

ARTICLE

Recurrent copy number changes in mentally retarded children harbour genes involved in cellular localization and the glutamate receptor complex

Martin Poot^{*1}, Marc J Eleveld¹, Ruben van 't Slot¹, Hans Kristian Ploos van Amstel¹ and Ron Hochstenbach¹

To determine the phenotypic significance of copy number changes (CNCs) in the human genome, we performed genome-wide segmental aneuploidy profiling by BAC-based array-CGH of 278 unrelated patients with multiple congenital abnormalities and mental retardation (MCAMR) and in 48 unaffected family members. In 20 patients, we found *de novo* CNCs composed of multiple consecutive probes. Of the 125 probes making up these probably pathogenic CNCs, 14 were also found as single CNCs in other patients and 5 in healthy individuals. Thus, these CNCs are not by themselves pathogenic. Almost one out of five patients and almost one out of six healthy individuals in our study cohort carried a gain or a loss for any one of the recently discovered microdeletion/microduplication loci, whereas seven patients and one healthy individual showed losses or gains for at least two different loci. The pathogenic burden resulting from these CNCs may be limited as they were found with similar frequencies among patients and healthy individuals ($P=0.165$; Fischer's exact test), and several individuals showed CNCs at multiple loci. CNCs occurring specifically in our study cohort were enriched for components of the glutamate receptor family (*GRIA2*, *GRIA4*, *GRIK2* and *GRIK4*) and genes encoding proteins involved in guiding cell localization during development (*ATP1A2*, *GIRK3*, *GRIA2*, *KCNJ3*, *KCNJ10*, *KCNK17* and *KCNK5*). This indicates that disease cohort-specific compilations of CNCs may aid in identifying loci, genes and biological processes that contribute to the phenotype of patients.

European Journal of Human Genetics (2010) 18, 39–46; doi:10.1038/ejhg.2009.120; published online 22 July 2009

Keywords: segmental aneuploidy profiling; BAC-based array-CGH; MCAMR; CNCs

INTRODUCTION

With a prevalence of ~2%, mental retardation is the most common reason for referral to medical genetic services.¹ Unlike in patients with X-linked mental retardation,² the genetic cause of mental retardation can rarely be pinpointed to a genetic locus or to a gene.³ Although the very genetic heterogeneity of mental retardation renders genetic linkage and association studies difficult, chromosomal aberrations, occurring with an incidence of 1 per 120 newborns,⁴ frequently provide a clue regarding the genetic locus underlying the phenotype of the affected child. Techniques allowing the detection of submicroscopic segmental aneuploidy have enabled us to pinpoint novel microdeletion and microduplication syndromes, such as CHARGE (coloboma, heart anomalies, chonal atresia, retardation, genital and ear anomalies),⁵ Peters Plus,⁶ recurrent 17q12 rearrangements,⁷ del(17)(q21),^{8,9} and the 22q13.3 deletion,¹⁰ Pitt–Hopkins¹¹ and thrombocytopenia-absent-radii syndromes¹² (for reviews see Lee and Lupski³ and Slavotinek¹³). In addition, clinically inconsequential segmental deletions^{14–16} and segmental duplications in healthy individuals¹⁷ have been reported. Studies involving array-based comparative genome hybridization (array-CGH),^{18,19} representational oligonucleotide microarray analysis,²⁰ fosmid library insert mapping²¹ and SNP inheritance analyses^{22–24} indicate that copy number changes (CNCs) are frequently occurring polymorphisms in the genome of healthy individuals.^{25–27} The

architecture of our chromosome complement, as it emerged during evolution, may be a driving force behind this type of genome diversity.^{26,28–31} Efforts to systematically compile CNCs have led to the identification of well over 6000 of such CNCs in DNA samples derived from lymphoblastoid cell lines of individuals of the HapMap population¹⁸ or from peripheral blood-derived DNA samples.^{32–34} The very existence of this type of structural genomic variability among apparently healthy individuals complicates a straightforward identification of pathogenetic segmental aneuploidy in patients.

To investigate whether CNCs may contribute to the complex spectrum of disorders consisting of combinations of mental retardation, developmental delay and multiple congenital anomalies (MCAMR), we determined the genomic locations and population frequencies of the CNCs shared among 278 patients and 48 of their unaffected family members seen in our institution. With this compilation of CNCs, we first determined the relative contribution of the microdeletion/microduplication syndromes that were recently discovered by array-based genome-wide segmental aneuploidy screening projects. Using gene prioritization tools (eg, Gene Annotation Tool to Help Explain Relationships (GATHER), Endeavour, Prioritizer), we secondly determined whether loci and genes pointing towards specific biochemical pathways were enriched in MCAMR patients by analyzing the 481 CNCs found exclusively in our cohort.

¹Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

*Correspondence: Dr M Poot, Department of Biomedical Genetics, University Medical Centre, PO Box 85090, Mail stop: KC.04.084.2, Utrecht, 3508 AB, The Netherlands. Tel: +31 88 756 8204; Fax: +31 88 756 8479; E-mail: m.poot@umcutrecht.nl

Received 21 January 2009; revised 14 May 2009; accepted 12 June 2009; published online 22 July 2009

MATERIALS AND METHODS

Patient and reference populations

A total of 278 unrelated patients showing a combination of mental retardation and skeletal or facial dysmorphisms, according to the criteria described by De Vries *et al*³⁵, and 48 of their unaffected parents were included in analyses of segmental aneuploidy by array-CGH. All patients were karyotyped at >550 band level. Patients with an abnormal karyotype, a sub-telomeric anomaly or a known microdeletion or microduplication syndrome detected by fluorescence *in situ* hybridization (FISH) or Multiplex Ligation-mediated Probe Amplification were excluded. To investigate whether a CNC has occurred *de novo* in an affected child, the parents were also analyzed by array-CGH. Thus, 20 *de novo* CNCs composed of multiple consecutive probes, presumed to be pathogenic, were detected in 20 (7.2%) out of 278 patients.

For this study, we generated two reference pools by combining DNAs derived from peripheral blood samples of 50 unrelated and healthy male and of 50 unrelated and healthy female donors (age range: 22–63 years). The blood samples were obtained through the Department of Hematology of the UMC Utrecht. The amount of each contributing DNA sample was adjusted in such a manner that it comprised 2% of the total reference pool. In this way, CNCs occurring at a frequency larger than 2% in the general population are likely to be included in the reference pools.

