Recurrent Reciprocal Genomic Rearrangements of 17q12 Are Associated with Renal Disease, Diabetes, and Epilepsy

Heather C. Mefford, Séverine Clauin, Andrew J. Sharp, Rikke S. Moller, Reinhard Ullmann, Raj Kapur, Dan Pinkel, Gregory M. Cooper, Mario Ventura, H. Hilger Ropers, Niels Tommerup, Evan E. Eichler, and Christine Bellanne-Chantelot

Most studies of genomic disorders have focused on patients with cognitive disability and/or peripheral nervous system defects. In an effort to broaden the phenotypic spectrum of this disease model, we assessed 155 autopsy samples from fetuses with well-defined developmental pathologies in regions predisposed to recurrent rearrangement, by array-based comparative genomic hybridization. We found that 6% of fetal material showed evidence of microdeletion or microduplication, including three independent events that likely resulted from unequal crossing-over between segmental duplications. One of the microdeletions, identified in a fetus with multicystic dysplastic kidneys, encompasses the *TCF2* gene on 17q12, previously shown to be mutated in maturity-onset diabetes, as well as in a subset of pediatric renal abnormalities. Fine-scale mapping of the breakpoints in different patient cohorts revealed a recurrent 1.5-Mb de novo deletion in individuals with phenotypes that ranged from congenital renal abnormalities to maturity-onset diabetes of the young type 5. We also identified the reciprocal duplication, which appears to be enriched in samples from patients with epilepsy. We describe the first example of a recurrent genomic disorder associated with diabetes.

Genomic disorders result from nonallelic homologous recombination (NAHR) between low-copy repeats and occur in ~1 in 1,000 live births.[1] The phenotypes of many of the known genomic disorders include developmental delay and mental retardation (e.g., velocardiofacial syndrome, Williams-Beurens syndrome, and Smith-Magenis syndrome). Therefore, many of the large screens for novel genomic disorders have focused on individuals with mental retardation.[2–7] Previously, we developed a BAC array targeted to 130 "rearrangement hotspots," defined as regions of the genome with an architecture suggestive of a susceptibility to recurrent microdeletion and/or duplication. Use of this array to evaluate 316 unaffected individuals[8,9] and 290 individuals with idiopathic mental retardation[7] resulted in the identification of copy-number polymorphisms, as well as novel genomic disorders.

Interestingly, despite an architecture that predicts a susceptibility to rearrangement, many of these 130 hotspot regions have never been observed as a copy-number variant in the unaffected and developmentally delayed individuals studied. We hypothesized that these "invariant" regions contain genes essential for normal morphogenesis and are involved in pathways other than those critical for normal cognitive development. To test this hypothesis, we evaluated DNA from fetal samples with one or more congenital anomalies for which the pathology was well

documented. We reasoned that rearrangements resulting in major malformations or incompatibility with life would be found in such prenatal cases. We show here that one of these microdeletions associated with a fetal sample with grossly abnormal, dysplastic multicystic kidneys is the result of a recurrent rearrangement of 17q12 mediated by segmental duplications. De novo deletions with identical breakpoints are found in living patients with a range of phenotypes, from those with renal abnormalities[10,11] to individuals whose primary diagnosis is maturity-onset diabetes of the young type 5 (MODY5 [MIM 137920]).[12] This is the first example of a recurrent genomic disorder associated with diabetes and one of the few examples of a contiguous gene-deletion syndrome without mental retardation. We also discovered the reciprocal duplication in individuals with epilepsy and/or mental retardation, as well as in unaffected individuals. These data suggest considerable variability in expressivity of the phenotype and widen the spectrum of diseases caused by genomic disorders.

Material and Methods

DNA Samples

DNA samples were obtained from prenatal autopsy specimens (*n* = 155) and from individuals with renal disease and/or MODY5,

Table 1. Abnormalities Detected by Array CGH in a Series of 155 Fetal Cases with One or More Congenital Anomalies

