Small rare recurrent deletions and reciprocal duplications in 2q21.1, including brain-specific ARHGEF4 and GPR148

Avinash V. Dharmadhikari1,2,†, Sung-Hae L. Kang1,‡, Przemyslaw Szafranski1, Richard E. Person1, Srirangan Sampath1, Siddharth K. Prakash3, Patricia I. Bader4, John A. Phillips III5, Vickie Hannig5, Misti Williams5, Sherry S. Vinson7, Angus A. Wilfong8, Tyler E. Reimschisel6, William J. Craig1, Ankita Patel1, Weimin Bi1, James R. Lupski1,9,10, John Belmont1, Sau Wai Cheung1 and Pawel Stankiewicz1,2,*

1Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Room R809, Houston, TX 77030, USA, 2Interdepartmental Program in Translational Biology and Molecular Medicine, Baylor College of Medicine, Houston, TX 77030, USA, 3Department of Internal Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA, 4Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA, 5Northeast Indiana Genetic Counseling Center, Fort Wayne, IN 46845, USA, 6Division of Medical Genetics and Genomic Medicine and 7Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA, 8Division of Medical Genetics and Genomic Medicine and 9Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA, 10Texas Children’s Hospital, Baylor College of Medicine, and

Received March 15, 2012; Revised and Accepted April 20, 2012

We have identified a rare small (∼450 kb unique sequence) recurrent deletion in a previously linked attention-deficit hyperactivity disorder (ADHD) locus at 2q21.1 in five unrelated families with developmental delay (DD)/intellectual disability (ID), ADHD, epilepsy and other neurobehavioral abnormalities from 17 035 samples referred for clinical chromosomal microarray analysis. Additionally, a DECIPHER (http://decipher.sanger.ac.uk) patient 2311 was found to have the same deletion and presented with aggressive behavior. The deletion was not found in either six control groups consisting of 13 999 healthy individuals or in the DGV database. We have also identified reciprocal duplications in five unrelated families with autism, developmental delay (DD), seizures and ADHD. This genomic region is flanked by large, complex low-copy repeats (LCRs) with directly oriented subunits of ∼109 kb in size that have 97.7% DNA sequence identity. We sequenced the deletion breakpoints within the directly oriented paralogous subunits of the flanking LCR clusters, demonstrating non-allelic homologous recombination as a mechanism of formation. The rearranged segment harbors five genes: GPR148, FAM123C, ARHGEF4, FAM168B and PLEKHB2. Expression of ARHGEF4 (Rho guanine nucleotide exchange factor 4) is restricted to the brain and may regulate the actin cytoskeletal network, cell morphology and migration, and neuronal function. GPR148 encodes a G-protein-coupled receptor protein expressed in the brain and testes. We suggest that small rare recurrent deletion of 2q21.1 is pathogenic for DD/ID, ADHD, epilepsy and other neurobehavioral abnormalities and, because of its small size, low frequency and more severe phenotype might have been missed in other previous genome-wide screening studies using single-nucleotide polymorphism analyses.

†These authors contributed equally to this work.

*To whom correspondence should be addressed at: Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Room R809, Houston, TX 77030, USA. Tel: +1 713 798 5370; Fax: +1 713 798 7418; Email: pawels@bcm.edu

© The Author 2012. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
INTRODUCTION

Several low-copy repeat (LCR)-rich regions in the human genome have been associated with common well-known recurrent microdeletion and reciprocal microduplication syndromes caused by non-allelic homologous recombination (NAHR) also known as genomic disorders (1,2). Some of these genomic disorders manifest variable expressivity and incomplete penetrance of neurodevelopmental and neuropsychiatric phenotypes, e.g. deletions and duplications in 1q21.1 (3,4), 10q11.21q11.23 (5), 15q11.2 (6), 15q13.3 (7,8), 16p11.2 (9,10), 16p12.1 (11) and 16p13.11 (12).

