Diagnostic Genome Profiling in Mental Retardation

Bert B. A. de Vries,[1,*] Rolph Pfundt,[1,*] Martijn Leisink,[2] David A. Koolen,[1]
Lisenka E. L. M. Vissers,[1] Irene M. Janssen,[1] Simon van Reijmersdal,[1] Willy M. Nillesen,[1]
Erik H. L. P. G. Huys,[1] Nicole de Leeuw,[1] Dominique Smeets,[1] Erik A. Sistermans,[1]
Ton Feuth,[3] Conny M. A. van Ravenswaaij-Arts,[1] Ad Geurts van Kessel,[1]
Eric F. P. M. Schoenmakers,[1] Han G. Brunner,[1] and Joris A. Veltman[1]

Departments of [1]Human Genetics, [2]Biophysics, and [3]Epidemiology and Biostatistics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

Mental retardation (MR) occurs in 2%–3% of the general population. Conventional karyotyping has a resolution of 5–10 million bases and detects chromosomal alterations in ~5% of individuals with unexplained MR. The frequency of smaller submicroscopic chromosomal alterations in these patients is unknown. Novel molecular karyotyping methods, such as array-based comparative genomic hybridization (array CGH), can detect submicroscopic chromosome alterations at a resolution of 100 kb. In this study, 100 patients with unexplained MR were analyzed using array CGH for DNA copy-number changes by use of a novel tiling-resolution genomewide microarray containing 32,447 bacterial artificial clones. Alterations were validated by fluorescence in situ hybridization and/ or multiplex ligation-dependent probe amplification, and parents were tested to determine de novo occurrence. Reproducible DNA copy-number changes were present in 97% of patients. The majority of these alterations were inherited from phenotypically normal parents, which reflects normal large-scale copy-number variation. In 10% of the patients, de novo alterations considered to be clinically relevant were found: seven deletions and three duplications. These alterations varied in size from 540 kb to 12 Mb and were scattered throughout the genome. Our results indicate that the diagnostic yield of this approach in the general population of patients with MR is at least twice as high as that of standard GTG-banded karyotyping.

Introduction

Mental retardation (MR) occurs in 2%–3% of newborns in the general population, but, in most cases, its cause has remained elusive (Roeleveld et al. 1997; Yeargin-Allsopp et al. 1997). Establishing the cause in a mentally retarded individual improves clinical management and facilitates genetic counseling of the family. Chromosome abnormalities are detectable by microscopic analysis of chromosomes isolated from peripheral blood lymphocytes in ~5% of patients with unexplained MR (Phelan et al. 1996; de Vries et al. 1997; Schinzel 2001). The resolution of microscopic techniques for detection of abnormal karyotypes is typically in the range of 5–10 Mb. Molecular cytogenetic techniques, such as FISH and multiplex ligation-dependent probe amplification (MLPA) (Schouten et al. 2002), have shown that causative submicroscopic rearrangements of the subtelomeric

regions can be found in ~5% of patients with human malformations and MR (Flint et al. 1995; Knight et al. 1999; De Vries et al. 2003; Koolen et al. 2004a). These results for the subtelomeric regions indicate that submicroscopic rearrangements in general may be a more common cause of MR than microscopically visible rearrangements.

Recent technological developments allow the investigation of the human genome at a resolution that is ~50–100 times higher than that of routine chromosome analysis by karyotyping (Pinkel et al. 1998; Lucito et al. 2003; Ishkanian et al. 2004; Ried 2004). Such DNA-based methods are collectively referred to as "molecular karyotyping." One such method is array-based comparative genomic hybridization (array CGH) (Solinas-Toldo et al. 1997; Pinkel et al. 1998; Veltman et al. 2002). Pilot studies using intermediate resolution (1-Mb) array CGH have indicated the potential of this technology in diagnosing patients with MR and congenital anomalies (Vissers et al. 2003; Shaw-Smith et al. 2004). Other recent studies have reported the presence of large-scale copy-number variations (LCVs) in healthy individuals (Iafrate et al. 2004; Sebat et al. 2004), which could complicate the clinical interpretation of these genomic alterations.

The goal of the underlying study was to estimate the significance of this whole-genome approach for the gen-

Table 1

Clinical Characteristics of the 100 Patients with MR

Characteristic	Value[a]
Sex:	
Male	60
Female	40
Age group:	
<10 years old	67
10–20 years old	23
>20 years old	10
Age (years):	
Median	7
Range	1–63
Level of MR:	
Mild	32
Moderate	27
Severe	32
Not assessable	9
Family history of MR:	
Yes:	
Compatible with Mendelian inheritance	13
Incompatible with Mendelian inheritance	4
No	83
Prenatal-onset growth retardation:	
Yes	18
No	82
Postnatal growth abnormalities:	
Yes:	
Head circumference or stature	42
Head circumference and stature	14
No	44
Facial dysmorphic features:	
Yes:	
1 Feature	19
≥2 Features	73
No	8
Nonfacial dysmorphism and congenital abnormalities:	
Yes:	
1 Abnormality	35
≥2 Abnormalities	18
No	47
Total clinical score:	
0	0
1	5
2	16
3	32
4	21
5	14
6	7
7	5
8–10	0

[a] Data are no. of patients, unless indicated otherwise.

eral population of patients with MR. Therefore, a recently published, 32,447-clone BAC array, with complete coverage of the entire human genome (Ishkanian et al. 2004), was used for the first time to test 100 mentally retarded patients, with or without additional congenital malformations, for the presence of submicroscopic chromosomal alterations. Novel algorithms were developed and validated for microarray normali-

zation and copy-number detection, and parental samples were used as a control population to distinguish between causative de novo alterations and inherited LCVs. A clinical evaluation of the patient sample used in this study, in comparison with that of the overall patient population seen in our clinical genetics center, allowed us to estimate the importance of this approach in medical practice.

Patients and Methods

Patients

The diagnostic value of the genomewide tiling-resolution array CGH approach was tested by screening 100 patients with unexplained MR (60 males and 40 females; age range 10 mo to 63 years) (table 1). All had normal GTG-banded chromosomes, and all were previously subjected to a specific subtelomeric screen using MLPA, with normal results (Koolen et al. 2004*a*). All patients were evaluated for additional clinical features by a board-certified clinical geneticist. A clinical scoring system based on the presence of family history of MR, prenatal-onset growth retardation, postnatal growth retardation or advanced growth, facial dysmorphic features, and other congenital abnormalities was applied, with a score from 0 to 10 (table 2) (Baralle 2001; de Vries et al. 2001).