Case	Fetal Anomalies	Alteration	hg17 Coordinates		Alteration Size	Verification[a]	Comments
			Start	Stop			
FA-142	Multicystic dysplastic kidneys	dup17q23	55061600	55218400	157 kb	Oligo array	Proximal breakpoint in *CLTC* gene; distal breakpoint in *TMEM49*; alteration within hotspot region but not mediated by segmental duplications (fig. 2*a*)
FA-180	Posthemorrhagic hydrocephalus and cranio-facial and vertebral anomalies	del16p13.11	14780000	16770000	1.9 Mb	Oligo array	Breakpoints in polymorphic segmental duplications (fig. 2*c*)
FA-275	Multicystic dysplastic kidneys	del17q12	31836750	33582750	1.8 Mb	Oligo array	Breakpoints in polymorphic segmental duplications (see text for details)
FA-328	Organomegaly and intrauterine demise	Trisomy 9p, mosaic			p-arm	None	Prenatal karyotype reported as normal; evidence of trisomy 9p in fetal liver tissue (average $\log_2$ ratio of 0.34 for 9p clones and 0.53 for X clones); no trisomy in placental tissue from same case.
FA-430	Multicystic dysplastic kidneys	47, XXY			X chromosome	None	Likely an incidental finding; fetal karyotype was not obtained because culture failed to grow
FA-441	Holoprosencephaly	dup5p15.2-pter del7q36.1-qter	1 148600000	12100000 158550000	12.1 Mb 10 Mb	32K	Likely unbalanced translocation; prenatal karyotype reported as normal; 7q deleted region contains *SHH* gene, known to cause holoprosencephaly
FA-457	Tetralogy of Fallot	del1p36.23-p36.13 dup1p36.13	8700000 16450000	16000000 16950000	7.3 Mb 500 kb	32K	Large interstitial deletion with additional 500-kb duplication at proximal breakpoint; breakpoints in unique sequence
FA-460	Craniosynostosis and gastrointestinal and limb anomalies	dupXq13	71600000	71800000	200 kb	32K	Proximal breakpoint in unique sequence; distal in large duplicated region
FA-602	Congenital diaphragmatic hernia	del15q25	81011700	83540700	2.5 Mb	Oligo array	Breakpoints in polymorphic segmental duplications; region contains 22 genes (fig. 2*b*)

