:362-7.
35. Blennow E. Reverse painting highlights the origin of chromosome aberrations. *Chromosome Res* 2004;**12**:25-33.
36. Nederlof PM, van der Flier S, Wiegant J, Raap AK, Tanke HJ, Ploem JS, van der Ploeg M. Multiple fluorescence in situ hybridization. *Cytometry* 1990;**11**:126-31.
37. Schrock E, du Manoir S, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA, Ning Y, Ledbetter DH, Bar-Am I, Soenksen D, et al. Multicolor spectral karyotyping of human chromosomes. *Science* 1996;**273**:494-7.
38. Speicher MR, Gwyn BS, Ward DC. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat Genet* 1996;**12**:368-75.
39. Tanke HJ, Wiegant J, van Gijlswijk RP, Bezrookove V, Pattenier H, Heetebrij RJ, Talman EG, Raap AK, Vrolijk J. New strategy for multi-colour fluorescence in situ hybridisation: COBRA: COmbined Binary RAtio labelling. *Eur J Hum Genet* 1999;**7**:2-11.
40. Wiegant J, Bezrookove V, Rosenberg C, Tanke HJ, Raap AK, Zhang H, Bittner M, Trent JM, Meltzer P. Differentially painting human chromosome arms with combined binary ratio-labeling fluorescence in situ hybridization. *Genome Res* 2000;**10**:861-5.
41. Uhrig S, Schuffenhauer S, Fauth C, Wirtz A, Daumer-Haas C, Apacik C, Cohen M, Muller-Navia J, Cremer T, Murken J, et al. Multiplex-FISH for pre- and postnatal diagnostic applications. *Am J Hum Genet* 1999;**65**:448-62.
42. Bezrookove V, Hansson K, van der Burg M, van der Smagt JJ, Hilhorst-Hofstee Y, Wiegant J, Beverstock GC, Raap AK, Tanke HJ, Breuning MH, et al. Individuals with abnormal phenotype and normal G-banding karyotype: improvement and limitations in the diagnosis by the use of 24-colour FISH. *Hum Genet* 2000;**106**:392-8.
43. Tonnies H. Modern molecular cytogenetic techniques in genetic diagnostics. *Trends Mol Med* 2002;**8**:246-50.
44. Engels H, Ehrbrecht A, Zahn S, Bosse K, Vrolijk H, White S, Kalscheuer V, Hoovers JM, Schwanitz G, Propping P, et al. Comprehensive analysis of human subtelomeres with combined binary ratio labelling fluorescence in situ hybridisation. *Eur J Hum Genet* 2003;**11**:643-51.
45. Nietzel A, Rocchi M, Starke H, Heller A, Fiedler W, Wlodarska I, Loncarevic IF, Beensen V, Claussen U, Liehr T. A new multicolor-FISH approach for the characterization of marker chromosomes: centromere-specific multicolor-FISH (cenM-FISH). *Hum Genet* 2001;**108**:199-204.
46. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992;**258**:818-21.
47. Ried T, Liyanage M, du Manoir S, Heselmeyer K, Auer G, Macville M, Schrock E. Tumor cytogenetics revisited: comparative genomic hybridization and spectral karyotyping. *J Mol Med* 1997;**75**:801-14.
48. Schrock E, Thiel G, Lozanova T, du Manoir S, Meffert MC, Jauch A, Speicher MR, Nurnberg P, Vogel S, Janisch W, et al. Comparative genomic hybridization of human malignant gliomas reveals multiple amplification sites and nonrandom chromosomal gains and losses. *Am J Pathol* 1994;**144**:1203-18.
49. Ried T, Petersen I, Holtgreve-Grez H, Speicher MR, Schrock E, du Manoir S, Cremer T. Mapping of multiple DNA gains and losses in primary small cell lung carcinomas by comparative genomic hybridization. *Cancer Res* 1994;**54**:1801-6.
50. Speicher MR, du Manoir S, Schrock E, Holtgreve-Grez H, Schoell B, Lengauer C, Cremer T, Ried T. Molecular cytogenetic analysis of formalin-fixed, paraffin-embedded solid tumors by comparative genomic hybridization after universal DNA-amplification. *Hum Mol Genet* 1993;**2**:1907-14.
51. Isola J, DeVries S, Chu L, Ghazvini S, Waldman F. Analysis of changes in DNA sequence copy number by comparative genomic hybridization in archival paraffin-embedded tumor samples. *Am J Pathol* 1994;**145**:1301-8.