Genomic DNA was isolated from blood leukocytes by use of routine procedures. For this study, two pools of reference DNA were made—one containing equal amounts of genomic DNA from 10 healthy male blood donors and one containing equal amounts of genomic DNA from 10 healthy female blood donors. Blood was collected for DNA isolation from a total of 72 parents for further analysis. This study was approved by the Medical Ethics Committee of the Radboud University Nijmegen Medical Centre.

Microarray Preparation

We prepared a tiling-resolution microarray consisting of 32,447 overlapping BAC clones, which were selected to cover the entire human genome (Ishkanian et al. 2004; Krzywinski et al. 2004) and are available at the BACPAC Resources Center Web site, using methodology essentially as described elsewhere (Veltman et al. 2004). In brief, genomic target DNAs were isolated from 1-ml bacterial cultures by use of an AutogenPrep 960 (Autogen), in accordance with the instructions of the manufacturer. Degenerate oligonucleotide-primed PCR (DOP-PCR)

Table 2

Clinical Checklist

The table is available in its entirety in the online edition of *The American Journal of Human Genetics.*

was performed on 50 ng of DNA from all clones, as described elsewhere (Telenius et al. 1992), with minor modifications (Veltman et al. 2002). DOP-PCR products were dissolved at a concentration of 1 μg/μl in a 30% dimethyl-sulfoxide solution and were spotted onto CMT-ultragaps coated glass slides (Corning) by use of an Omnigrid 100 arrayer (Genomic Solutions). All 32,447 clones were printed once on a single microarray slide.

Isolation of DNA and Microarray-Based Genome Profiling

Isolation of genomic DNA, DNA labeling, hybridization of labeled DNA to the 32,447-BAC array, and spot identification were performed as described elsewhere (Vissers et al. 2003). In brief, genomic DNA from patients and controls was isolated in accordance with standard procedures and was purified using a QIAamp kit (QIAgen), in accordance with the instructions of the manufacturer. Next, 500 ng of genomic DNA from each patient was labeled by random priming with Cy3-dUTP and Cy5-dUTP (Amersham Biosciences) and was hybridized in duplicate with dye-swap against the sex-mismatched reference pool. Parental samples were hybridized once against the same reference pool as their child. Test and reference samples were mixed with 120 μg of human Cot-1 DNA (Roche), were coprecipitated, and were resuspended in 120 μl of a hybridization solution containing 50% formamide, 10% dextran sulfate, 2 × saline sodium citrate (SSC), 4% SDS, and 10 μg/μl of yeast tRNA (Invitrogen). Hybridization and posthybridization washing procedures were performed using a GeneTac Hybridization Station (Genomic Solutions). An 18-h hybridization at 37°C with active circulation of the probe was performed, followed by five posthybridization wash cycles in 50% formamide and 2 × SSC at 45°C and five wash cycles in phosphate buffer at 20°C. Slides were dried by centrifugation and were scanned using a GenePix Autoloader 4200AL laser scanner (Axon Instruments). Spot identification and two-color fluorescence intensity measurements were obtained using the GenePix 5.0 software, and all data were entered into a database for subsequent analysis.

Data Analysis

The $\log_2$-transformed test:reference ratio for each clone was normalized by subtracting its local mean $\log_2$ test:reference ratio obtained by a weighted median filter (Yin and Yang 1996). This filter averages the log ratios of autosomal clones weighted by a Gaussian distribution with its mean at the spot to be normalized and an SD of 1 physical mm on the array. The normalized ratios were analyzed for loss and gain regions by a standard hidden Markov model (HMM) (Rabiner 1989). This HMM has three hidden states (s_i) for each clone: normal,

loss, and gain. The probability of finding the HMM in a certain state, given normalized log ratios x_i, is given by

$$p(s|x) \propto p(x|s)p(s) = \prod_i p(x_i|s_i)\, p(s_i|s_{i-1}) , \qquad (1)$$

where the transition probability is given by

$$p(s_i|s_{i-1}) = \begin{Bmatrix} (1 + 2e^{-\Delta_i})/3 & \text{if } s_i = s_{i-1} \\ (1 - e^{-\Delta_i})/3 & \text{if } s_i \neq s_{i-1} \end{Bmatrix} ,$$

where Δ_i is the distance on the chromosome between the adjacent clones i and $i-1$. This equation has the following properties: (1) clones with identical chromosomal positions have identical states; (2) clones at an infinite distance from each other (e.g., located on different chromosomes) are noncorrelated; and (3) the transition probability for a distance $2\Delta_i$ is the square of that for a distance Δ_i. In this study, we used units of 150 Mb to measure the distances Δ_i. The observed log ratio x_i is assumed to be Gaussian distributed, with SD σ and a mean that depends on its hidden state s_i, as follows:

$$p(x_i|\text{normal}) = N(0,\sigma)$$

$$p(x_i|\text{loss}) = N(\mu,\sigma)$$

$$p(x_i|\text{gain}) = N(-\mu,\sigma) .$$

The sex-mismatched hybridizations allowed us to use the median of the log ratios of clones mapped to the chromosome X as an internal estimate of a single-copy gain or loss for each experiment and allowed a quick objective general assessment of the quality of individual experiments. For this reason, the sex chromosomes were excluded from further analysis.

The SD of the log ratios of all autosomal clones was used as an estimate for σ. For duplicate experiments, the means and SDs were computed per array, and the probability $p(x_i^1, x_i^2|s_i)$ is the product of the two individual Gaussian distributions. To find the hidden state that best matches the observations, we used a standard Viterbi algorithm that optimizes equation (1) (Forney 1973). This results in a path in which we find regions of consecutive clones that are in the hidden gain or loss state. To quantify the likelihood of a copy-number gain or loss in a certain region, indexed by n, we compared the optimal log likelihood of the Viterbi path, $\log p^*$, with the log-likelihood in the absence of the region, $\log p^n$. The "certainty" that a region contains a real copy-number change was defined as

$$\text{certainty} = \log_{10}(\log p^* - \log p^n) .$$

The experimental threshold for this certainty was set at 1.5 for the experiments done in duplicate. To exclude individual false-positive results, we furthermore required

that a region consist of at least three clone positions. This HMM method was validated by randomly permuting all autosomal data and rerunning the algorithm on the entire set. For single-slide experiments (parental samples), this analysis resulted in the detection of eight aberrations, whereas in the experiments in duplicate (patient samples), only two small aneusomic segments were identified. By randomly permuting all data, we created a database of microarray data for which no spatial correction is expected. Any aberration found can, therefore, be interpreted as a false positive. By doing so, a valid estimate of the expected number of false positives in each experiment is obtained. In addition, an experiment in duplicate was performed in which the male reference pool was hybridized against the female reference pool. Again, the HMM was run, and no significant regions with a copy-number gain or loss were found. From this, we conclude that the reference pools themselves did not contain pronounced gains or losses. From all these analyses, we concluded that the criteria are highly conservative and that, especially for the experiments in duplicate, almost all regions that meet the criteria are excellent candidates for being a true copy-number alteration.