Several rare copy-number variations (CNVs) have been identified in up to 8% of patients with attention-deficit hyperactivity disorder (ADHD) and some of these CNVs contain genes important in psychological and neurological functions (13–16). A subset of the identified loci have been previously in patients with other neurodevelopmental and neuropsychiatric phenotypes such as autism and schizophrenia, e.g. recurrent CNVs in 1q21.1, 15q11.2, 15q13.3, 16p11.2 and 16p13.11 (14,16), as well as CNVs encompassing the ASTN2, CNTN5, GABRG1 (16), A2BP1, AUTS2, CNTNAP2 and IMMP2L (13) genes. Importantly, several new candidate susceptibility genes: GRM5 and PTPRD (13), BCHE, PLEKHB1, NDUFAF2, SLC2A3 and NPY (15), CPLX2, ZBBX and PTPRN2 (16) and BDNF (17) have been reported. Recently, CNVs involving glutamate receptor gene networks (GRM1, GRM5, GRM7 and GRM8) have been reported in patients with ADHD (18).

Since conventional genome-wide association studies (GWAS) had not yielded any statistically significant loci (19–21), meta-analyses of family-based and case–control GWA screenings in patients with ADHD have been performed (22–25). However, no significant associations have been found, likely because the effects of the individual risk variants are very small and include multiple rare alleles. A number of candidate genes playing a role in the dopamine and serotonin pathways have been identified, including SLC6A3 [dopamine transporter gene (DAT1)], DRD4 (dopamine receptor 4 gene), DRD5 (dopamine receptor 5 gene), HTR1B (serotonin receptor gene) and SNAP25 (26–32). The SLC9A9 (solute carrier family 9, member 9, sodium/hydrogen exchanger) gene has also been implicated using association studies (19,33–36). Recently, integration of the various GWAS results has identified a list of 85 ADHD candidate genes, 45 of which defined a neurodevelopmental protein network involved in directed neurite outgrowth (37).

Here, using clinical array comparative genomic hybridization (aCGH) analysis, we have identified a small (450 kb unique sequence) rare recurrent deletion and reciprocal duplication in 2q21.1, including brain-specific ARHGEF4 and GPR148, in 10 unrelated families with DD/intellectual disability (ID), ADHD, autism and epilepsy.

RESULTS

aCGH analysis

We have detected four ~417–631 kb (patients 1–4) and one ~1.2 Mb (patient 5) (450 kb unique sequence) microdeletions flanked by large complex LCR clusters on chromosome 2 at 2q21.1q21.2, proximal LCR2q21.1 (~800 kb, chr2: 130,394,886–131,193,978) and distal LCR2q21.1q21.2 (~1.2 Mb, chr2: 131,646,166–132,838,104) (hg18) (Figs 1 and 2A and B) using chromosomal microarray analysis (CMA) V8.0 OLIGO (patients 2 and 4) and V8.1 OLIGO (patients 1, 3 and 5). The deletion coordinates are chr2: 131,204,290–131,621,552 (patients 1–4) and chr2: 131,204,290–132,356,498 (patient 5). Patient 5 also has a 390 kb duplication in 17p13.3 (chr17:2,942,086–3,331,777),
involving the \textit{ASPA} gene. Both CNV’s were inherited from an apparently healthy mother. Mutations in \textit{ASPA} are associated with Canavan’s disease (OMIM: #271900), which is inherited as an autosomal recessive trait (38,39). The 17p13.3 duplication involves only a 5’ portion of the \textit{ASPA} gene and thus, is rather unlikely to be deleterious. In addition, patient 5 does not manifest features of Canavan’s disease.

CMA in patient 5’s maternal grandparents did not show any evidence of the deletion and duplication, thus they most likely arose in his mother \textit{de novo}. The fact that patient 5 inherited...
the 2q21.1 deletion from his apparently healthy mother could be explained by sex differences in expressivity or paternally inherited or de novo changes in the modifier gene(s). Using CMA, patient 3’s brother and mother have also been found to carry the 2q21.1 deletion.

Additionally, we have identified five patients with reciprocal duplications at 2q21.1 using CMA V8.0 OLIGO (patients 6 and 7) and V8.1 OLIGO (patients 8–10) (Fig. 1). The duplication coordinates are chr:2:131,204,290–131,621,552. All nine of patient 9’s siblings were tested by CMA; only one had the 2q21.1 duplication.