Array-CGH

For array-CGH, 1 μ g of sonicated genomic DNA from each patient (Tester) and a 1 μ g sample of sonicated genomic DNA from the sex-mismatched reference pool were labelled with Cy3-dUTP and Cy5-dUTP (Amersham Biosciences, Little Chalfont, UK), respectively, using the BioPrime DNA Labeling System (Invitrogen, Carlsbad, CA, USA). These labelled DNAs were co-hybridized to the array-slides using a GeneTAC Hybstation (Genomic Solutions, Ann Arbor, MI, USA).⁵ The array has been created by spotting in triplicate DOP-PCR products of 3783 BAC DNA probes that cover the entire human genome with an average spacing of \sim 0.7 Mb.⁵ Only BACs with a unique chromosomal location, as ascertained by FISH, have been included.³⁶ After hybridization and washing, the slides were scanned and imaged on a ScanArray Express HT (Perkin Elmer, Wellesley, MA, USA) using ScanArray Express software (version 2.1). Data were analyzed with the GenePix Pro 5.0 software package (Axon Instruments, Union City, CA, USA), followed by LOWESS normalization using the Acuity 3.1 software package (Axon Instruments). After analysis, the averages of all triplicate spots were calculated. Subsequently, data were analyzed essentially according to Wong *et al*.³² Briefly, all spots with an aberrant shape, signal to noise ratios <5.0 or with an SD above 0.3 were excluded. Using the signals of all autosomal probes, the 95% confidence interval (95% CI) was calculated. All autosomal BACs outside the 95% CI with a *P*-value of <0.05 for being nonaberrant (two-sided Student's *t*-test) of all participants were tabulated in a single Excel worksheet. Using a confidence interval of 3.3 or even 4.0 SD gave smaller numbers of aberrant signals, but the same BACs were eventually identified (results not shown). Subsequently, the significantly aberrant signal populations of all participants thus obtained were compared with each other. As all precautions notwithstanding singly occurring aberrant signals could still represent false positives, these were excluded from further analyses. Thus, on an average, 1.96 aberrant signals per experiment were eliminated. This amounts to 0.052% of all BAC probes on the array. All aberrant signals occurring more than once were identified and their precise genomic locations were retrieved from the UCSC Genome Browser (Human Genome Build 36, March 2006). As we used a sex-mismatched reference sample, X and Y chromosome BACs are not informative for our analyses. Hence, our approach does not permit any inferences regarding CNCs of sex chromosomes. The array data analysis procedure is outlined in Figure 1.

Illumina Infinium HumanHap300 Genotyping BeadChip SNP array

For the Infinium HumanHap300 Genotyping BeadChip SNP array analysis, we used 750 ng of patient DNA and followed the protocol as described by the manufacturer (Illumina, San Diego, CA, USA).

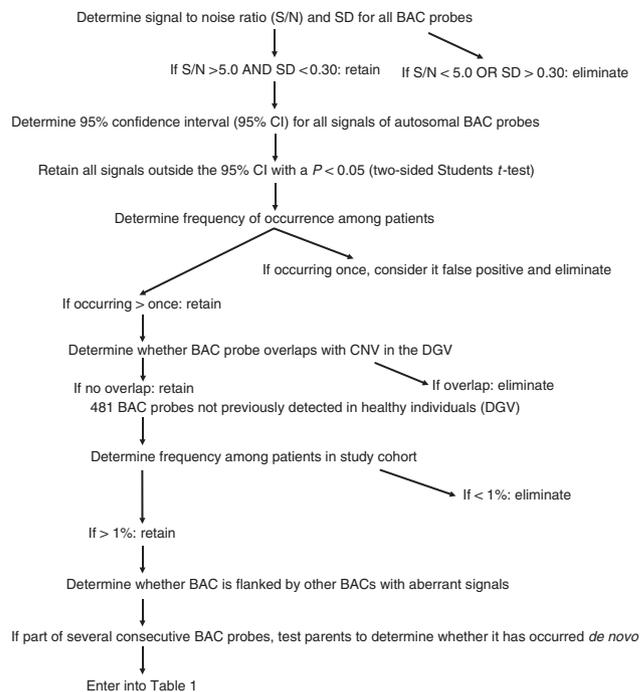


Figure 1 Flowchart of the data-triage procedure. For further explanation see text.

RESULTS

We performed genome-wide segmental aneuploidy profiling using a 3783 BAC-array with a functional genome coverage of 21.0% (for a definition of functional genome coverage see Coe *et al*³⁷) of 278 unrelated MCAMR patients and 48 of their unaffected family members (see Materials and methods). In 326 hybridizations, we found a total of 12 486 significantly aberrant signals, which amounted to an average of 38.3 (SD of 19.5; range: 14–96) and a median of 34.0 aberrant signals per participant. After excluding false positives (see Materials and methods), we retained aberrant signals representing a total of 1308 distinct loci in our study cohort. As we compared DNA samples of a single proband with a reference composed of a mixture of 50 DNA samples (see Materials and methods), these loci represent directly observed recurrent CNCs. The genomic locations of all CNCs and their frequency of occurrence as losses and gains in our study population are shown in online Table 1. As an example, the data for chromosome 1 are displayed in Figure 2.

As part of our study, we detected CNCs involving multiple consecutive BAC probes in 20 out of 278 patients, which have arisen *de novo* and were therefore presumed to be pathogenic (Table 1)³⁸. Several of these cases have been described in detail.^{39–44} In addition, we found nine inherited segmental aneuploidies. The 20 *de novo* aberrations encompassed 125 BAC probes, 14 of which were also identified as single CNCs shared among at least 2 other patients and 5 among the 48 healthy individuals of our study population. Of these 125 BAC probes, 65 have been detected among healthy individuals and are listed in the Database of Genomic Variants (DGV, <http://projects.tcag.ca/variation>). This emphasizes that *de novo* segmental aneuploidies and recurrent single-probe CNCs represent two intrinsically linked sets of genomic loci and that CNCs at these loci are not *per se* pathogenic, although they may contribute to the patient's phenotype.