[a] 32K = whole-genome tiling-path BAC array (n = 32,433 BAC clones) (see the "Material and Methods" section).
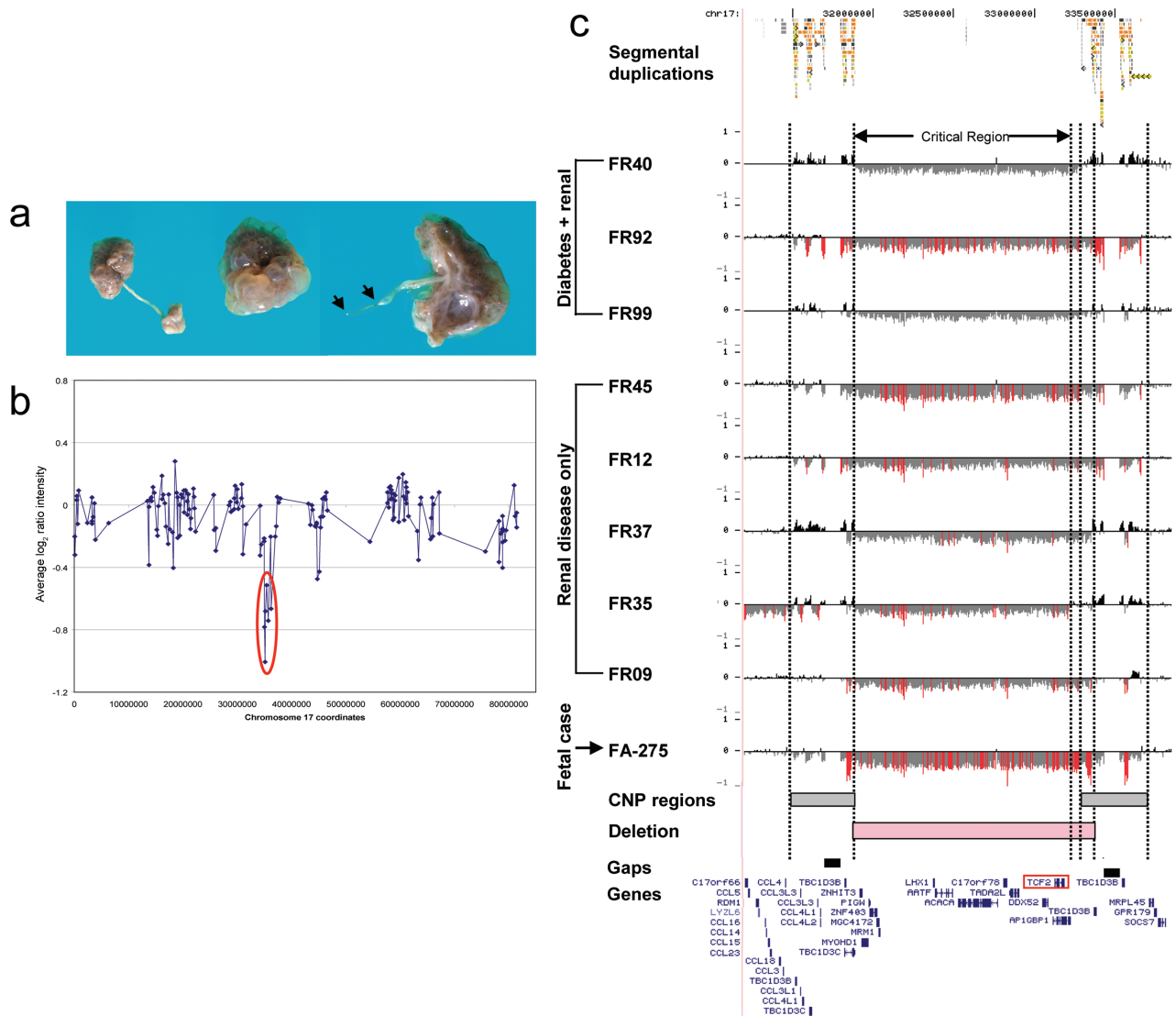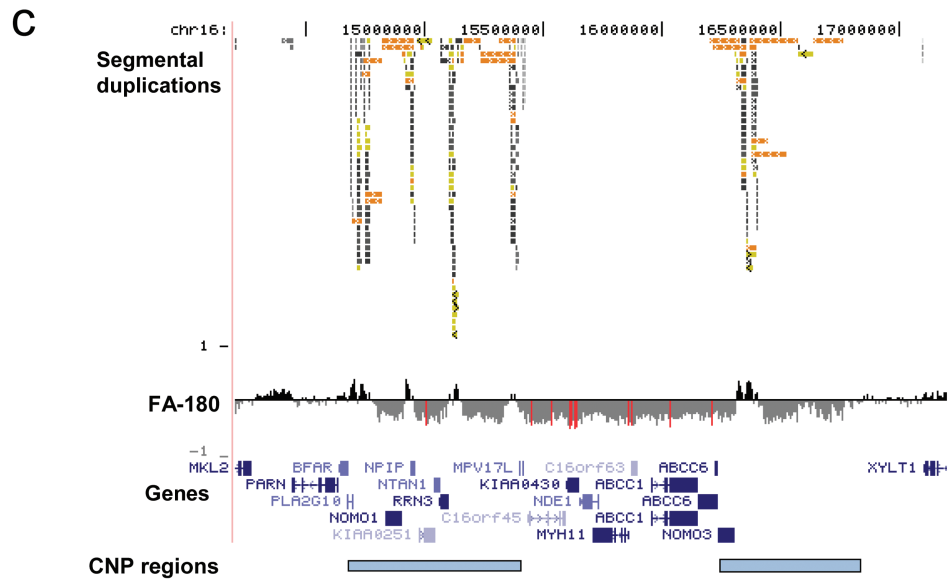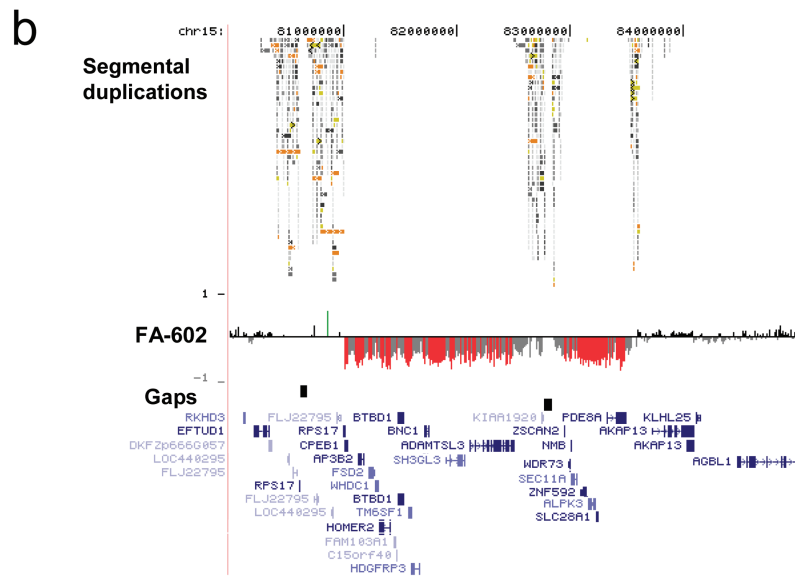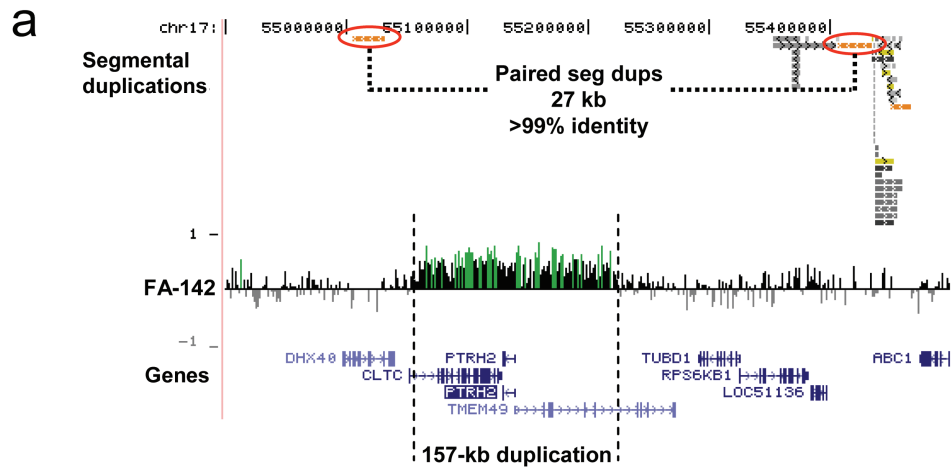
Figure 1. Recurrent microdeletion of 17q12. *a,* Kidney photographs from case FA-275, showing multiple cysts bilaterally with enlargement of the left kidney, hypoplastic bladder, and atresia of the left ureter. *b,* BAC array data for case FA-275, showing deletion of a chromosomal region represented by seven chromosome 17 BAC clones. *c,* Structural resolution of a 1.5-Mb microdeletion region of 17q12 by use of a high-density oligonucleotide microarray. Eight of the nine cases have proximal and distal breakpoints mapping to flanking segmental duplications. Plots show the $\log_2$ ratio (*Y*-axis) for 48,256 probes (*X*-axis) in this region of 17q12 (hg17 coordinates chr17:31200000–33850000). For each individual, deviations of probe $\log_2$ ratios from zero are depicted by gray and black lines, with those exceeding a threshold of 0.9 SDs (duplications) and 1.5 SDs (deletions) from the mean probe ratio colored green and red to represent relative gains and losses, respectively. Bars define regions of copy-number variation (*shaded gray*) in controls and common recurrent deletion (*shaded pink*).