Confirmation Experiments

De novo chromosomal alterations were validated in the individual patients by at least one other method: MLPA (Schouten et al. 2002), with specifically designed synthetic probe sets, and/or FISH. It was similarly confirmed that parental samples contained a normal copy number for these chromosomal regions. For MLPA, we developed one or four uniquely sized probes for each region to be tested, in accordance with a protocol provided by MRC-Holland (all reagents for MLPA reaction and subsequent PCR amplification were purchased from MRC-Holland). Ten probes were combined in one MLPA assay, in combination with four standard control probes. MLPA analysis was performed as described elsewhere, with minor modifications (Koolen et al. 2004a). In brief, 200–400 ng of DNA in a final volume of 8 μl was heated for 5 min at 98°C. Equal amounts of probe mix and SALSA hybridization buffer were added to each sample, followed by heat denaturation for 1 min at 97°C and overnight incubation at 60°C. Next, Ligation-65 mix was added and was incubated for 15 min at 54°C. After heat inactivation, the ligation products were amplified by PCR with the use of the 6-FAM–labeled primer set. Amplification products were identified and quantified by capillary electrophoresis on an ABI 3730 genetic analyzer, using GeneMapper software (Applied Biosystems). Data were normalized by dividing each probe's signal strength by the average signal strength of the sample. This normalized peak pattern was divided by the average peak pattern of all samples in the same experiment. Values for control wild-type peaks ($n = 1,667$)

were centered at 1.0, with an SD of 0.06. The mean ± 5 SD was used as the threshold for copy-number changes (1.3 for gains and 0.7 for losses).

In addition, each de novo DNA copy-number alteration identified by array CGH was validated by FISH analysis on fixed metaphase spreads from the respective patient and from his or her parents. In each region of significant DNA copy-number alteration, four clones—also present on the array—were selected as probes in the FISH validation assay. In brief, BAC DNAs were amplified by DOP-PCR and were labeled with biotin by use of random priming. Subsequent hybridizations were performed overnight at 37°C on an OmniSlide Thermal Cycler System (Thermo Hybaid; Thermo Life Sciences), in accordance with standard protocols. Detection of the biotin-labeled probes was achieved using either streptavidin-Cy3 (Amersham Biosciences) or avidin-FITC followed by goat-a-avidin-FITC (Vector Laboratories). The FISH slides were analyzed using a Leica DM RA fluorescence microscope.
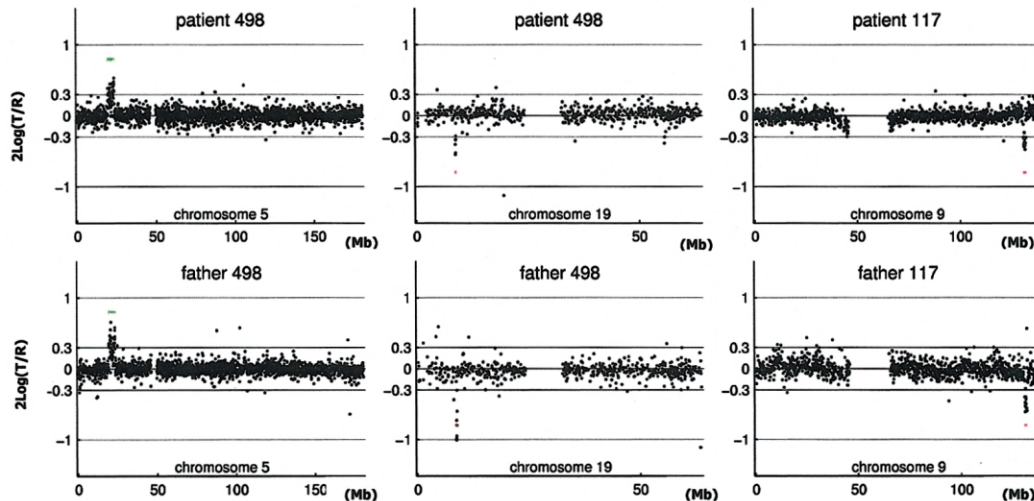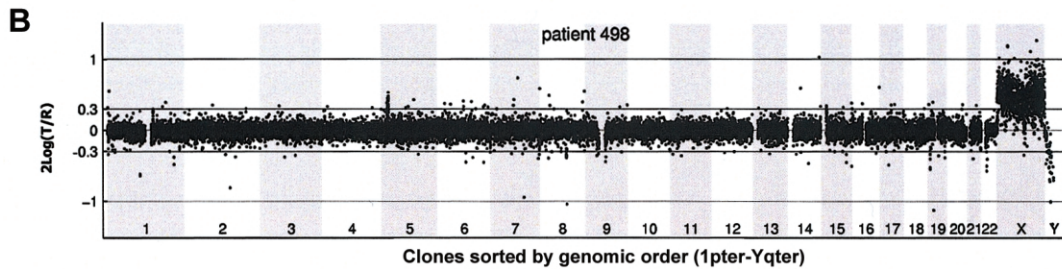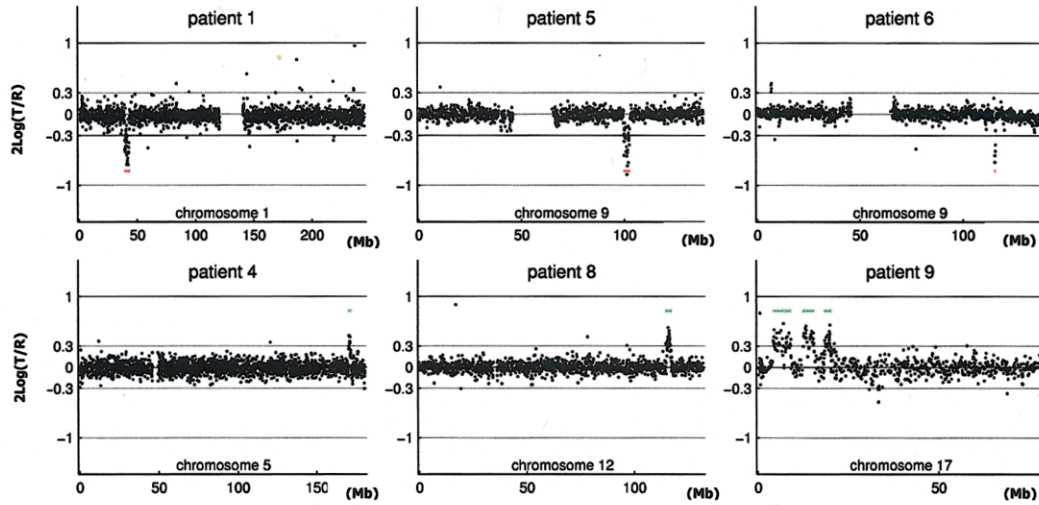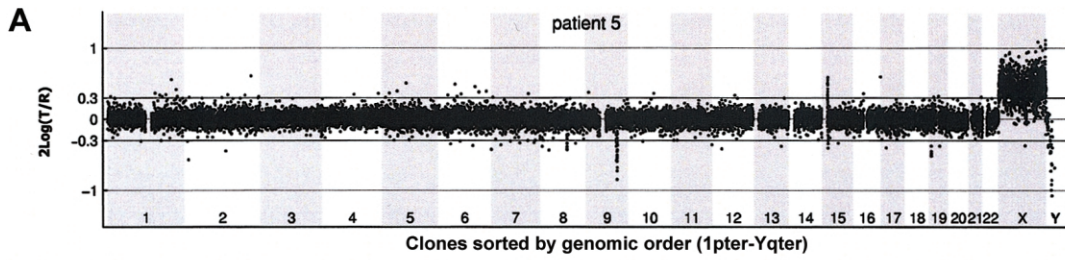
Results