52. Bryndorf T, Kirchhoff M, Rose H, Maahr J, Gerdes T, Karhu R, Kallioniemi A, Christensen B, Lundsteen C, Philip J. Comparative genomic hybridization in clinical cytogenetics. *Am J Hum Genet* 1995;**57**:1211-20.

53. Levy B, Dunn TM, Kaffe S, Kardon N, Hirschhorn K. Clinical applications of comparative genomic hybridization. *Genet Med* 1998;**1**:4-12.
54. Brown WT. Perspectives and molecular diagnosis of the fragile X syndrome. *Clin Lab Med* 1995;**15**:859-75.
55. Curry CJ, Stevenson RE, Aughton D, Byrne J, Carey JC, Cassidy S, Cunniff C, Graham JM, Jr., Jones MC, Kaback MM, et al. Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics. *Am J Med Genet* 1997;**72**:468-77.
56. van Karnebeek CD, Scheper FY, Abeling NG, Alders M, Barth PG, Hoovers JM, Koevoets C, Wanders RJ, Hennekam RC. Etiology of mental retardation in children referred to a tertiary care center: a prospective study. *Am J Ment Retard* 2005;**110**:253-67.
57. Shevell MI, Majnemer A, Rosenbaum P, Abrahamowicz M. Etiologic yield of subspecialists' evaluation of young children with global developmental delay. *J Pediatr* 2000;**136**:593-8.
58. Crawford DC, Acuna JM, Sherman SL. FMR1 and the fragile X syndrome: human genome epidemiology review. *Genet Med* 2001;**3**:359-71.
59. Hecht F, Jacky PB, Sutherland GR. The fragile X chromosome: current methods. *Am J Med Genet* 1982;**11**:489-95.
60. Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, Schlessinger D, Sutherland GR, Richards RI. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* 1991;**252**:1711-4.
61. Garber KB, Visootsak J, Warren ST. Fragile X syndrome. *Eur J Hum Genet* 2008;**16**:666-72.
62. Down JLH. Observations of an ethnic classification of idiots. *Clinical Lecture Reports, London Hospital* 1866;**3**:259-62.
63. Lejeune J, Lafourcade J, De Grouchy J, Berger R, Gautier M, Salmon C, Turpin R. Partial deletion of the short arm of chromosome 5. Indivdualization of a new morbid state. *Sem Hop* 1964;**40**:1069-79.
64. de Lange C. Sur un type nouveau de degeneresccence (typus Amstelodamensis). *Arch Med Enfants* 1933;**36**:713-9.
65. Sotos JF, Dodge PR, Muirhead D, Crawford JD, Talbot NB. Cerebral gigantism in childhood. A syndrome of exessively rapid growth and acromegalic features and a nonprogressive neurologic disorde. *N Engl J Med* 1964;**271**:109-16.
66. Kurotaki N, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, Ohashi H, Naritomi K, Tsukahara M, Makita Y, et al. Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat Genet* 2002;**30**:365-6.
67. Craig AP, Luzzatti L. Translocation in de Lange's Syndrome? *Lancet* 1965;**2**:445-6.
68. Falek A, Schmidt R, Jervis GA. Familial de Lange syndrome with chromosome abnormalities. *Pediatrics* 1966;**37**:92-101.
69. Breslau EJ, Disteche C, Hall JG, Thuline H, Cooper P. Prometaphase chromosomes in five patients with the Brachmann-de Lange syndrome. *Am J Med Genet* 1981;**10**:179-86.
70. Tonkin ET, Wang TJ, Lisgo S, Bamshad MJ, Strachan T. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat Genet* 2004;**36**:636-41.
71. Musio A, Selicorni A, Focarelli ML, Gervasini C, Milani D, Russo S, Vezzoni P, Larizza L. X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat Genet* 2006;**38**:528-30.
72. Deardorff MA, Kaur M, Yaeger D, Rampuria A, Korolev S, Pie J, Gil-Rodriguez C, Arnedo M, Loeys B, Kline AD, et al. Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am J Hum Genet* 2007;**80**:485-94.
73. Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 1997;**20**:399-407.
74. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998;**20**:207-11.
75. Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, et al. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* 2001;**29**:263-4.
76. Hodgson G, Hager JH, Volik S, Hariono S, Wernick M, Moore D, Nowak N, Albertson DG, Pinkel D, Collins C, et al. Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. *Nat Genet* 2001;**29**:459-64.
77. Fiegler H, Carr P, Douglas EJ, Burford DC, Hunt S, Scott CE, Smith J, Vetrie D, Gorman P, Tomlinson IP, et al. DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer* 2003;**36**:361-74.
78. Ishkanian AS, Malloff CA, Watson SK, deLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, et al. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet* 2004;**36**:299-303.