with the appropriate institutional-review-board approval. Fetal liver tissue was obtained from Children's Hospital and Regional Medical Center. All samples were from fetuses that underwent autopsy after elective termination or fetal demise between 1995 and 2006. DNA was extracted using Gentra PureGene DNA extraction kit. The renal pediatric and MODY5 cases represent patients who were analyzed previously for *TCF2* molecular abnormalities.[10–12] Two control groups were used to assess the extent of normal copy-number variation. The first control group consisted of 316 unrelated individuals (HapMap and Diversity pop-

ulation) who had been tested using the same BAC duplication microarray.[8,9] A second control population, comprising 960 unrelated white adults (aged 40–70 years) from the United States, was genotyped using the HumanHap300 Genotyping BeadChips (Illumina), which consists of ~317,000 HapMap SNPs spread throughout the genome. Each individual was enrolled in the Pharmacogenomics and Risk of Cardiovascular Disease (PARC) study, which aims to identify genetic contributors to the variable efficacy of statin drugs for cardiovascular disease (PharmGKB: PARC Profile). Hybridizations, data analysis, and copy-number

a

Segmental duplications

Paired seg dups
27 kb
>99% identity

FA-142

Genes

DHX40 PTRH2 TUBD1 ABC1
CLTC PTRH2 RPS6KB1
TMEM49 LOC51136

157-kb duplication

b

Segmental duplications

FA-602

Gaps

RKHD3 FLJ22795 BTBD1 KIAA1920 PDE8A KLHL25
EFTUD1 RPS17 BNC1 ZSCAN2 AKAP13
DKFZp666G057 CPEB1 ADAMTSL3 NMB AKAP13
LOC440295 AP3B2 SH3GL3 WDR73
FLJ22795 FSD2 SEC11A AGBL1
RPS17 WHDC1 ZNF592
FLJ22795 BTBD1 ALPK3
LOC440295 TM6SF1 SLC28A1
HOMER2
FAM103A1
C15orf40
HDGFRP3

c

Segmental duplications

FA-180

Genes

MKL2 BFAR NPIP MPV17L C16orf63 ABCC6 XYLT1
PARN NTAN1 KIAA0430 ABCC1
PLA2G10 RRN3 NDE1 ABCC6
NOMO1 C16orf45 MYH11 ABCC1
KIAA0251 NOMO3