Identification of Chromosomal Alterations by Genomewide Tiling-Resolution Array CGH

Our sample set consisted of 100 patients with unexplained MR, all with normal chromosomes, as determined by routine karyotyping, and with normal subtelomeres, as determined by MLPA. DNA from each patient was labeled and hybridized in duplicate (with label swap) against a sex-mismatched reference pool onto microarrays containing 32,447 human BAC clones, which completely covered the entire human genome. For each patient, fluorescence ratios of the scanned images were quantified and normalized. The two experiments for each patient were merged, clones were ordered according to their position on the human genome sequence, and copy-number states (normal, gain, or loss) were assigned to each clone by use of a validated HMM (see the "Patients and Methods" section). In figure 1, examples are shown of genome profiles obtained using the tiling-resolution BAC arrays. A total of 274 DNA copy-number alterations were identified in the 100 patients, with a range from 0 to 9 alterations per patient (fig. 2). Only three patients in this series did not show any copy-number alterations. Aneusomic segments varied in size from 50 kb to 12 Mb and were covered by 3–118 adjacently mapped clones present on the microarrays.

Distinguishing between Inherited and de Novo Alterations

To distinguish clinically relevant copy-number alterations from normal LCVs, 72 clinically normal parents of the 100 patients were tested using the same tiling-resolution microarrays. These parental samples served

A

patient 5

patient 1 — chromosome 1

patient 5 — chromosome 9

patient 6 — chromosome 9

patient 4 — chromosome 5

patient 8 — chromosome 12

patient 9 — chromosome 17

B

patient 498

patient 498 — chromosome 5

patient 498 — chromosome 19

patient 117 — chromosome 9

father 498 — chromosome 5

father 498 — chromosome 19

father 117 — chromosome 9

2Log(T/R)

Clones sorted by genomic order (1pter-Yqter)

as a control population and, in addition, provided valuable information on the inheritance of specific copy-number changes. Eleven alterations, detected in 10 individual patients (including two adjacent alterations in patient 9), were present in neither of the parental DNA samples nor in DNA from any of the 70 other unrelated parents tested (table 3). These de novo alterations most likely represent clinically relevant alterations. For five other alterations, de novo occurrence could not be established, as one or both of the patients' parents were not available; however, these alterations were never observed in any of the other parents tested (table 3). For these alterations, it remains unclear whether they are clinically relevant. All 258 remaining alterations detected in this study were shown to occur in at least one of the normal parents tested (table 4). Such inherited alterations most likely represent LCVs that are not related to the disorder under investigation. De novo aneusomic segments ranged in genomic size from 0.54 to 12.37 Mb (median 2.76 Mb), whereas LCVs ranged in size from 0.05 to 2.16 Mb (median 0.43 Mb) ($P = .015$, by two-sided independent student T test) (fig. 3). This suggests that clinically relevant aberrations are, in general, larger than normal genomic variants. The largest polymorphic variant detected in this study consisted of two adjacently located duplications on chromosome 5p14.3, involving a total of 3.55 Mb.

Confirmation Studies and Molecular Description of Clinically Relevant Aberrations

All de novo chromosomal alterations were validated in the individual patients by at least one other method: FISH and/or MLPA. Parental samples were similarly checked to confirm that they contained a normal copy-number for these chromosomal regions (table 2 and figs. A1–A10 in appendix A [online only]). In addition, the presence of a copy-number alteration was confirmed in four patient-parents trios with inherited copy-number variations, in addition to four alterations detected in patients for whom no parents were available for testing. In all cases, these MLPA and FISH studies confirmed the aberrations predicted by array CGH analysis.