Copy-number verification by fluorescence in situ hybridization and multiplex ligation-dependent probe amplification (MLPA) analyses

Deletions in patients 3, 4 and 5 and parents of patient 5 were verified by fluorescence in situ hybridization (FISH) analysis with the chromosome 2q21.1-specific BAC clone RP11-675I14 (Fig. 2C). The 17p13.3 duplication in patient 5 was verified with the BAC clone RP11-64J4. Duplications in patients 6, 7 and 9 were verified with the BAC clone RP11-581I23. In patients 7 and 9, the 2q21.1 duplications were found to be inherited from their fathers. Inheritance of the 2q21.1 duplications in patients 6, 8 and 10 is not known.

The 2q21.1 deletions in patients 1 and 2 and 2q21.1 duplications in patients 8 and 10 were verified by multiplex ligation-dependent probe amplification (MLPA) (Fig. 2E–H).

Bioinformatics and in silico sequence analysis

Bioinformatic analyses of the 2q21.1 deletion-flanking LCRs, including the Miropeats program, revealed two directly oriented paralogous subunits ~109 kb in size and of 97.7% overall DNA sequence identity that can serve as a potential substrate for NAHR (chr:2:131,084,465–131,193,978 and chr:2:131,691,307–131,800,522) (Figs 1 and 3).

We have identified 15 and 13 7-mer CCTCCCT and 3 and 4 degenerate 13-mer CCNCTNNCNCC recombination motifs known to be associated with hotspots for homologous recombination crossover sites (40), in the directly oriented proximal and distal ~109 kb subunits, respectively (Fig. 4A).

Long-range polymerase chain reaction and DNA sequencing

Using long-range polymerase chain reaction (PCR) primers F: 5′-TGTGCACTTTACACTAATCCTGCTGA, specific to the proximal ~109 kb subunit LCR2q21.1 and R: 5′-ACCAATGCTATATCACGCCACAGTTTGCCTAGA, specific to the distal paralogous copy within LCR2q21.1q21.2, we have obtained a patient-specific junction fragment in patient 2 (Fig. 2D). To locate potential NAHR sites within the LCRs, we sequenced the patient-specific junction fragment PCR product. Sequence analysis revealed NAHR sites within a 181 bp GC-rich interval between two informative cis-morphic nucleotides chr:2:131,161,844–131,162,024 and chr:2:131,768,533–131,768,713 in the proximal and distal LCR subunits, respectively (Fig. 4B and C). Both the proximal and distal NAHR sites align in AluS q SINE repeats.

We did not identify the NAHR sites in patients 1, 3 and 4 because of insufficient numbers of cis-morphisms that precluded the primer design for the remaining portion of the ~109 kb LCR subunits. The non-recurrent deletion in patient 5 was likely not mediated by NAHR, given the absence of directly oriented LCR subunits in LCR2q21.1q21.2 (Fig. 1).

Deletion and duplication frequency in affected and control populations

We have identified a total number of five 2q21.1 deletions and five reciprocal duplications from 17,035 patients, including those with ADHD (~2.5%), autism spectrum disorders (ASDs) (~9.2%), epilepsy (~5.5%) and DD/DD (~28.0%). The percentage of cases with no indications is 17.0%. The presence of DD/DD in all our patients may reflect an ascertainment bias, because individuals are often referred for CMA for general indications such as ID/DD.

We have compared the frequency of the 2q21.1 deletion (between 131.19 and 131.65 Mb) among our study population to the frequency in 14,000 controls to determine whether it predisposes individuals to an abnormal phenotype. No 2q21.1 deletions have been found in five control groups consisting of 2792 individuals (41) (British origin, Affymetrix GeneChip 500 K Mapping array), 2493 individuals (42) (Worldwide and European descent, Illumina 317K arrays, Illumina 550 + 317 K supplemented with 240S single-nucleotide polymorphism (SNP) arrays, Illumina 650Y arrays), 1152 individuals (43) (Caucasian, African American, Asian and Hispanic origin, Affymetrix GeneChip 100 and 500 K arrays), 450 individuals (44) (African, Asian and European descent, Agilent 105 K arrays) and 5088 individuals in dbGaP. There was only one case of an 828 kb overlapping deletion (chr:2:131,316,199–132,144,820) identified in an additional fifth control group consisting of 2025 individuals (45) (Caucasian, African American and Asian Americans, Illumina Infinium II HumanHap550 BeadChip), who was later reevaluated as affected and thus, inappropriate for inclusion in an ‘unaffected control’ group (H. Hakonarson, personal communication). In addition, there was no 2q21.1 deletion or duplication in the Database of Genomic Variants (http://projects.tcag.ca/variation/).