CNP regions

analysis, with particular reference to chromosome 17q12 (137 probes within the critical region), were performed according to published protocols.[13]

Array-Based Comparative Genomic Hybridization (Array CGH)

DNA was hybridized to a custom BAC array consisting of 2,007 clones targeted to regions of the genome flanked by segmental duplications, as described elsewhere.[9] This array includes all regions associated with known genomic disorders and an additional ~105 regions with similar genomic architecture. Because of the targeted nature of this array, we will not detect rearrangements that are not mediated by segmental duplications. Regions were scored as copy-number variant if the $\log_2$ ratio of two or more consecutive clones each exceeded twice the SD of the autosomal clones in dye-swap replicate experiments.[7] We used a whole-genome tiling path array containing 32,433 BAC clones[14] or a custom oligonucleotide array (NimbleGen Systems), consisting of 385,000 isothermal probes (length 45–75 bp) covering several chromosomal regions, including a 3-Mb region of chromosome 17q12 (55,888 probes; mean density 1 probe per 53 bp) to refine the breakpoints. Hybridizations were performed as described elsewhere,[15] and a single unaffected male (GM15724 [Coriell]) was used as reference. Oligonucleotide array data are available at the Human Genome Structural Variation Project Web site.

Quantitative PCR

Oligonucleotides for real-time quantitative PCR assays were selected using Primer3 software.[16] Primer sequences (5′→3′) for exon 1 primers are as follows: forward primer ATTTCCTGGTGCGAGT-TTTG and reverse primer CAGGGGATGACTCCTGAAGA. PCRs were performed in an LC480 machine (Roche) by use of SYBR Green PCR Master Mix (Applied Biosystems). PCR conditions were 95°C for 5 min, followed by 40 cycles at 95°C for 15 s, 55°C for 20 s, and 72°C for 20 s. Reactions were performed in triplicate, with a final reaction volume of 10 μl containing 10 ng DNA, 1.6 μM primer, and 5 μl SYBR Green Master Mix. Control primers were placed in the albumin gene.

FISH

Metaphase spreads and nuclei were obtained from phytohemagglutinin-stimulated peripheral lymphocytes of individuals 6498 and 6840 by standard procedures. FISH experiments were performed using fosmid WIBR2-6022g13 (NCBI May 2004 assembly [hg17] coordinates chr17:32364308–32399464), directly labeled by nick-translation with Cy3-dCTP (Perkin-Elmer), as described elsewhere,[17] with minor modifications. Fosmids WIBR2-3001A12 (hg17 coordinates chr17:32309158–32347025; not duplicated in either case) and WIBR2-946N02 (hg17 coordinates chr17:

33156592–33197819; duplicated in both cases) were used in control experiments.

Results
Fetal Anomalies and Genomic Rearrangements

We evaluated 155 fetuses with one or more congenital anomalies and no known cytogenetic abnormalities in this study (128 with normal karyotype, 12 of whom also had normal FISH results for 22q11 deletion, and 27 of whom had no karyotype information reported). To minimize false-positive results, we focused on microdeletions and microduplications defined by two or more adjacent BAC clones on our targeted array that were not observed as copy-number polymorphisms in >400 control individuals either on our BAC array platform[8,9] or in studies of full-tiling path microarrays.[18,19] We identified nine individuals (6%) with detectable deletions or duplications, eight of which may be pathogenic (table 1). Analysis using whole-genome tiling-path BAC arrays or high-density oligonucleotide arrays allowed refinement of the breakpoints. Three of the alterations are microdeletions with breakpoints in flanking segmental duplications: a 1.8-Mb deletion at 17q12 in a fetus with multicystic renal dysplasia (FA-275 in fig. 1), a 2.5-Mb deletion at 15q25.2 in a fetus with congenital diaphragmatic hernia and mild hydrocephalus (fig. 2b), and a 2-Mb deletion at 16p13.11 in a fetus with posthemorrhagic hydrocephalus, cleft lip, preauricular tags, and two cleft vertebrae (fig. 2c). In each case, the regions flanking the microdeletions are polymorphic in copy number (on the basis of an analysis of 270 unaffected individuals). The smallest rearrangement we detected was a 157-kb duplication (case FA-142 in fig. 2a).