DNA from 10 patients contained validated de novo segmental aneusomies: seven deletions and three duplications. The aneusomic segments identified varied in genomic size from 540 kb to 12 Mb, and they were scat-
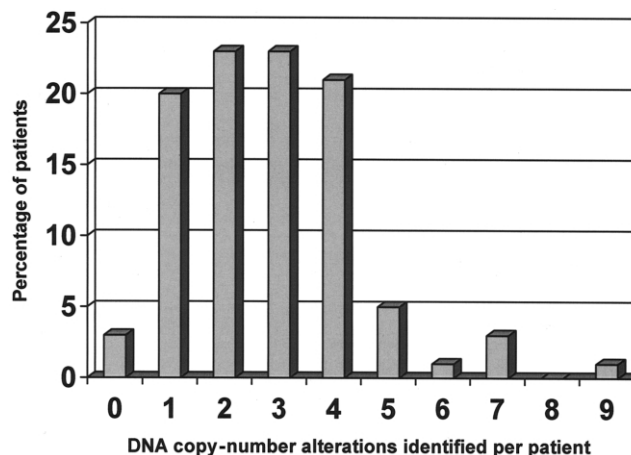
Figure 2 Frequency distribution of aneuploid segments per patient

tered throughout the genome. In all but one patient, the de novo alteration involved a single genomic segment of adjacent DNA clones. DNA from patient 9 contained a complex duplication on the short arm of chromosome 17, consisting of four interspersed regions covering a total of 8.93 Mb (table 3 and figs. 1 and A9 [online only]). For this patient, one of the regions involved encompasses a known microduplication of the *PMP22* gene associated with Charcot-Marie-tooth disease (CMT [MIM 118220]). Another known microdeletion region was observed in patient 10 at 22q11, the region associated with classical DiGeorge syndrome (MIM 188400) and velocardiofacial syndrome (VCFS [MIM 192430]). The other chromosomal alterations did not map to known microdeletion/duplication regions. Two of these alterations had a genomic size <1 Mb—one deletion of 920 kb on 2q23.1-q23.2 (in patient 2) and one deletion of 540 kb on 9q33.1 (in patient 6).

Clinical Implications of Tiling-Resolution Copy-Number Screening

Clinical details of the individual patients are presented in appendix A (online only) and are summarized in table 3. The study group included 32 patients with mild MR, 27 with moderate MR, and 32 with severe MR. In nine patients, the level of MR was not accurately assessable, mainly because of young age. The distribution of the

Figure 1 Genomic profiles obtained by tiling-resolution array CGH for patients with MR. The arrays contained 32,447 human BAC clones (indicated by small circles representing the log$_2$-transformed and normalized test:reference intensity ratios ["2Log(T/R)"]), ordered from 1pter to Yqter in the two genome profiles and from pter to qter for individual chromosomes, on the basis of physical-mapping positions obtained from the UCSC Genome Browser (May 2004 freeze). *A,* Examples of de novo alterations, including deletions and duplications, in patients 1, 4, 5, 6, 8, and 9. *B,* Inherited copy-number variations in patients 117 and 498, including the matching chromosome profiles of one of their parents. The individual chromosome plots show small horizontal lines to indicate the presence of a copy-number alteration (gain shown in green; loss shown in red) detected by HMM analysis. Single-clone changes were ignored.

Table 3

Patients with de Novo Alterations (Both Parents Tested) or Candidate Alterations (One or Both Parents Unavailable for Testing)

Type of Submicroscopic Aberration and Chromosome Band	Patient	Gain or Loss	Start (Mb)	End (Mb)	Length (Mb)	No. of Clones[a]	Confirmed[b]	Clinical Findings (Score)[c]
De novo:								
1p34.3-1p34.2	1	Loss	39.22	43.15	3.93	42	Yes	Severe MR, FD, GR, MC, and delayed brain myelinization (7)
2q23.1-2q23.2	2	Loss	149.17	150.09	.92	11	Yes	Severe MR, FD, GR, MC, BP, and epilepsy (3)
3q27.1-3q29	3	Loss	184.43	196.8	12.37	124	Yes	Severe MR, FD, GR, MC, hypogenitalism, hypoplastic kidneys, and deafness (6)
5q35.1	4	Gain	170.52	171.76	1.24	16	Yes	Mild MR, semilobular holoprosencephaly, finger-like thumbs, and polydactyly (3)
9q31.1	5	Loss	99.74	102.58	2.85	30	Yes	Moderate MR, FD, MC, and transposition of the great vessels (4)
9q33.1	6	Loss	115.3	115.84	.54	6	Yes	Mild MR, FD, macrocephaly, and autistic spectrum (3)
11q14.1-11q14.2	7	Loss	78.12	85.61	7.49	64	Yes	Mild MR and FD (5)
12q24.21-12q24.23	8	Gain	114.91	117.21	2.3	39	Yes	Severe MR, FD, GR, MC, and BP (7)
17p13.2-17p13.1	9	Gain	4.27	7.16	2.89	28	Not done	Moderate MR and FD (3)[d]
17p13.1	9	Gain	7.67	9.10	1.43	14	Yes	Moderate MR and FD (3)[d]
17p12	9	Gain	12.65	15.54	2.88	30	Yes	Moderate MR and FD (3)[d]
17p11.2	9	Gain	18.55	20.03	1.48	24	Yes	Moderate MR and FD (3)[d]
22q11.21	10	Loss	17.1	19.75	2.66	35	Yes	Mild MR, FD, and GR (3)
Candidate:								
1q21.1	11	Gain	143.25	145.38	2.12	21	Yes	Moderate MR, FD, and GR (5)
3p14.1	12	Loss	67.59	68.15	.56	5	Yes	Moderate MR (1)
7q11.21	13	Loss	64.23	64.58	.35	18	Not done	Mild MR, FD, and BP (3)
9p24.3	14	Gain	.21	.45	.23	5	Yes	Moderate, MR, and FD (2)
15q24.1-15q24.2	15	Loss	72.21	73.86	1.65	16	Yes	Mild MR, FD, GR, and MC (6)

NOTE.—None of the genomic regions were present in the Database of Genomic Variants (see Web Resources).

[a] Number of BAC clones located in the copy-number alteration.

[b] Confirmed by MLPA and/or FISH.

[c] For detailed clinical descriptions, see appendix A (online only). The clinical checklist score is given in parentheses. BP = behavioral problems; FD = facial dysmorphism; GR = growth retardation; MC = microcephaly.