No 2q21.1 duplications were found in four control groups consisting of 6120 individuals (42–45). There was only one case of a 367.6 kb duplication (chr:2:131,262,791–131,630,364) identified in the control group consisting of 2792 individuals (41).

DISCUSSION

Our data suggest that small rare recurrent deletions in 2q21.1 are pathogenic for DD/ID, ADHD, epilepsy and other neurobehavioral abnormalities. In support of this notion, a DECIPHER (http://decipher.sanger.ac.uk) patient 2311 presenting with aggressive behavior was found to have the same deletion. Interestingly, the chromosome 2q21.1 region has been identified previously as linked in patients diagnosed with ADHD using an endophenotype approach for linkage...
Figure 3. DNA sequence homology between the proximal LCR2q21.1 (chr2:130,394,886–131,193,978) and distal LCR2q21.1q21.2 (chr2:131,646,166–132,838,104) clusters for paralogous subunits larger than 1 kb in size (hg18). (Top) UCSC segdup track representing the proximal LCR2q21.1 cluster. (Middle) Results of Miropeats program analysis between the proximal LCR2q21.1 and distal LCR2q21.1q21.2 clusters. (Bottom) UCSC segdup track representing the distal LCR2q21.1q21.2 cluster.
analysis (46). However, this 2q21.1 deletion has not been found in larger-sized genome-wide CNV studies in 335 (13), 366 + 825 (14), 99 (15), 248 (16) and 1013 (18) individuals with ADHD. Given the small size (450 kb) of the unique sequence between the flanking LCRs of the 2q21.1 deletion region, low frequency and a more severe phenotype than isolated ADHD, it is possible that patients with this deletion might have been missed (or not included) in the aforementioned studies. For instance, the study performed by Williams et al. (14) using SNP arrays analyzed only CNVs 500 kb in size.

We have also identified five individuals with reciprocal duplications involving of 2q21.1. Given the parental data and the fact that the 2q21.1 duplication was present in an apparently healthy individual (41), we have concluded that the clinical consequences of the 2q21.1 duplication are unknown at this time and more patients need to be studied for better genotype–phenotype correlations.

The rearranged region in our patient cohort includes five RefSeq genes: GPR148, FAM123C, ARHGEF4, FAM168B and PLEKHB2. ARHGEF4 (Rho guanine nucleotide exchange factor 4; also called ASEF) is expressed only in the brain and has been implicated in neural morphogenesis and connectivity, regulating the activity of small Rho GTPases by catalyzing the exchange of bound GDP by GTP (48). Defects in guanine nucleotide exchange factor genes are known to contribute to a number of disease conditions such as: slowed nerve conduction velocity (ARHGEF10, MIM 608236), X-linked mental retardation (ARHGEF6, MIM 300436), early infantile epileptic encephalopathy (ARHGEF9, MIM 300607), Charcot–Marie–Tooth disease type 4H (FGD4 MIM 609311), coronary heart disease (KALRN, MIM 608901), faciogenital dysplasia with ADHD (FGD1, MIM 305400) and breast cancer (ARHGEF5, MIM 114480). Given that expression of ARHGEF4 is restricted to the brain and defects in guanine nucleotide exchange factor genes have been implicated in a number of neurological diseases, including ADHD, we suggest that ARHGEF4 is a good candidate gene for neurodevelopmental and neurobehavioral disorders.