Table 1 Array-CGH results of all patients with a *de novo* or an inherited CNC composed of multiple consecutive BAC probes

Case number	Array-CGH Results	Chromosome	Begin	End	Width	Inheritance
<i>De novo losses</i>						
1	46,XY arr cgh 1p32p31.3(RP11-265C4→RP11-89D5)x1	01	64 110 562	72 272 289	8 161 727	
2	46,XY arr cgh 1q44qter(RP11-88H4→CTB-167K11)x1	01	242 447 174	245 522 847	3 075 673	
3	46,XX arr cgh 3pterp26(RP5-1186B18→RP11-28P14)x1	03	1	7 009 709	7 009 708	
4	46,XX arr cgh 3qter(RP1-196F4→CTB-56H22)x1	03	198 377 055	198 491 701	114 646	
5	46,XY arr cgh 4pterp16(RP11-97H19→RP1-36p21)x1	04	1	6 813 350	6 813 349	
6	46,XY arr cgh 4q35qter(RP11-196K13→CTB-31J3)x1	04	185 011 645	190 554 259	5 542 614	
7	46,XX arr cgh 9q34(RP11-451O4→RP11-90B11)x1	09	136 429 910	137 468 380	1 038 470	
8	46,XY arr cgh 12q15(RP11-1024C4→RP11-89M22)x1	12	67 478 956	70 736 075	3 257 119	
9	46,XY arr cgh 15q25q26(RP11-116I7→RP11-57P9)x1	15	92 786 170	97 247 309	4 461 139	
10	46,XX arr cgh 15q26.2qter(RP11-120N1→RP1-154P1)x1	15	93 473 059	100 338 915	6 865 856	
11	46,XY arr cgh 16p12p13(CTA-991H5→RP11-109D4)x1	16	16 553 897	18 747 519	2 193 622	
12	46,XX arr cgh 16q24(RP11-110I8→RP11-140k16)x1	16	82 189 776	82 429 929	240 153	
<i>De novo gains</i>						
13	46,XX arr cgh 1p35p33(RP11-104J13→RP11-207P5)x3	01	31 000 823	37 176 819	6 175 996	
14	46,XY arr cgh 6q24(RP11-89a10→RP11-15h7)x3	06	139 955 509	141 971 532	2,016 023	
15	46,XX arr cgh 7pter(RP11-23D23)x3	07	1	954 584	954 583	
16	46,XY arr cgh 8p21(RP11-89M8→RP11-275K7)x3	08	22 779 034	28 923 242	6 144 208	
17	46,XY arr cgh 12pterp13.31(CTB-124K20→RP11-166G2)x3	12	1	5 783 146	5 783 145	
18	46,XX arr cgh 17p12(RP11-42F12→RP11-90G12)x3	17	14 045 669	15 390 352	1 344 683	
19	46,XY arr cgh 19q13.3(RP11-17I20→RP11-236B14)x3	13	53 172 856	55 797 960	2 625 104	
20	46,XY arr cgh 22q13.33(RP11-784H5→RP11-732O9)x3	22	47 825 728	48 357 762	532 034	
<i>Inherited losses</i>						
21	46,XX arr cgh 1q22(RP11-10P13→RP11-91G14)x1	01	157 624 308	158 600 366	976 058	Maternal
22	46,XX 2qter(CTB-172I13)x1	02	243 365 001	243 615 958	250 957	Maternal
23	46,XX arr cgh 3pter(RP5-1186b18→CTC-228k22)x1	03	1	955 897	955 896	Paternal
24	46,XX arr cgh 4pterp16.3(CTB-118B13→RP11-262P20)x1	04	1	752 620	752 619	Paternal
25	46,XY arr cgh 7p11(RP11-91C9→RP11-251I15)x1	07	52 135 582	57 922 250	5 786 668	Maternal
<i>Inherited gains</i>						
26	46,XY arr cgh 1q21.1;1q21.2(CTD-2122I24;RP11-30F17)x3	01	144 106 312	144 458 820	352 508	Paternal
27	46,XX arr cgh 22q12.1(RP11-13J4,RP11-29I20)x3	22	23 439 027	25 078 910	1 639 883	Paternal

Case number	Genes (OMIM numbers) linked to MCAMR phenotypes	Clinical features	Reference
<i>De novo losses</i>			
1	PDE4B (600127); NTRKR1 (602336)	Psychomotor retardation; agenesis of the corpus callosum	
2	HNRNPU (602869); ZNF124 (194631)	Agenesis of the corpus callosum; Dandy Walker; speech delay; facial dysmorphisms	Poot <i>et al</i> ⁴¹
3	CHL1 (607416); CNTN6 (607220); CNTN4 (607280); SUMF1 (272200); ITPR1 (147265); MRT2A (609262)	Epilepsy from age 1 onwards; vision impairments; obstipation	
4	DLG1 (601014)	Psychomotor retardation; facial dysmorphisms	
5	HD (143100); CPLX1 (605032); PDE6B (268000); IDUA (607016); FGFR1 (605830); WHSC1 (194190); WHSC2 (606026); FGFR3 (610474); EVC2 (607261); MSX1 (142983); MARLIN1 (611195); PPP2R2C (605997); WFS1 (606201)	Wolf Hirschhorn syndrome, diabetes insipidus, diabetes mellitus	Flipsen-ten Berg <i>et al</i> ⁴³
6	SLC25A4 (103220); CYP4V2 (608614)	Retrognathia; delayed motoric development; bilateral hearing impairment; frequent airway infections	
7	RXRA (180245)	Perimembranous VSD, secundum ASD, bicuspid aortic valve; dysplastic pulmonic valve with stenosis;	
8	FRS2 (607743); KCNMB4 (605223); FBXO16 (608516); TPH2 (607478)	Facial dysmorphisms; severe motoric developmental delay Intrauterine and postnatal growth delay; developmental delay	