79. de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM, Reijmersdal S, Nillesen WM, Huys EH, de Leeuw N, et al. Diagnostic genome profiling in mental retardation. *Am J Hum Genet* 2005;**77**:606-16.
80. Fiegler H, Redon R, Andrews D, Scott C, Andrews R, Carder C, Clark R, Dovey O, Ellis P, Feuk L, et al. Accurate and reliable high-throughput detection of copy number variation in the human genome. *Genome Res* 2006;**16**:1566-74.
81. Carvalho B, Ouwerkerk E, Meijer GA, Ylstra B. High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. *J Clin Pathol* 2004;**57**:644-6.
82. Zhao X, Li C, Paez JG, Chin K, Janne PA, Chen TH, Girard L, Minna J, Christiani D, Leo C, et al. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 2004;**64**:3060-71.
83. Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R, Tsang P, Curry B, Baird K, Meltzer PS, et al. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc Natl Acad Sci U S A* 2004;**101**:17765-70.
84. Froyen G, Van Esch H, Bauters M, Hollanders K, Frints SG, Vermeesch JR, Devriendt K, Fryns JP, Marynen P. Detection of genomic copy number changes in patients with idiopathic mental retardation by high-resolution X-array-CGH: important role for increased gene dosage of XLMR genes. *Hum Mutat* 2007;**28**:1034-42.
85. Veltman JA, Yntema HG, Lugtenberg D, Arts H, Briault S, Huys EH, Osoegawa K, de Jong P, Brunner HG, Geurts van Kessel A, et al. High resolution profiling of X chromosomal aberrations by array comparative genomic hybridisation. *J Med Genet* 2004;**41**:425-32.
86. Lugtenberg D, Veltman JA, van Bokhoven H. High-resolution genomic microarrays for X-linked mental retardation. *Genet Med* 2007;**9**:560-5.
87. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;**437**:376-80.
88. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, et al. Paired-end mapping reveals extensive structural variation in the human genome. *Science* 2007;**318**:420-6.
89. Wilke K, Duman B, Horst J. Diagnosis of haploidy and triploidy based on measurement of gene copy number by real-time PCR. *Hum Mutat* 2000;**16**:431-6.
90. Armour JA, Sismani C, Patsalis PC, Cross G. Measurement of locus copy number by hybridisation with amplifiable probes. *Nucleic Acids Res* 2000;**28**:605-9.
91. White S, Kalf M, Liu Q, Villerius M, Engelsma D, Kriek M, Vollebregt E, Bakker B, van Ommen GJ, Breuning MH, et al. 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.
92. 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.
93. White SJ, Vink GR, Kriek M, Wuyts W, Schouten J, Bakker B, Breuning MH, den Dunnen JT. Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses. *Hum Mutat* 2004;**24**:86-92.
94. Bentley DR. Whole-genome re-sequencing. *Curr Opin Genet Dev* 2006;**16**:545-52.
95. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME. Microarray-based genomic selection for high-throughput resequencing. *Nat Methods* 2007;**4**:907-9.
96. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, et al. Direct selection of human genomic loci by microarray hybridization. *Nat Methods* 2007;**4**:903-5.
97. Olson M. Enrichment of super-sized resequencing targets from the human genome. *Nat Methods* 2007;**4**:891-2.
98. Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO, Straatman H, van der Vliet W, Huys EH, van Rijk A, et al. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am J Hum Genet* 2003;**73**:1261-70.
99. Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H, Firth H, Sanlaville D, Winter R, Colleaux L, et al. Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J Med Genet* 2004;**41**:241-8.
100. Menten B, Maas N, Thienpont B, Buysse K, Vandesompele J, Melotte C, de Ravel T, Van Vooren S, Balikova I, Backx L, et al. Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports. *J Med Genet* 2006;**43**:625-33.

101. Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM, Nordenskjold M. Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH). *J Med Genet* 2005;**42**:699-705.
102. Engels H, Brockschmidt A, Hoischen A, Landwehr C, Bosse K, Walldorf C, Toedt G, Radlwimmer B, Propping P, Lichter P, et al. DNA microarray analysis identifies candidate regions and genes in unexplained mental retardation. *Neurology* 2007;**68**:743-50.
103. Rosenberg C, Knijnenburg J, Bakker E, Vianna-Morgante AM, Sloos W, Otto PA, Kriek M, Hansson K, Krepischi-Santos AC, Fiegler H, et al. Array-CGH detection of micro rearrangements in mentally retarded individuals: clinical significance of imbalances present both in affected children and normal parents. *J Med Genet* 2006;**43**:180-6.