GPR148 encodes a G-protein-coupled receptor (GPR) protein and its expression is restricted to the brain and testes (49). GPRs contain extracellular N terminus, seven transmembrane regions, and an intracellular C terminus and transduce extracellular signals through heterotrimeric GTP-binding proteins, G-proteins. To date, a couple of genes encoding GPRs have been shown to be pathogenic in humans. GPR56 is
responsible for bilateral frontoparietal polymicrogyria (MIM 606854) and GPR98 is responsible for familial febrile seizures (MIM 604352) and Usher syndrome type IIC (MIM 605472). In addition, GPR85 has been shown to be involved in determining brain size, modulating diverse behaviors and potentially in vulnerability to schizophrenia (50) and haploinsufficiency of GPR63 was postulated to play a role in the etiology of the ASD (51).

Studies in mice revealed expression of Fam123c in the central and peripheral nervous system, suggesting an involvement in murine neurogenesis (52). However, the function of this gene in humans is unclear at this time. Similarly, very little is known about PLEKHB2 (Pleckstrin homology domain-containing family B member 2) (53) and FAM168B. However, it is interesting to note that Lesch et al. (15) reported an ≈500 kb duplication in 11q13.4, harboring the brain-specific PLEKHB1 gene expressed in primary sensory neurons, in two siblings with ADHD, who inherited the CNV from their affected mother.

The proximal chromosome 2 region 2q11.2-q21.1 is LCR-rich. Homozygous LCR-flanked recurrent microdeletions involving NPHP1 in 2q13 have been reported in patients with familial juvenile nephronophthisis 1 (NPHP1, OMIM 256100). The more distally adjacent segment in 2q13 encompasses a 1.71 Mb recurrent microdeletion associated with ID and dysmorphism (54) and chromosome region 2q13q14.1 harbors the evolutionary breakpoint of the ancestral centric fusion of two chromosomes in non-human primates (55). Bioinformatic analyses of the 2q21.1 flanking large and complex LCR clusters revealed about 109 kb directly oriented paralogous LCR subunit pair that can serve as substrates for NAHR. In patient 2, we identified NAHR sites in a GC-rich portion of these LCRs (Fig. 4). Interestingly, polyurine and pyrimidine sequences with the GC content of at least 50% have also been associated with human recombination hotspots (56,57).

Our findings indicate that small rare recurrent deletions of 2q21.1 between LCR2q21.1 and LCR2q21.1q21.2 are pathogenic for DD/ID, ADHD, epilepsy and other neurobehavioral abnormalities. This segment harbors five annotated genes with ARHGFE4 and GPR148 as possible candidate genes. We propose that additional studies in patients with these neurodevelopmental phenotypes should include analysis of the ARHGFE4 and GPR148 genes for copy-number changes and/or sequence mutations.

**MATERIALS AND METHODS**

**Subject recruitment**

Individuals with deletions and duplications involving 2q21.1 were identified after referral for clinical chromosomal microarray testing to the Medical Genetics Laboratories (MGL) at Baylor College of Medicine (BCM).

**Patients**

Patient 1 is a 12.5-year-old boy with ADHD, social anxiety and language impairment evidenced by spelling and pronunciation difficulties. He had memory deficits and atypical behaviors such as head banging. He was able to perform fourth-grade class work during his home schooling. Despite a normal gestational and birth history and size, his weight and height accelerated at about 3.5 years. Gynecomastia was noted at 4 years. He developed diabetes type 2 at 10 years, and acanthosis nigricans, obesity, tall stature and continued gynecomastia at the time of genetic evaluation. His bone age was 15.5–16 years. He was also found to have an occipital cystic hygroma. Methylation testing to rule out Prader–Willi syndrome was normal. There is a positive paternal family history for diabetes type 2 and polycystic kidney disease. His father is deceased from liver cancer. Normal parental intelligence is reported. His mother is not available for genetic testing.

Patient 2 is a 5.5-year-old female referred for evaluation of possible Williams–Beuren syndrome. She was born to a 29-year-old G4P1Ab2 mother at 42 weeks of gestation. Her birth weight was 7 pounds 4 ounces, and she was 21 inches in length. The pregnancy was complicated by maternal diabetes. She smiled at 3 months, rolled at 4 months, sat at 6 months, crawled at 5–6 months, stood at 7 months and walked at 10 months. Her mother reports a history of easy distraction, hyperactivity and resistance to instructions. Family history is significant for ADHD, including father and two paternal half siblings. Upon physical exam, she was noted to have mild dysmorphic features, including hypertelorism, wide spaced teeth and clinodactyly. Her parents are unavailable for genetic testing.