MODY5: A Recurrent 17q12 Microdeletion Disorder

Case FA-275 represents a fetus with bilateral multicystic renal dysplasia, bilateral ureteropelvic junction stenosis, atretic right ureter, and hypoplastic bladder (fig. 1a). We detected a deletion of seven BACs spanning nearly 2 Mb of sequence on our targeted BAC array (fig. 1b). Fine mapping with a customized oligonucleotide array (average probe spacing 53 bp) showed that the deleted region spans 1.8 Mb and involves 19 known genes, including *CCL3L1* (MIM 601395), *CCL4L1* (MIM 603782), and *TBC1D3* (MIM 607741), which are present in multiple copies (fig. 1c). Because of copy-number polymorphism in the flank-

---

Figure 2. Structural resolution of additional rearrangements in the fetal autopsy series. *a,* A 157-kb duplication of chromosome 17q23 in FA-142, a fetus with multicystic dysplastic kidneys. Paired segmental duplications (seg dups) with >99% identity are indicated. Breakpoints of this duplication are in *CLTC* (proximal) and *TMEM49* (distal). *PTHR2* (*BIT1*) is entirely within the duplicated region. *b,* A 2.5-Mb microdeletion of chromosome 15q25.2 (hg17 coordinates chr15:81011700–83540700) in FA-602, a fetus with congenital diaphragmatic hernia and mild hydrocephalus. *c,* A 2-Mb deletion of 16p13.12-p12.3 in FA-180 (hg17 coordinates chr16:14200000–17200000). Regions of known copy-number polymorphism (CNP) in unaffected individuals are indicated (*blue bars*). Plots show the $\log_2$ ratio (*Y*-axis) for probes (*X*-axis) in the region depicted. Deviations of probe $\log_2$ ratios from zero are depicted as in figure 1c.
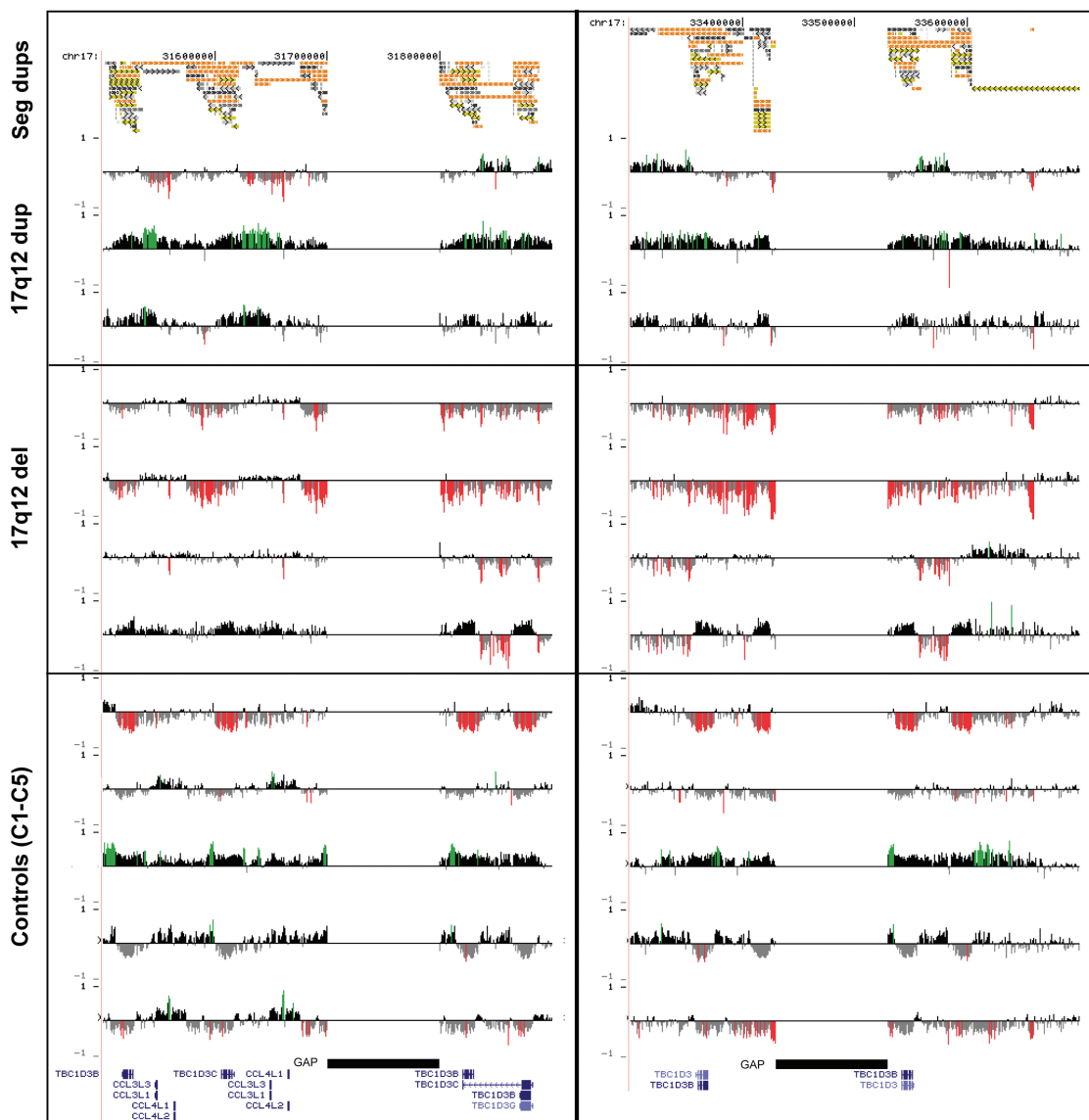