[d] The same patient.

level of MR did not differ significantly between the group of patients with a de novo copy-number alteration ($n = 10$) and the patients with reliable cognitive assessment ($n = 81$) ($P = 1.00$ by Mann-Whitney U test). In addition to MR, all patients were evaluated for other clinical features by use of a clinical checklist (tables 1 and 2). The 10 patients with de novo alterations had a mean score of 4.4 ± 1.7, which was not significantly different from that of the remaining 90 patients without detectable de novo aneusomic segments (mean score 3.6 ± 1.4) ($P = .17$ by Mann-Whitney U test). This implies that, on average, the patients with de novo alterations did not have more clinical features suggestive of a chromosomal disorder than did patients without such an alteration. The distribution of the clinical scores of the 100 patients is shown in table 5.

Table 4

All Polymorphic Variants Found in the 100 Patients

The table is available in its entirety in the online edition of *The American Journal of Human Genetics.*

The 100 patients tested by array CGH in the present study were part of a series of 297 mentally retarded patients previously tested for subtelomeric abnormalities by use of MLPA. A subtelomeric aberration was found in 14 (4.7%) of 297 patients (table 5). To ascertain the representativeness of the current sample set for all patients with MR who were referred to our clinical genetics center for cytogenetic analysis, we gathered information on the 710 patients with unexplained MR (excluding Down syndrome) who were referred for karyotyping during a 2-year period (September 2002 to September 2004). Within this group, 34 patients (4.8%) had a microscopically visible chromosomal abnormality. A total of 120 unselected patients from the remaining 676 were clinically scored (table 5). Their mean score of 2.2 ± 1.7 was significantly less than that of the 100 patients tested in the current study ($P < .0001$ by Mann-Whitney U test). Thus, the current detection rate of 10% in our sample partly reflects clinical preselection based on phenotypes with additional dysmorphic features and/or growth abnormalities. The use of this genomewide tiling-resolution array CGH approach in an unselected

sample of patients with MR will likely have a lower yield of de novo microdeletions/duplications. To estimate the expected yield of submicroscopic alterations over the entire genome, including the subtelomeric regions, within an unselected population of mentally retarded patients, we calculated the frequencies of submicroscopic abnormalities for each clinical score among the group of 100 individuals tested by array CGH and the group of 297 tested by MLPA. These frequencies were then used to estimate the expected frequency of the submicroscopic aberrations for each clinical score among 120 unselected patients with MR. On the basis of bootstrapping with 1,500 replications, the overall estimated frequency of microdeletions/duplications in this unselected series of mentally retarded patients who were referred to our department for karyotyping is 7.3% (95% CI 4.2%–10.3%). This is considerably higher than the rate of microscopically visible deletions and duplications in this patient group (4.8%).

Discussion

In this study, we demonstrate that submicroscopic interstitial chromosomal abnormalities are a frequent cause of MR in patients with additional phenotypic features. In a set of 100 patients who have unexplained MR with or without congenital anomalies, 10 de novo aneusomic segments were detected by use of a tiling-resolution genomewide BAC array. For an additional five alterations detected, it is unknown whether the chromosome abnormality identified was de novo, because one or both parents were unavailable for testing. The 10 validated de novo copy-number alterations detected in this study by genomewide tiling-resolution array CGH were not detected by routine chromosome analysis or by quantitative analysis of the subtelomeric regions with the use of MLPA. This can be explained by the fact that all de novo alterations were interstitial and that 7 of the 10 abnormalities were <4 Mb (2 were <1 Mb). The tiling-resolution array CGH approach used in this study has a 10-fold increase in coverage over that of routinely used 1-Mb resolution BAC arrays with ~3,000 clones (Vissers et al. 2003; Shaw-Smith et al. 2004; Rosenberg et al. 2005). Therefore, it can be expected that the application of this approach leads to an increase in diagnostic yield for MR. Although, indeed, many more aberrations were detected by this approach (on average, three per patient), only two de novo aberrations <1 Mb were detected. This result might suggest that a reliable 1-Mb copy-number screening could identify the majority of submicroscopic anomalies related to MR. However, additional data are needed before the percentage of deletions and duplications <1 Mb in patients with MR can be reliably assessed. One added value of tiling-path resolution arrays, moreover, lies in the robustness of the detection. A 1-Mb duplication, for instance, will be detected by 10

clones present on the tiling-path resolution array, compared with only 1 clone on a 1-Mb resolution array. This results in a significantly lower false-positive rate. In fact, no false positives were detected in this study. Furthermore, a tiling-path array approach results in a more precise mapping of the start and endpoints of each rearrangement, thus allowing detailed genotype-phenotype correlation assessments.

Conventional GTG-banded karyotyping detects chromosomal alterations in ~5% of individuals with unexplained MR. Recent studies involving techniques such as telomeric FISH and MLPA have revealed that small subtelomeric deletions and duplications occur in ~5% of clinically preselected individuals with unexplained MR whose routine chromosome analysis was normal (Biesecker 2002; de Vries et al. 2003; Koolen et al. 2004a). Combined, these studies suggest that chromosomal alterations, both microscopic and submicroscopic and interstitial as well as subtelomeric, are responsible for 20% of cases of unexplained MR. However, we need to take into account that our series of 100 patients is not representative of all patients with unexplained MR who are referred to our clinical genetics department. Correction for this clinical preselection suggests an overall incidence of ~12% of chromosomal alterations in this patient group, of which ~7% are submicroscopic (interstitial and subtelomeric combined). We conclude that molecular karyotyping by tiling-resolution array CGH will increase the diagnostic yield of chromosomal abnormalities among mentally retarded individuals by approximately twofold. This is probably a conservative estimate, since aberrations on the X chromosome could not be analyzed for the presence of genomic rearrangements because of the sex-mismatched hybridizations needed for quality control (see the "Patients and Methods" section). At present, the robustness of the tiling-resolution array CGH procedure is such that these internal quality assessments are no longer required, thus
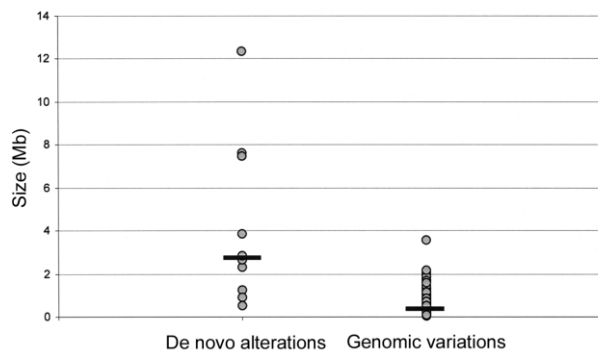
Figure 3 Size distribution of chromosomal aberrations, divided into de novo alterations (*n* = 10) and polymorphic variants (*n* = 258). Black horizontal bars indicate median size (de novo alteration median 2.76 Mb; genomic variant median 0.43 Mb).