Patient 3 is a 6.5-year-old boy with mild DD, ADHD, behavioral problems, seizures, abnormal EEG and dysmorphic features. During the neuropsychological evaluation at age 5 years 9 months, he was found to have a Full Scale IQ of 83 with 16.1 percentile (Wechsler Preschool and Primary Scale of Intelligence-III, WPPSI-III). His language was impaired with elements of dyslexia, dyscalculia and dysnomia. His speech was rapid and not understandable. He could not function in a regular classroom and required supervision. He displayed poor memory, poor organizational skills and distractibility. His behavior was aggressive, self-focused, defiant and anxious. There is a positive family history of bipolar disease (paternal grandmother and paternal great aunt), schizophrenia (paternal uncle), depression and epilepsy (paternal aunts). Maternal family history is positive for drug addiction and alcoholism. His 11-year-old brother has been diagnosed with ADHD and his mother has a diagnosis of depression. She was reported to be unable to make decisions, has low self-esteem and is anxious, fearful and paranoid. Upon physical exam, he was at the 10th percentile for height, 25th percentile for weight and 25th percentile for occipital-frontal circumference (OFC). He was noted to have slanted palpebral fissures, prominent lips, high palate, cupped ears with flat antihelix, upturned nose with bulbous tip, broad thumbs, clinodactyly of fifth fingers, joint hyper mobility, flat feet, hockey stick Palmer creases and hirsutism. He was prescribed lamotrigine for seizures.

Patient 4 is a 17-year-old adopted male with ADHD, encephalopathy, borderline cognitive ability and conduct disorder. His biological parents had cognitive delays and disabilities. He was prescribed methylphenidate HCL extended release for ADHD.

Patient 5 is a 3-year-old boy with profound DD, intractable epilepsy and microcephaly. There was a gestational history of
poor fetal movement. He was delivered by C-section with normal birth weight (7 pounds 6 ounces) and normal Apgars. The newborn history was characterized by temperature instability, feeding difficulties (G-tube placement) and jaundice. At 3 weeks, he had acute encephalopathy and began to have recurrent seizures that were initially tonic and tonic-clonic. He subsequently developed refractory flexor infantile spasms having up to 500 per day. He has failed medication trials with phenobarbital, levetiracetam, topiramate, oxcarbazepine, lamotrigine, phenytoin and prednisolone. He is currently on the ketogenic diet and is prescribed zonisamide. His evaluation included normal test results for Fragile X, ARX (including sequence analysis), plasma amino acids and urinary organic acids. Cerebral spinal fluid studies included normal results for glucose, protein, amino acids, lactate, pyruvate and neurotransmitters. Brain magnetic resonance imaging (MRI) showed bilateral symmetric posterior occipital cystic encephalomalacia with cortical laminar necrosis and cerebral atrophy consistent with hypoxic ischemia. An MRI SPECT (single photon emission computed tomography) revealed a global decrease in NAA (N-acetyl-aspartate, left more than right) without lactate peak as well as an increased myoinositol, most likely secondary to gliosis. Upon physical exam, he is at the 50th percentile for height, 50th percentile for weight and OFC was at 40.4 cm at 19 months (−6 SD). He is non-verbal and does not follow objects when they are presented. He shows axial hypotonia with mild distal spasticity. Reflexes show hypertonia. He has severe global DD with cortical blindness, cannot roll-over or sit and is taking nothing by mouth, all nutrition is received through a gastrostomy tube.

### Duplication patients

We have also identified five unrelated families with reciprocal duplications (patients 6-10, ages 6.5, 7.0, 9.5, 9.5, and 1.0 years, respectively). Although detailed clinical information was not obtained, the indication provided for testing for most patients included ASDs (patients 6, 7 and 9), DD/ID (patients 7, 8 and 9), ADHD (patient 7), organic encephalopathy (patient 7), seizures (patient 9) and cardiac defect (patient 10). Patient 8 was adopted, but the biological father was reported to be unaffected. Patient 9 has significant family history with six of nine siblings affected with autism, Asperger syndrome or epilepsy. The father of patient 9 also has epilepsy. Parental samples for the other three patients were unavailable for testing (Table 1).