Figure 3. Copy-number variation of the segmental duplications (Seg dups) at breakpoint regions. An expanded view of oligonucleotide array CGH within the segmental duplications located at the proximal (*left*) and distal (*right*) breakpoints of 17q12 rearrangements. Control set includes five unaffected individuals analyzed on same custom array as the patients with 17q12 deletion or duplication. For each individual, deviations of probe log$_2$ ratios from zero are depicted as in figure 1*c*. Control individuals show increased and decreased copy number in the segmental duplications, highlighting the difficulty in determining precise rearrangement breakpoints. Individuals shown are IMR379, 5812, 6498, FR12, FR92, FR09, FR37, C1, C2, C3, C4, and C5.

ing regions (figs. 3 and 4), it is difficult to predict the exact breakpoints of the deletion; however, both proximal and distal breakpoints are within segmental-duplication clusters with multiple regions of high identity—the most significant has >99% identity across 76 kb (hg17 coordinates chr17:31808172–31883983 and chr17:33529232–33604211).

The deleted region in case FA-275 includes the *TCF2* gene (MIM 189907), mutations of which are known to cause MODY5,[20] as well as both pediatric[11,21] and prenatally detectable[10] cystic renal disease. One third of a series

of MODY5-affected individuals with *TCF2* alterations have deletions of the entire *TCF2* gene and surrounding sequence.[12] Similar deletions have been found in pediatric cystic renal disease.[10,11,21] Because the deletion in case FA-275 appeared to overlap the deletions reported in patients with MODY5, we used oligonucleotide arrays to refine the deletion breakpoints in five patients with pediatric renal disease without diabetes and three patients with MODY5 (diabetes and renal disease) who had previously been shown to have deletions encompassing the *TCF2* gene (table 2).

Table 2. Phenotypes Associated with 17q12 Rearrangements

Individual or Case	Del or Dup	Renal Phenotype	Diabetes, Age at Onset (years)	Cognitive Ability	Seizures	Other	De Novo or Inherited
FA-275	Del	MCDK, prenatal onset	NA	NA	NA	...	Unknown
FR40	Del	Single kidney with isolated cysts, onset at age 7 years	Yes, 22	Normal	None	Cryptorchidism, pancreatic atrophy, and elevated liver enzymes (2.5 ×)	De novo
FR92	Del	Bilateral cysts and hypoplasia, onset at age 1 year	Yes, 10	Normal	None	Elevated liver enzymes (2 ×)	De novo
FR99	Del	Abnormal renal function, onset at age 27 years	Yes, 23	Normal	None	Elevated liver enzymes (4 ×)	De novo
FR45	Del	Bilateral MCDK, prenatal onset; GFR < 80	No[a]	Normal	None	...	Inherited
FR12	Del	Prenatal hyperechogenic kidneys and postnatal MCDK; GFR=95	No	Normal	None	...	De novo
FR37	Del	Prenatal hyperechogenic kidneys and postnatal MCDK, pelvic dilatation, and hyperuricemia	No	Normal	None	...	De novo
FR35	Del	Prenatal hyperechogenic kidneys and bilateral MCDK; GFR=96	No	Normal	None	...	De novo
FR09	Del	Bilateral MCDK, prenatal onset	No	Normal	None	...	De novo
IMR379	Dup	None known	No	MR	Unknown	...	Unknown[b]
5812	Dup	None known	No	MR	Focal complex	...	Inherited
6498	Dup[c]	None known	No	MR	Focal complex	...	Inherited[d]

NOTE.—Del = deletion; Dup = duplication; GFR = glomerular filtration rate (in mL/min/1.73 m^2); MCDK = multicystic dysplastic kidneys; MR = mental retardation.

[a] The father of FR45 has diabetes (onset at age 27 years) and renal cysts, and the brother of FR45 also has MCDK (prenatal onset).

[b] Presumed inheritance from the father, who was unavailable for analysis.