Table 5

Estimated Frequencies of Submicroscopic Aberrations for Each Clinical Score, in the Study Cohorts and an Unselected Cohort of Patients with MR

	STUDY COHORTS							UNSELECTED COHORT (n = 120)	
	Array CGH Tested (n = 100)			MLPA Tested (n = 297)					Estimated
CLINICAL SCORE	No. of Patients	No. of Aberrations	Prevalence	No. of Patients	No. of Aberrations	Prevalence	Combined Prevalence	No. of Patients	Frequency of Aberrations
0	0	0	.00	1	0	.00	.00	20	.0
1	5	0	.00	36	1	.03	.03	25	.7
2	16	0	.00	58	1	.02	.02	33	.6
3	32	5	.16	86	5	.06	.21	17	3.5
4	21	1	.05	62	1	.02	.06	14	.9
5	14	1	.07	32	4	.13	.19	4	.8
6	7	1	.14	15	1	.07	.20	5	1.0
7	5	2	.40	6	0	.00	.40	1	.4
8	0	0	.00	1	1	1.00	1.00	1	1.0
9	0	0	.00	0	0	.00	.00	0	.0
10	0	0	.00	0	0	.00	.00	0	.0
Total	100	10	.10	297	14	.05	.14	120	8.8

NOTE.—The estimations are based on the combined prevalences of the submicroscopic anomalies in the study cohorts.

allowing sex-matched hybridizations in a routine diagnostic setting.

The clinical consequences of submicroscopic de novo copy-number alteration are determined by the kind of alteration (deletion or duplication) and by the number and function of genes affected by the dosage alteration. Three patients had a de novo duplication varying in size from 1.2 to 8.9 Mb, leading to trisomy of 7–50 genes. These patients were mildly to severely mentally retarded, and two had brain abnormalities, which varied from microcephaly to semilobar holoprosencephaly. The duplication on chromosome 17p11.2-p13.2 (patient 9) was complex, since four separate regions were involved, including the CMT critical region (Suter and Scherer 2003). At age 6 years, this moderately retarded girl had not yet developed signs of CMT. The seven de novo deletions varied in size from 540 kb to 12.4 Mb. The gene content of the affected chromosomal areas varied from monosomy of 1 gene in patient 6 to monosomy of >70 genes in patient 3. The microdeletions at 2q23.1-q23.2 (patient 2) and 22q11.21 (patient 10) have been described previously. The latter deletion is involved in the DiGeorge syndrome and VCFS, which is a well-known, common, and clinically recognizable syndrome (Lindsay 2001). The 4-year-old boy with this deletion had mild MR, short stature, and some suggestive facial features but lacked diagnostic VCFS features, such as cardiac anomalies, cleft palate, long/thin fingers, hypocalcemia, or thymic hypoplasia. This shows that array CGH also aids in diagnosing atypical clinical presentation of common microdeletion syndromes. The 920-kb microdeletion on 2q23.1-q23.2 (in patient 2)

partly overlaps a microdeletion previously reported by our group and determined using a 1-Mb resolution BAC array (Vissers et al. 2003; Koolen et al. 2004*b*). Both patients had severe MR, postnatal growth retardation, microcephaly, coarse facies, and epilepsy. The overlap in clinical features between the two patients with these microdeletions may point to a new syndrome. Three genes are located in the overlapping deletion region— *MBD5, EPC2,* and *KIF5C.* Mice lacking the gene for KIF5C, a neuronal kinesin enriched in motor neurons, show a smaller brain size and a loss of peripheral neurons (Kanai et al. 2000). MBD5 is a member of the methyl CpG-binding-domain protein family specifically expressed in the brain (Nagase et al. 2000; Roloff et al. 2003).

In addition to clinically relevant DNA copy-number alterations, we detected 258 LCVs in our series of 100 patients. All these variations were also observed in one or more normal parental samples. This finding confirms and extends recent reports of the frequency of LCVs in normal individuals (Iafrate et al. 2004; Sebat et al. 2004). Such variation represents a novel class of polymorphisms within the human genome whose exact frequency in different ethnic groups remains to be established. In general, normal genomic variants were smaller than clinically relevant copy-number alterations. It is essential to rule out such submicroscopic variation, by studying parental samples, before drawing any conclusions about whether an aneusomic segment is causative for the MR. In addition, it is important to check whether such an alteration has been described previously in normal individuals, since this will reduce the likelihood that

it is causative. Moreover, clinical comparison with previously described patients who have overlapping alterations will aid in the clinical interpretation. Systematic collection of information on the phenotypic effects of rare submicroscopic alterations is urgently needed, now that molecular karyotyping is becoming part of clinical patient evaluation (see ECARUCA–European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations Web site).

In conclusion, this study demonstrates, for the first time, the use of a tiling-resolution BAC array in the diagnosis of unexplained MR in patients. We were able to distinguish de novo alterations from inherited DNA copy-number alterations, and we confirmed all findings identified by the array CGH method. After correction for clinical preselection, our data show that the percentage of submicroscopic rearrangements among individuals with unexplained MR exceeds that of microscopically visible alterations. Offering molecular karyotyping by array CGH to patients who have unexplained MR, with or without congenital anomalies, will therefore significantly improve diagnostic yield, leading to accurate diagnosis and genetic counseling.

Note added in proof.—The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE3191.