Table 1 Continued

Case number	Genes (OMIM numbers) linked to MCAMR phenotypes	Clinical features	Reference
9	IGF1R (270450)	Developmental delay; growth retardation	
10	NR2F2 (107773); IGF1R (270450); TM2D3 (610014); ALDH1A (600463)	IUGR, proportional growth retardation, ASD type II, open oval window, valvular stenosis of the pulmonary artery, mild mental retardation; oculocutaneous albinism without misrouting	Poot <i>et al</i> ⁴²
11	SMG1 (607032)	Hydrocephalus; symptomatic epilepsy	
12	CDH13 (601364)	Microcephaly; cheilognathoschisis; low-set ears; upslant; hearing impairment; severe mental retardation	
<i>De novo gains</i>			
13	YARS (603623); DLGAP3 (611413); GRIK3 (138243); CCDC28B (610162); SYNC1 (611750); HPCA (142622); ADC (608353); GJB3 (603324); GJA (121012); COL8A2 (120252);	Facial dysmorphisms; delayed speech development	
14	no MCAMR related genes	Hypertelorism; frontal bossing; blepharophimosis; hypoplasia of the bile duct; urogenital tract abnormalities	
15	CENTA1 (608114); FAM20C (259775)	Generalized epilepsy of the Lennox Gastaut type	
16	NEF3 (162280); CHRNA2 (610353); DPYSL2 (602463); ADRA1C (104221); FZD3 (606143); RHOBTB2, LOXL2, ENTPD4; DOCK5, GNRH1; KCTD9, CDCA2	Postnatal growth retardation; severe speech delay; severe cognitive and motor delay; self mutilation; Temper tantrums; autism	Özgen <i>et al</i> ⁴⁴
17	CACNA1C (114205); B4GALNT3 (612220); WNK1 (605232); CACNA2D4 (608171); TULP3 (604730); SLC6A12 (603080); HSN2 (201300); WNT5B (606361); KCNA6 (176257); KCNA1 (176260); KCNA5 (176267); NTF3 (162660)	Developmental (in particular speech) delay; hyperluxation of joints	
18	COX10 (602125); HS3ST3B1 (604058); PMP22 (601097)	Tetralogy of Fallot; motoric development delayed; poor balance; anorectal malformation	
19	GRIN2D (602717); BCAT2 (113530); KCNA7 (176268); MYH14 (608568); CABP5 (607316); PLA2G4C (603602); KCNJ14 (603953); SULT2B1 (604125); SPACA4 (609932); CA11 (604644); FTL (134790); GYS1 (138570); LHB (152780); NTF5 (162662); SLC6A16 (607972); SLC17A7 (605208); CPT1C (608846); NUP62 (271930); KCNC3 (605259)	Cleft palate; scrotal hypoplasia; strabismus divergens; psychomotor delay; epilepsy; growth retardation	
20	No MCAMR related genes	Hypotelorism; dysplastic ears; trigonocephaly	
<i>Inherited losses</i>			
21	APCS (104770); PIGM (610293); KCNJ10 (602208); GIRK3 (600932); ATP1A2 (182340); ATP1A4 (607321); CASQ1 (114250); PXF (600279)	Developmental retardation, limited expressive speech; cavum septum pellucidum; facial dsymorphisms	
22	no MCAMR related genes	Postnatal growth retardation; developmental delay; eye abnormalities	
23	CHL1 (607416)	Intractable grand mal epilepsy; dilated ventricles and sulci, severe psychomotor retardation	
24	ZNF141 (194648); PDE6B (180072); MYL5 (160782)	Congenital nephrotic syndrome; congenital meiosis; developmental delay	
25	PSP (172480)	Microcephaly; epilepsy; pyramidal tetraparesis; bilateral neuron migration impairment	
<i>Inherited gains</i>			
26	PEX11B (603867); ITGA1 (604042)	Idiopathic generalized epilepsy; normal growth	Mefford <i>et al</i> ⁴⁶
27	SGSM1 (611417); ADRBK2 (109636); MYO18B (607295); CRYBB3 (123630); CRYBB2 (123620)	ASD; VSD; coarctation of the aorta; hypoplastic aortic arch; patent ductus arteriosus Facial and skeletal dysmorphic features	

Description of array-CGH according to the International System for Human Cytogenetic Nomenclature, version 2005 (ISCN 2005). The numbers in the columns, begin, end and width, demarcate the minimal size of the CNCs based on the affected BAC or SNP probes, according to the Human Genome Build 36.1 (March 2006). Coordinates based on SNP array data are shown in bold face. For further explanation see text

Out of the 3783 BACs on our array, 23 correspond to 11 of the microdeletion/microduplication syndromes that were recently discovered by genome-wide array-based screening efforts^{13,45,46} (Table 2). However, our array did not cover the 17q21.3 and the 16p11p12 microdeletion syndromes.^{8,9} Thus, 23 BAC probes revealed 84 gains and losses (0.67% of all aberrant signals) in hybridizations of 76 DNA samples from 69 patients (21.6% of all patients) and 7 unaffected parents (14.8% of all parents). These frequencies among patients and healthy individuals are not significantly different ($P=0.165$; Fischer's

exact test). Thus, close to one out of five individuals of our study cohort carried gains or losses at recurrent microdeletion/microduplication loci (Table 2). Hybridizations of seven patients and one healthy individual showed a loss or gain for two different microdeletion/microduplication loci. For five loci, at 1q21, 9q22.3, 15q13.3, 16p13 and 20q13.13q13.2, frequencies of gains and losses in our study cohort exceeded 1% (Table 2). These data suggest that some of the recurrent microdeletion/microduplication loci are relative hotspots for segmental aneuploidy among patients with MCAMR and their