### DNA isolation

Genomic DNA was extracted from peripheral blood via the Puregene DNA isolation kit (Gentra System, Minneapolis, MN, USA).

### Table 1. Summary of genetic and clinical findings in patients 1–10

<table>
<thead>
<tr>
<th>Pt</th>
<th>Age (years)</th>
<th>Sex</th>
<th>CNV coordinates (hg18)</th>
<th>Gain/loss</th>
<th>Parental Inheritance</th>
<th>2q21.1 CNVs in other family members</th>
<th>Other reported CNVs</th>
<th>Clinical phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.5</td>
<td>M</td>
<td>chr2:131,204,290−131,621,552</td>
<td>Loss</td>
<td>Parental DNA unavailable</td>
<td>Unknown</td>
<td>None</td>
<td>ADHD, social anxiety, language impairment, memory deficits, atypical behaviors, gynecomastia, diabetes type 2, Acanthosis nigricans, occipital cystic hygroma, obesity and tall stature</td>
</tr>
<tr>
<td>2</td>
<td>5.5</td>
<td>F</td>
<td>chr2:131,204,290−131,621,552</td>
<td>Loss</td>
<td>Parental DNA unavailable</td>
<td>Unknown</td>
<td>None</td>
<td>Easy distraction, hyperactivity and resistance to instructions</td>
</tr>
<tr>
<td>3</td>
<td>6.5</td>
<td>M</td>
<td>chr2:131,204,290−131,621,552</td>
<td>Loss</td>
<td>Inherited from affected mother</td>
<td>Brother with ADHD</td>
<td>None</td>
<td>Mild DD, ADHD, language impairment, behavioral problems, seizures, abnormal EEG and dysmorphic features</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>M</td>
<td>chr2:131,204,290−131,621,552</td>
<td>Loss</td>
<td>Adopted</td>
<td>Unknown</td>
<td>None</td>
<td>ADHD, encephalopathy, borderline cognitive ability and conduct disorder</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>M</td>
<td>chr2:131,204,290−132,356,498</td>
<td>Loss</td>
<td>Inherited from unaffected mother</td>
<td>Unknown</td>
<td>chr17:2,942,086-3,331,777 (gain)</td>
<td>None</td>
</tr>
<tr>
<td>6</td>
<td>6.5</td>
<td>M</td>
<td>chr2:131,204,290−131,621,552</td>
<td>Gain</td>
<td>Parental DNA unavailable</td>
<td>Unknown</td>
<td>None</td>
<td>ASDs</td>
</tr>
<tr>
<td>7</td>
<td>9.5</td>
<td>M</td>
<td>chr2:131,204,290−131,621,552</td>
<td>Gain</td>
<td>Inherited from unaffected father</td>
<td>Unknown</td>
<td>None</td>
<td>DD, ADHD, ASDs, organic encephalopathy</td>
</tr>
<tr>
<td>8</td>
<td>7.0</td>
<td>F</td>
<td>chr2:131,204,290−131,621,552</td>
<td>Gain</td>
<td>Adopted</td>
<td>Unknown</td>
<td>None</td>
<td>DD/ID</td>
</tr>
<tr>
<td>9</td>
<td>9.5</td>
<td>F</td>
<td>chr2:131,204,290−131,621,552</td>
<td>Gain</td>
<td>Inherited from affected father</td>
<td>Sister with Asperger syndrome</td>
<td>None</td>
<td>DD/ID, ASDs, seizures</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>M</td>
<td>chr2:131,204,290−131,621,552</td>
<td>Gain</td>
<td>Parental DNA unavailable</td>
<td>Unknown</td>
<td>None</td>
<td>Cardiac defect</td>
</tr>
</tbody>
</table>
aCGH analysis

A total of 17,035 unrelated subjects referred for CMA at the MGL at BCM were screened using custom-designed exon-targeted aCGH oligonucleotide microarrays (V8.0 and V8.1 Oligo, 180 K) designed by the MGL at BCM (http://www.bcm.edu/geneticlabs/) and manufactured by Agilent Technology (Santa Clara, CA, USA) as previously described (58).