Web Resources

The URLs for data presented herein are as follows:

BACPAC Resources Center, http://bacpac.chori.org/
Database of Genomic Variants, http://projects.tcag.ca/variation/
ECARUCA–European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations, http://www.ecaruca.net/
MRC-Holland, http://www.mlpa.com/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for CMT, DiGeorge syndrome, and VCFS)

References

Baralle D (2001) Chromosomal aberrations, subtelomeric defects, and mental retardation. Lancet 358:7–8

Biesecker LG (2002) The end of the beginning of chromosome ends. Am J Med Genet 107:263–266

de Vries BB, van den Ouweland AM, Mohkamsing S, Duivenvoorden HJ, Mol E, Gelsema K, van Rijn M, Halley DJ, Sandkuijl LA, Oostra BA, Tibben A, Niermeijer MF, for the Collaborative Fragile X Study Group (1997) Screening and diagnosis for the fragile X syndrome among the mentally retarded: an epidemiological and psychological survey. Am J Hum Genet 61:660–667

de Vries BB, White SM, Knight SJ, Regan R, Homfray T, Young ID, Super M, McKeown C, Splitt M, Quarrell OW, Trainer AH, Niermeijer MF, Malcolm S, Flint J, Hurst JA, Winter RM (2001) Clinical studies on submicroscopic subtelomeric rearrangements: a checklist. J Med Genet 38:145–150

de Vries BB, Winter R, Schinzel A, Ravenswaaij-Arts C (2003) Telomeres: a diagnosis at the end of the chromosomes. J Med Genet 40:385–398

Flint J, Wilkie AO, Buckle VJ, Winter RM, Holland AJ, McDermid HE (1995) The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. Nat Genet 9:132–140

Forney GD (1973) The Viterbi algorithm. Proc IEEE Inst Electr Electron Eng 61:268–278

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. Nat Genet 36:949–951

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

Kanai Y, Okada Y, Tanaka Y, Harada A, Terada S, Hirokawa N (2000) KIF5C, a novel neuronal kinesin enriched in motor neurons. J Neurosci 20:6374–6384

Knight SJ, Regan R, Nicod A, Horsley SW, Kearney L, Homfray T, Winter RM, Bolton P, Flint J (1999) Subtle chromosomal rearrangements in children with unexplained mental retardation. Lancet 354:1676–1681

Koolen DA, Nillesen WM, Versteeg MH, Merkx GF, Knoers NV, Kets M, Vermeer S, van Ravenswaaij CM, de Kovel CG, Brunner HG, Smeets D, de Vries BB, Sistermans EA (2004a) Screening for subtelomeric rearrangements in 210 patients with unexplained mental retardation using multiplex ligation dependent probe amplification (MLPA). J Med Genet 41:892–899

Koolen DA, Vissers LE, Nillesen W, Smeets D, van Ravenswaaij CM, Sistermans EA, Veltman JA, de Vries BB (2004b) A novel microdeletion, del(2)(q22.3q23.3), in a mentally retarded patient, detected by array-based comparative genomic hybridization. Clin Genet 65:429–432

Krzywinski M, Bosdet I, Smailus D, Chiu R, Mathewson C, Wye N, Barber S, Brown-John M, Chan S, Chand S, Cloutier A, Girn N, Lee D, Masson A, Mayo M, Olson T, Pandoh

P, Prabhu AL, Schoenmakers E, Tsai M, Albertson D, Lam W, Choy CO, Osoegawa K, Zhao S, de Jong PJ, Schein J, Jones S, Marra MA (2004) A set of BAC clones spanning the human genome. Nucleic Acids Res 32:3651–3660

Lindsay EA (2001) Chromosomal microdeletions: dissecting del22q11 syndrome. Nat Rev Genet 2:858–868

Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KC, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M (2003) Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res 13:2291–2305

Nagase T, Kikuno R, Ishikawa K, Hirosawa M, Ohara O (2000) Prediction of the coding sequences of unidentified human genes. XVII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res 7:143–150

Phelan MC, Crawford EC, Bealer DM (1996) Mental retardation in South Carolina. In: Saul RA, Phelan MC (eds) Proceedings of the Greenwood Genetic Center. Greenwood Genetic Center, Greenwood, SC, pp 45–60

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 20:207–211

Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE Inst Electr Electron Eng 77:257–286

Ried T (2004) Cytogenetics—in color and digitized. N Engl J Med 350:1597–1600

Roeleveld N, Zielhuis GA, Gabreels F (1997) The prevalence of mental retardation: a critical review of recent literature. Dev Med Child Neurol 39:125–132

Roloff TC, Ropers HH, Nuber UA (2003) Comparative study of methyl-CpG-binding domain proteins. BMC Genomics 4:1

Rosenberg C, Knijnenburg J, Bakker E, Vianna-Morgante A, Sloos WC, Otto PA, Kriek M, Hansson K, Krepisch-Santos AC, Fiegler H, Carter NP, Bijlsma EK, van Haeringen A, Szuhai K, Tanke HJ (2005) 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 (http://jmg.bmjjournals.com/cgi/rapidpdf/jmg.2005.032268v1) (electronically published June 24, 2005; accessed August 22, 2005)

Schinzel A (2001) Catalogue of unbalanced chromosome aberrations in man. 2nd ed. Walter de Gruyter, Berlin

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

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305:525–528

Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H, Firth H, Sanlaville D, Winter R, Colleaux L, Bobrow M, Carter NP (2004) 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 41:241–248

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer 20:399–407

Suter U, Scherer SS (2003) Disease mechanisms in inherited neuropathies. Nat Rev Neurosci 4:714–726

Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics 13:718–725

Veltman JA, Schoenmakers EF, Eussen BH, Janssen I, Merkx G, van Cleef B, van Ravenswaaij CM, Brunner HG, Smeets D, Geurts van Kessel A (2002) High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. Am J Hum Genet 70:1269–1276

Veltman JA, Yntema HG, Lugtenberg D, Arts H, Briault S, Huys EH, Osoegawa K, de Jong P, Brunner HG, Geurts van Kessel A, van Bokhoven H, Schoenmakers EF (2004) High resolution profiling of X chromosomal aberrations by array comparative genomic hybridisation. J Med Genet 41:425–432

Vissers LELM, de Vries BBA, Osoegawa K, Janssen IM, Feuth T, Choy CO, Straatman H, van der Vliet W, Huys EHLPG, van Rijk A, Smeets D, van Ravenswaaij-Arts CMA, Knoers NV, van de Burgt I, de Jong PJ, Brunner HG, Geurts van Kessel A, Schoenmakers EFPM, Veltman JA (2003) Array-based comparative genomic hybridization for the genome-wide detection of submicroscopic chromosomal abnormalities. Am J Hum Genet 73:1261–1270

Yeargin-Allsopp M, Murphy CC, Cordero JF, Decoufle P, Hollowell JG (1997) Reported biomedical causes and associated medical conditions for mental retardation among 10-year-old children, metropolitan Atlanta, 1985 to 1987. Dev Med Child Neurol 39:142–149

Yin L, Yang R (1996) Weighted median filters: a tutorial. IEEE Trans Circuits Syst 43:157–192