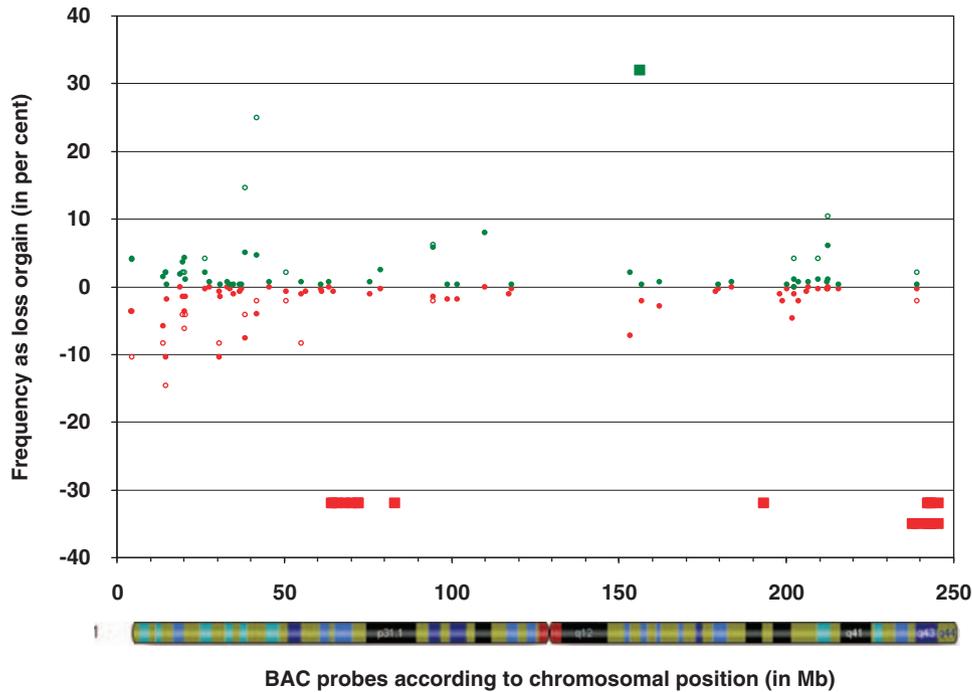


Figure 2 Recurrent CNCs specifically found in our MCAMR cohort. All CNCs mentioned in the DGV database have been omitted. Each dot represents a BAC probe according to its nucleotide position, following the March 2006 human reference sequence (NCBI Build 36.1). The ordinate position of each symbol represents the frequency of each CNC among MCAMR patients (closed symbol) and healthy individuals (open symbols). For each *de novo* loss and *de novo* gain found in our patients, red bars in the bottom decade and green bars in the top decade, respectively, have been added.

Table 2 Frequencies of CNCs corresponding to the loci of the recently described microdeletion/microduplication syndromes

Locus	Number of patients	Number of healthy individuals	Total	% of total	Number of losses	Number of gains
1q21	3	1	4	1.23	1	3
1q41q42	—	—	0	0	—	—
2p15p16	—	—	n.a.	n.a.	—	—
3q29	—	—	n.a.	n.a.	—	—
8q21.3q22.1	—	—	0	0	—	—
9q22.3	17	3	20	6.13	9	15
15q11	1	0	1	0.31	1	—
15q13.3	2	2	4	1.23	4	—
15q24	—	—	n.a.	n.a.	—	—
16p13	8	0	8	2.45	8	—
16p11p12	—	—	n.a.	n.a.	—	—
17q11q12	2	0	2	0.61	2	—
17q12	—	—	0	0	—	—
17q21.3	—	—	n.a.	n.a.	—	—
20q13.13q13.2	36	1	37	11.35	23	18
Sum	69	7	76	—	48	36

n.a., no corresponding BAC probe are available on our array

parents. As CNCs at these recurrent microdeletion/microduplication loci occurred in healthy individuals, the pathogenic burden resulting from these CNCs may be limited, as has already been proposed by Ledbetter.⁴⁷

Subsequently, we eliminated all CNCs that have previously been found as copy number variants in cohorts of healthy individuals (DGV; <http://projects.tcag.ca/variation>) and in those found among patients and healthy individuals in a study of amyotrophic lateral sclerosis.⁴⁸ Thus, we retained 481 CNCs that were exclusively found among the 278 patients in our study cohort. Of these, 203 CNCs were shared at a >1% frequency among the patients. Subsequently, we determined whether the frequencies of losses and gains among these 203 retained CNCs were distributed according to a random generation and elimination model or were subjected to selective pressures. Out of these 203 CNCs, 30 (14.8%) occurred only as a loss and 31 (15.3%) only as a gain. This distribution of the 203 CNCs shared at >1% among the participants in our study cohort is not consistent with a random generation and elimination model, and thus indicates that these CNCs may have been subjected to selective pressures.

In an analogy to Wong *et al*,³² we assumed that CNCs are identical to the BAC probe representing them. This has been confirmed by SNP array mapping of break points in a small sample from our patient cohort (Table 1). On the basis of this assumption, we analyzed the gene content of the CNCs found in our study population. The 278 patients in our cohort shared 481 CNCs; 78 of these were found as a loss only and 89 as a gain only. No genes were covered by 26 of those CNCs occurring as a loss only and by 35 of those CNCs occurring as a gain only. All other CNCs were associated with annotated genes. For instance, RP11-96F24, which was not previously reported as a CNC, showed a gain in 22 patients, but not in any healthy individual of our study cohort, and no hemizygous loss. This BAC probe covers the glutathione S-transferase M3 (*GSTM3*) gene. Although an increased dosage for this gene is apparently well tolerated, haploinsufficiency may not be compatible with life.

To systematically determine whether the CNCs that were specifically found among the members of our MCAMR patient cohort covered genes that encoded shared cellular components, biological processes or molecular functions, we first determined to which Gene Ontologies (<http://www.geneontology.org/GO.doc.shtml>) these belong. To do so, we analyzed our data using GATHER developed by Chang and Nevins.⁴⁹ Regardless of the frequencies of CNCs in our study cohort, each gene name was entered once. This allowed circumventing potential biases due to multiple listings of genes. We found that genes involved in potassium ion transport (GO:0006813) were specifically enriched for in the CNCs shared among members of our patient cohort with a $P < 0.0001$ (χ^2 -test) and a Bayes factor of 9 (likelihood of association vs nonassociation). These genes were *ABCC9*, *ATP1A2*, *ATP1A4*, *ATP1B2*, *GIRK3*, *GRIA2*, *GRIA4*, *GRIK2*, *GRIK4*, *KCNJ10*, *KCNJ3*, *KCNJ8*, *KCNK16*, *KCNK17*, *KCNK5*, *KCNK6*, *KCNN2*, *KCNN4*, *KCNV2*, *KCTD13*, *KCTD17*, *KCTD3* and *SLC24A3*. Of these 23 genes associated with potassium transport, seven (*ATP1A2*, *GIRK3*, *GRIA2*, *KCNJ10*, *KCNJ3*, *KCNK17* and *KCNK5*) were also associated with the establishment of localization (GO:0051234). Thus, genes involved in guiding cell localization during development by controlling potassium transport seem to be selectively enriched among CNCs shared among MCAMR patients, and are not found in healthy individuals. Analysis of our CNC set using the Endeavour gene prioritization tool developed by Aerts *et al*⁵⁰ and the Prioritizer tool developed by Franke *et al*,⁵¹ using different combinations of databases, corroborated this conclusion, although

P-values remained slightly above a significance level of 0.05 (results not shown).