104. Stankiewicz P, Beaudet AL. Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. *Curr Opin Genet Dev* 2007;**17**:182-92.
105. Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, van der Vliet W, Huys EH, de Jong PJ, Hamel BC, et al. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* 2004;**36**:955-7.
106. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, et al. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 2005;**77**:78-88.
107. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. *Trends Genet* 2002;**18**:74-82.
108. Varela MC, Krepischi-Santos AC, Paz JA, Knijnenburg J, Szuhai K, Rosenberg C, Koiffmann CP. A 17q21.31 microdeletion encompassing the MAPT gene in a mentally impaired patient. *Cytogenet Genome Res* 2006;**114**:89-92.
109. Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R, Kooy RF, Reyniers E, Romano C, Fichera M, et al. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet* 2006;**38**:999-1001.
110. Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, Curley R, Cumming S, Dunn C, Kalaitzopoulos D, et al. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* 2006;**38**:1032-7.
111. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, et al. A common inversion under selection in Europeans. *Nat Genet* 2005;**37**:129-37.
112. Slavotinek AM. Novel microdeletion syndromes detected by chromosome microarrays. *Hum Genet* 2008;**124**:1-17.
113. Kirchhoff M, Bisgaard AM, Duno M, Hansen FJ, Schwartz M. A 17q21.31 microduplication, reciprocal to the newly described 17q21.31 microdeletion, in a girl with severe psychomotor developmental delay and dysmorphic craniofacial features. *Eur J Med Genet* 2007;**50**:256-63.
114. Lu X, Shaw CA, Patel A, Li J, Cooper ML, Wells WR, Sullivan CM, Sahoo T, Yatsenko SA, Bacino CA, et al. Clinical implementation of chromosomal microarray analysis: summary of 2513 postnatal cases. *PLoS ONE* 2007;**2**:e327.
115. Shaffer LG, Kashork CD, Saleki R, Rorem E, Sundin K, Ballif BC, Bejjani BA. Targeted genomic microarray analysis for identification of chromosome abnormalities in 1500 consecutive clinical cases. *J Pediatr* 2006;**149**:98-102.
116. A Proposed Standard System on Nomenclature of Human Mitotic Chromosomes. *Am J Hum Genet* 1960;**12**:384-8.
117. Wyandt HE. Introduction. In: Wyandt HE, Tonk VS, editors. Atlas of Human Chromosome Heteromorphisms.Dordrecht: Kluwer Academic Publishers; 2004. p. 3-10.
118. Donahue RP, Bias WB, Renwick JH, McKusick VA. Probable assignment of the Duffy blood group locus to chromosome 1 in man. *Proc Natl Acad Sci U S A* 1968;**61**:949-55.
119. Hansson K, Szuhai K, Knijnenburg J, van Haeringen A, de Pater J. Interstitial deletion of 6q without phenotypic effect. *Am J Med Genet A* 2007;**143A**:1354-7.
120. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet* 2001;**27**:234-6.
121. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001;**409**:928-33.
122. The International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;**437**:1299-320.
123. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;**449**:851-61.

124. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, et al. Large-scale copy number polymorphism in the human genome. *Science* 2004;**305**:525-8.
125. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. *Nat Genet* 2004;**36**:949-51.
126. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, et al. Fine-scale structural variation of the human genome. *Nat Genet* 2005;**37**:727-32.
127. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, et al. Common deletion polymorphisms in the human genome. *Nat Genet* 2006;**38**:86-92.
128. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 2006;**38**:75-81.
129. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. Global variation in copy number in the human genome. *Nature* 2006;**444**:444-54.
130. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, et al. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet* 2007;**80**:91-104.
131. Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, Tran CW, Scheffer A, Steinfeld I, Tsang P, Yamada NA, et al. The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* 2008;**82**:685-95.
132. de Smith AJ, Tsalenko A, Sampas N, Scheffer A, Yamada NA, Tsang P, Ben Dor A, Yakhini Z, Ellis RJ, Bruhn L, et al. Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. *Hum Mol Genet* 2007;**16**:2783-94.
133. Carter NP. As normal as normal can be? *Nat Genet* 2004;**36**:931-2.
134. Shaw CJ, Lupski JR. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet* 2004;**13 Spec No 1**:R57-R64.
135. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, et al. Copy number variation: new insights in genome diversity. *Genome Res* 2006;**16**:949-61.
136. Rodriguez-Revenga L, Mila M, Rosenberg C, Lamb A, Lee C. Structural variation in the human genome: the impact of copy number variants on clinical diagnosis. *Genet Med* 2007;**9**:600-6.