FISH analysis

Confirmatory and parental FISH analyses were performed using standard procedures.

MLPA analysis

Probes for MLPA analysis were designed using the freely available H-MAPD Web server (http://genomics01.arcane.stony-brook.edu/mlpa/cgi-bin/mlpa.cgi) and mapped in exons for the following genes: GPR148, FAM123C, ARHGEF4, FAM168B and PLEKHB2. MLPA reactions were carried out using SALSA MLPA reagents and P300 reference probe mix as per instructions (MRC-Holland, Amsterdam). MLPA product (1.1 μl) and 0.25 μl of GS500 Liz Size Standard were added to 10 μl of formamide and loaded onto an ABI 3730x1 capillary electrophoresis machine (Applied Biosystems, Foster City, CA, USA). Data were analyzed using Genemarker MLPA analysis software (SoftGenetics, State College, PA, USA).

Bioinformatics and in silico sequence analysis

Genomic sequences based on oligonucleotide coordinates from the aCGH experiments were downloaded from the UCSC genome browser (NCBI build 36, March 2006, http://www.genome.ucsc.edu) and assembled using BLAT (UCSC browser), BLAST2 (NCBI, http://blast.ncbi.nlm.nih.gov/) and the Sequencher v4.8 software (Gene Codes, Ann Arbor, MI, USA). Interspersed repeat sequences were identified using RepeatMasker (http://www.repeatmasker.org). DNA GC content was analyzed with cpgplot (www.ebi.ac.uk/Tools/emboss/cpgplot).

To better visualize the chromosome architecture in the 2q21.1 region, we have used the ICAass (v 2.01) (The Genome Institute at Washington University, St Louis, MO, USA). The graphical display was performed using emboss/miropeats (v 2.01) (The Genome Institute at Washington University, St Louis, MO, USA). The program was run using two thresholds of 1000 bp.

Long-range PCR and DNA sequencing

Given the recurrent breakpoint mapping to flanking LCRs, we hypothesized NAHR as a most likely molecular mechanism of formation of the 2q21.1 deletions. Thus, we designed long-range PCR primers to harbor at least three nucleotide mismatches based on paralogous sequence variants (also known as cis-morphisms) (59,60), forward primer specific to the directly oriented ~109 kb subunit in the proximal 2q21.1 LCR cluster and reverse primer in the paralogous copy in the distal 2q21.1q21.2 LCR cluster. This allowed preferential amplification of the predicted junction fragment, generated in the recombinant LCR as a product of NAHR. Primers were designed using the Primer 3 software (http://frodo.wi.mit.edu/primer3/). Amplification of ~10–20 kb fragments was performed using Takara LA Taq Polymerase (TaKaRa Bio USA, Madison, WI, USA) following the manufacturer’s protocol. Amplification of the GC-rich region in the LCRs was performed in the presence of 4% dimethyl sulfoxide, with the following conditions: 94°C for 1 min, followed by 30 cycles (94°C for 30 s, 68°C for 7–12 min) and 72°C for 10 min. PCR products were treated with ExoSAP-IT (USB, Cleveland, OH, USA) to remove unconsumed dNTPs and primers, and directly sequenced by Sanger di-deoxynucleotide sequencing using the initial primers and primers specific for both proximal and distal copies of the LCRs (Lone Star Labs, Houston, TX, USA).

ACKNOWLEDGEMENTS

We thank Dr Joanna Wiszniewska for helpful discussion.

Conflict of Interest statement. J.R.L. is a consultant for Athena Diagnostics, owns stock in 23andMe and Ion Torrent Systems Inc. and is a co-inventor on multiple US and European patents for DNA diagnostics. Furthermore, the Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from molecular diagnostic testing (MGL, http://www.bcm.edu/geneticlabs/).

FUNDING

This study was supported in part by the Intellectual and Developmental Disabilities Research Center (IDDRC, grant P30 HD024064) and the National Institute of Neurological Disorders and Stroke (National Institutes of Health, grant R01 NS058529) to J.R.L.

REFERENCES