It should be noted that the BAC-array used in this study affords a relatively low-resolution mapping of the break points of CNCs. Thus, it is conceivable that assuming that the break points of CNCs coincide with the borders of BAC clones (in analogy to Wong *et al*³²) may lead to a spurious inclusion of genes not involved in the observed CNC. This would lead to 'noise' in the gene prioritization analyses in such a way that no significance may be reached. In our analyses, and in those reported by Cuscó *et al*,⁵² still significant enrichment of certain Gene Ontologies has been found. Yet, in future studies, higher resolution platforms are recommended.

To determine whether some of these genes are part of one or several macromolecular networks, we used the Search Tool for Recurring Instances of Neighbouring Genes (STRING) developed by Snel *et al*,⁵³ and found that genes from the glutamate receptor family (*GRIA2*, *GRIA4*, *GRIK2* and *GRIK4*) form, on the basis of protein-protein interactions, a network with *NSE*, *GRIP1*, *PRKCABP*, *GRIN1*, *GSR*, *RIPK1*, *PRKCG* and *GLUR6*. In addition, the ATP-sensitive inward rectifier potassium channel genes *KCNJ10*, *KCNJ3*, *KCNJ8* and *GIRK3* may form heteromeric protein complexes with each other and with *PSD95*, *KCNJ16*, *KCNJ5* and *KCNJ6*. These gene prioritization analyses indicate that the MCAMR-associated CNCs identified in our study may to some extent represent a distinct population of dosage-sensitive genes, and point towards shared biological processes in children with MCAMR.

DISCUSSION

In this study, we addressed three questions. First, we determined whether the loci of *de novo* segmental aneuploidy in MCAMR patients are unique or overlap with the recurrent CNCs found among healthy relatives as well. Second, we tried to assess the pathogenic burden resulting from the microdeletion/microduplication syndromes recently discovered by genome-wide array-based segmental aneuploidy profiling efforts in our cohort of MCAMR patients. Third, we investigated whether the recurrent patient-specific CNCs in our study cohort harbour genes that may be involved in common biochemical pathways relevant to the phenotypes of MCAMR.

To do so, we compiled recurrent CNCs in a cohort of 278 unrelated MCAMR patients and 48 of their unaffected family members with a 3783 BAC-array. We used rigorous criteria to identify BAC probes showing aberrant signals without resorting to preset assumptions of a putative rate of aberration in each individual participant. In addition, similar to Wong *et al*,³² we took several precautions to exclude false-positive signals. These attempts to obtain a 'clean' compilation of CNCs in our study cohort may have led to some false negatives. The patient cohort for our study was chosen such as to exclude aberrant karyotypes and subtelomeric anomalies or identifiable 'classical' microdeletion or microduplication syndromes (eg, the Smith-Magenis syndrome, the 22q11 syndrome). No additional selection criteria were applied. Therefore, our study cohort constitutes a reasonable representation of patients with multiple congenital anomalies and mental or motor retardation for which no diagnosis could be reached with standard diagnostic methods. To minimize potential population-based differences in haplotype compositions, we used a mixture of DNA samples from 50 healthy male or female volunteers. Finally, to exclude a spurious generation of CNCs by Epstein-Barr virus transformation,^{33,54} we used DNA samples directly derived from freshly drawn peripheral blood samples only. In view of the precautions detailed above, the 1308 loci found are close to representing the true CNC content of our study cohort, as can be ascertained at the

level of the functional genome coverage afforded by the array used (21%).³⁷ As has been pointed out,²⁵ a higher probe resolution and genome coverage may detect more CNCs, but the relevance of these smaller CNCs to the pathology in patients with multiple congenital anomalies and mental retardation remains unknown.

First, we determined whether the loci of *de novo* segmental aneuploidy in MCAMR patients are unique or overlap with the recurrent CNCs found among healthy relatives as well. Of the 125 BAC probes representing 20 *de novo* segmental aneuploidies among our 278 patients, 65 have been detected among healthy individuals in other studies (see DGV; <http://projects.tcag.ca/variation>) and 14 have also been identified as single CNCs shared among at least two other patients and five among the 48 healthy relatives of our study population. This emphasizes that *de novo* segmental aneuploidies and recurrent CNCs represent two intrinsically linked sets of genomic loci. In addition, these findings indicate that a mere CNC at any of these loci is neither *per se* pathogenic nor phenotypically neutral. However, hemizyosity for a certain gene, as provoked by a copy number loss, may 'unmask' an autosomal-recessive mutation in the retained allele.^{12,43} CNCs showing a genomic gain, on the other hand, may contribute to the patient's phenotype if the genes contained therein encode proteins that form heteromeric complexes with other protein(s). Then, the balance of the participating protein components may be perturbed such that nonfunctional complexes with pathogenic consequences may be formed. This molecular mechanism may account for the apparent gene dosage sensitivity of certain genes, which are covered by these CNCs detected as both gains and losses in our study (eg, *KCNJ10*, *KCNJ3* and *GIRK3*). Enrichment of this set of CNCs for these genes encoding components of the heteromeric ATP-sensitive inward rectifier potassium channels may represent such a case in point.

In addition, CNCs may affect the expression of genes in diploid chromosomal regions flanking the CNC.^{55,56} Hence, CNCs may affect the phenotype by complex mechanisms in addition to and possibly different from altered gene dosage. Therefore, CNCs shared among unrelated patients with related phenotypes remain pathogenically relevant and merit further scrutiny. Thus, further investigations are needed to determine a potential contribution of a given CNC to the phenotype of an MCAMR patient.

Second, we tried to assess the relative pathogenic burden resulting from the recently reported novel microdeletion/microduplication syndromes.^{13,45,46} Only one out of the seven inherited and none of the *de novo* CNCs represented an example of these novel microdeletion/microduplication syndromes (case number 26 in Table 1), whereas all the other CNCs seem to be novel and unique. The latter has been a common finding in previous studies of cohorts of MCAMR patients (reviewed in Hochstenbach *et al*, 2009).³⁷ Close to one out of five individuals in our study cohort carried gains or losses at recurrent microdeletion/microduplication loci. This indicates that the novel microdeletion/microduplication syndromes are frequently found among CNCs shared by our MCAMR patients and healthy individuals in our study cohort. As 76 out of 12 486 (0.61%) CNCs covered any of these 11 loci, the pathogenic burden of these CNCs may be relatively modest.⁴⁷ As seven patients and one healthy individual showed a loss or gain for at least two different loci, these are not likely to be the sole cause of the patient's phenotype. In addition, patients with CNCs at some of these loci show rather divergent phenotypes.^{46,58} Therefore, we infer that these loci may represent instances of 'genomic unrest', provoking a limited contribution to the aetiology of MCAMR and to the pathogenic burden in the population at large.

Third, our stepwise gene prioritization analyses indicate that the MCAMR-associated CNCs identified in our study point towards shared biological processes in children with MCAMR, including the glutamate receptor family (*GRIA2*, *GRIA4*, *GRIK2* and *GRIK4*), which, by protein-protein interactions, forms a network with *NSF*, *GRIPI*, *PRKCABP*, *GRIN1*, *GSR*, *RIPK1*, *PRKCG* and *GLUR6*. In addition, the ATP-sensitive inward rectifier potassium channel genes, *KCNJ10*, *KCNJ3*, *KCNJ8* and *GIRK3*, may form heteromeric protein complexes with each other and with *PSD95*, *KCNJ16*, *KCNJ5* and *KCNJ6*. This suggests that the phenotypically diverse cohort of MCAMR patients may share a limited set of biochemical aetiologies.

In summary, our study of 278 patients with multiple congenital malformations and mental or motor retardation and 48 of their healthy relatives has unveiled widespread CNC. The CNCs occurring at a higher than 1% frequency in our study cohort seemed to occur as deletions only and as duplications only at less than expected rates. This indicates that the more frequently occurring CNCs have been subjected to selective pressure, suggesting that some of them may affect the phenotype. Therefore, these CNCs are not to be discarded, even if their contribution to the phenotype is not immediately obvious. Our finding that the set of CNCs shared among patients with MCAMR is enriched for genes encoding components of the ATP-sensitive inward rectifier potassium channel genes involved in guiding cell localization during development and members of the glutamate receptor family suggests that these biological processes may contribute to the phenotypes of patients with MCAMR. This hypothesis merits testing in replicating studies of other cohorts of patients and by using other biological pathway analysis algorithms. Eventually, patient-cohort-specific compilations of CNCs may serve as direct markers for genes that may help to identify loci, genes and biological processes contributing to the phenotype of patients.

ACKNOWLEDGEMENTS

We kindly acknowledge the help of Hylke M Blauw, Jan H Veldink and Leonard H van den Berg (Department of Neurology, University Medical Centre Utrecht) with the processing of the data set.

- 1 Newman WG, Hamilton S, Ayres J *et al*: Array comparative genomic hybridization for diagnosis of developmental delay: an exploratory cost-consequences analysis. *Clin Genet* 2007; **71**: 254-259.
- 2 Ropers HH: X-linked mental retardation: many genes for a complex disorder. *Curr Opin Genet Dev* 2006; **16**: 260-269.
- 3 Lee JA, Lupski JR: Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* 2006; **52**: 103-121.
- 4 Nielsen J, Wohler M: Chromosome abnormalities found among 34 910 newborn children: results from a 13-year incidence study in Aarhus, Denmark. *Hum Genet* 1991; **87**: 81-83.
- 5 Vissers LE, de Vries BB, Osoegawa K *et al*: Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am J Hum Genet* 2003; **73**: 1261-1270.
- 6 Lesnik Oberstein SA, Kriek M, White SJ *et al*: Peters Plus syndrome is caused by mutations in B3GALT1, a putative glycosyltransferase. *Am J Hum Genet* 2006; **79**: 562-566.
- 7 Mefford HC, Clauin S, Sharp AJ *et al*: Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am J Hum Genet* 2007; **81**: 1057-1069.
- 8 Koelen DA, Vissers LE, Pfundt R *et al*: A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet* 2006; **38**: 999-1001.
- 9 Sharp AJ, Hansen S, Selzer RR *et al*: Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* 2006; **38**: 1038-1042.
- 10 Cusmano-Ozog K, Manning MA, Hoyme HE: 22q13.3 deletion syndrome: a recognizable malformation syndrome associated with marked speech and language delay. *Am J Med Genet C Semin Med Genet* 2007; **145**: 393-398.

- 11 Zweier C, Peippo MM, Hoyer J *et al*: Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am J Hum Genet* 2007; **80**: 994–1001.
- 12 Klopocki E, Schulze H, Strauss G *et al*: Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. *Am J Hum Genet* 2007; **80**: 232–240.
- 13 Slavotinek AM: Novel microdeletion syndromes detected by chromosome microarrays. *Hum Genet* 2008; **124**: 1–17.
- 14 Hand JL, Michels W, Marinello MJ *et al*: Inherited interstitial deletion of chromosomes 5p and 16q without apparent phenotypic effect: further confirmation. *Prenat Diagn* 2000; **20**: 144–148.
- 15 Sumption ND, Barber JC: A transmitted deletion of 2q13 to 2q14.1 causes no phenotypic abnormalities. *J Med Genet* 2001; **38**: 125–127.
- 16 Barber JC: Directly transmitted unbalanced chromosome abnormalities and euchromatic variants. *J Med Genet* 2005; **42**: 609–629.
- 17 Horvath JE, Bailey JA, Locke DP *et al*: Lessons from the human genome: transitions between euchromatin and heterochromatin. *Hum Mol Genet* 2001; **10**: 2215–2223.
- 18 Iafrate AJ, Feuk L, Rivera MN *et al*: Detection of large-scale variation in the human genome. *Nat Genet* 2004; **36**: 949–951.
- 19 Sharp AJ, Locke DP, McGrath SD *et al*: Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 2005; **77**: 78–88.
- 20 Sebat J, Lakshmi B, Troge J *et al*: Large-scale copy number polymorphism in the human genome. *Science* 2004; **305**: 525–528.
- 21 Tuzun E, Sharp AJ, Bailey JA *et al*: Fine-scale structural variation of the human genome. *Nat Genet* 2005; **37**: 727–732.
- 22 Conrad DF, Andrews TD, Carter NP *et al*: A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 2006; **38**: 75–81.
- 23 Hinds DA, Kloek AP, Jen M *et al*: Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet* 2006; **38**: 82–85.
- 24 McCarroll SA, Hadnott TN, Perry GH *et al*: Common deletion polymorphisms in the human genome. *Nat Genet* 2006; **38**: 86–92.
- 25 Eichler EE: Widening the spectrum of human genetic variation. *Nat Genet* 2006; **38**: 9–11.
- 26 Freeman JL, Perry GH, Feuk L *et al*: Copy number variation: new insights in genome diversity. *Genome Res* 2006; **16**: 949–961.
- 27 McCarroll SA, Altshuler DM: Copy-number variation and association studies of human disease. *Nat Genet* 2007; **39**: S37–S42.
- 28 Bailey JA, Gu Z, Clark RA *et al*: Recent segmental duplications in the human genome. *Science* 2002; **297**: 1003–1007.
- 29 Stankiewicz P, Lupski JR: Genome architecture, rearrangements and genomic disorders. *Trends Genet* 2002; **18**: 74–82.
- 30 Feuk L, Carson AR, Scherer SW: Structural variation in the human genome. *Nat Rev Genet* 2006; **7**: 85–97.
- 31 Lupski JR: Genomic rearrangements and sporadic disease. *Nat Genet* 2007; **39**: S43–S47.
- 32 Wong KK, deLeeuw RJ, Dosanjh NS *et al*: A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet* 2007; **80**: 91–104.
- 33 Simon-Sanchez J, Scholz S, Fung HC *et al*: Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum Mol Genet* 2007; **16**: 1–14.
- 34 de Smith AJ, Tsalenko A, Samps N *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**: 2783–2794.
- 35 de Vries BB, White SM, Knight SJ *et al*: Clinical studies on submicroscopic subtelomeric rearrangements: a checklist. *J Med Genet* 2001; **38**: 145–150.
- 36 Cheung VG, Nowak N, Jang W *et al*: Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* 2001; **409**: 953–958.
- 37 Coe BP, Ylstra B, Carvalho B *et al*: Resolving the resolution of array CGH. *Genomics* 2007; **89**: 647–653.
- 38 Aston E, Whitby H, Maxwell T *et al*: Comparison of targeted and whole genome analysis of postnatal specimens using a commercially available array based comparative genomic hybridisation (aCGH) microarray platform. *J Med Genet* 2008; **45**: 268–274.
- 39 de Pater JM, Nikkels PG, Poot M *et al*: Striking facial dysmorphisms and restricted thymic development in a fetus with a 6-megabase deletion of chromosome 14q. *Pediatr Dev Pathol* 2005; **8**: 497–503.
- 40 de Pater JM, Poot M, Beemer FA *et al*: Virilization of the external genitalia and severe mental retardation in a girl with an unbalanced translocation 1;18. *Eur J Med Genet* 2006; **49**: 19–27.
- 41 Poot M, Kroes HY, V D Wijst SE *et al*: Dandy-Walker complex in a boy with a 5 Mb deletion of region 1q44 due to a paternal t(1;20)(q44;q13.33). *Am J Med Genet* 2007; **143A**: 1038–1044.
- 42 Poot M, Eleveld MJ, van 't Slot R *et al*: Proportional growth failure oculocutaneous albinism in a girl with a 6.87 Mb deletion of region 15q26.2 → qter. *Eur J Med Genet* 2007; **50**: 432–440.
- 43 Flipsen-ten Berg K, van Hasselt PM, Eleveld MJ *et al*: Unmasking of a hemizygous WFS1 gene mutation by a chromosome 4p deletion of 8.3 Mb in a patient with Wolf-Hirschhorn syndrome. *Eur J Hum Genet* 2007; **15**: 1132–1138.
- 44 Özgen HM, Staal WG, Barber JC *et al*: A novel 6.14 Mb duplication of chromosome 8p21 in a patient with autism and self mutilation. *J Autism Dev Disord* 2009; **39**: 322–329.
- 45 Sharp AJ, Mefford HC, Li K *et al*: A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet* 2008; **40**: 322–328.
- 46 Mefford HC, Sharp AJ, Baker C *et al*: Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 2008; **359**: 1685–1699.
- 47 Ledbetter DH: Cytogenetic technology—genotype and phenotype. *N Engl J Med* 2008; **359**: 1728–1730.
- 48 Blauw HM, Veldink JH, van Es MA *et al*: Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen. *Lancet Neurol* 2008; **7**: 319–326.
- 49 Chang JT, Nevins JR: GATHER: a systems approach to interpreting genomic signatures. *Bioinformatics* 2006; **22**: 2926–2933.
- 50 Aerts S, Lambrechts D, Maity S *et al*: Gene prioritization through genomic data fusion. *Nat Biotechnol* 2006; **24**: 537–544.
- 51 Franke L, van Bakel H, Fokkens L *et al* (2006) Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *Am J Hum Genet* **78**: 1011–1025.
- 52 Cuscó I, Medrano A, Gener B *et al*: Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. *Hum Mol Genet* 2009; **18**: 1795–1804.
- 53 Snel B, Lehmann G, Bork P *et al*: STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* 2000; **28**: 3442–3444.
- 54 Redon R, Ishikawa S, Fitch KR *et al*: Global variation in copy number in the human genome. *Nature* 2006; **444**: 444–454.
- 55 Merla G, Howald C, Henrichsen CN *et al*: Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. *Am J Hum Genet* 2006; **79**: 332–341.
- 56 Reymond A, Henrichsen CN, Harewood L *et al*: Side effects of genome structural changes. *Curr Opin Genet Dev* 2007; **17**: 381–386.
- 57 Hochstenbach R, van Binsbergen E, Engelen J *et al*: Array analysis and karyotyping: workflow consequences based on a retrospective study of 36325 patients with idiopathic developmental delay in the Netherlands. *Eur J Med Genet* 2009; e-pub ahead of print 9 April 2009.
- 58 Koolen DA, Sharp AJ, Hurst JA *et al*: Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. *J Med Genet* 2008; **45**: 710–720.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)