[c] Duplication of TCF2 and LHX1 genes (see text).

[d] TCF2 duplication inherited from the unaffected mother; duplication of LHX1 is likely also inherited from the mother, who is mosaic (see text).
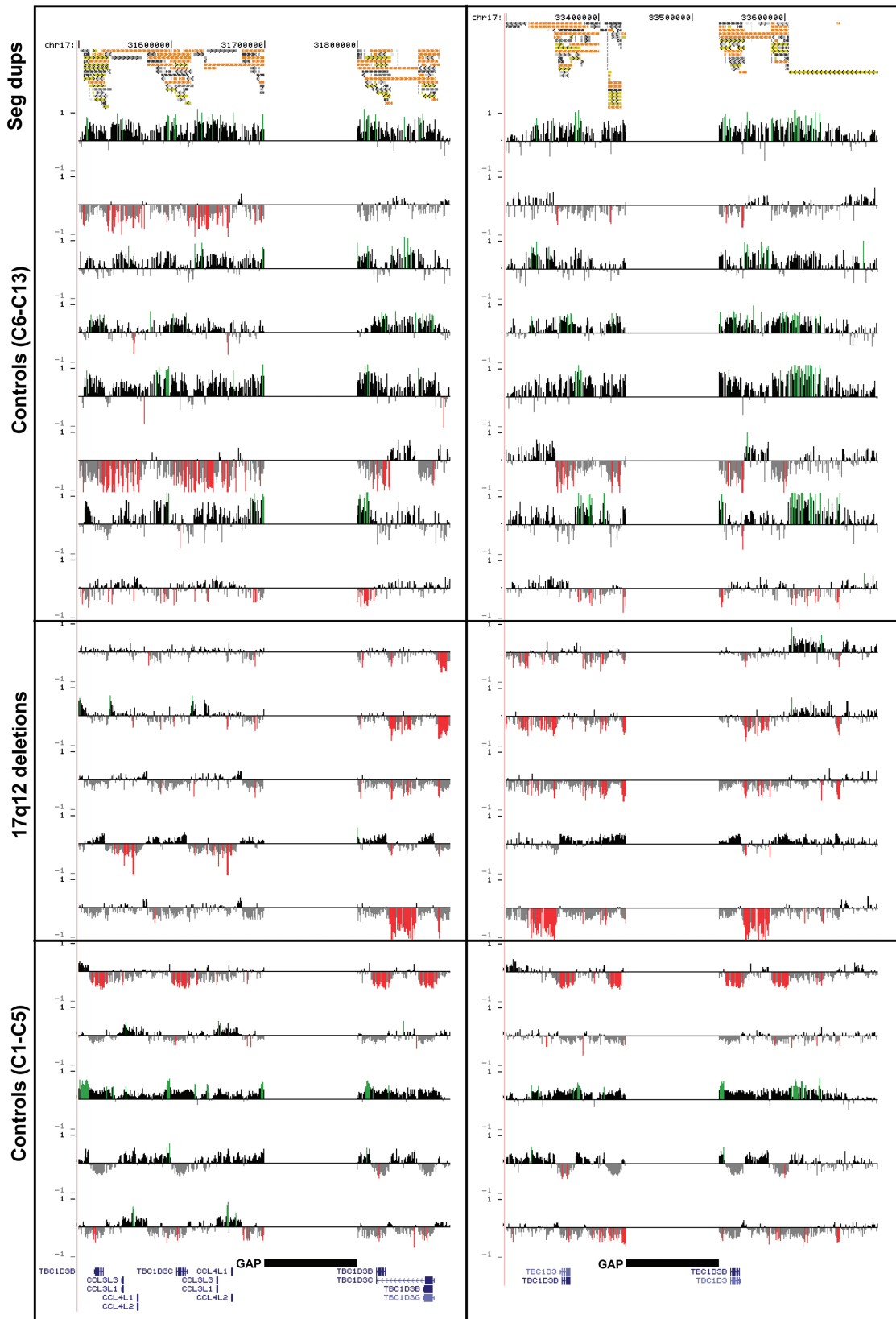
Figure 4. Copy-number polymorphism of the segmental duplications at the breakpoint regions. An expanded view of figure 2, showing eight additional control individuals (controls C6–C13) analyzed on a custom oligonucleotide array targeted to segmental duplications (Seg dups) and five additional patients with 17q12 deletion. Controls C1–C5 are the same as in figure 2. For each individual, deviations of probe log$_2$ ratios from zero are depicted as in figures 1 and 2. The patients with 17q12 deletion shown are FR40, FR99, FR45, FR35, and FA-275.
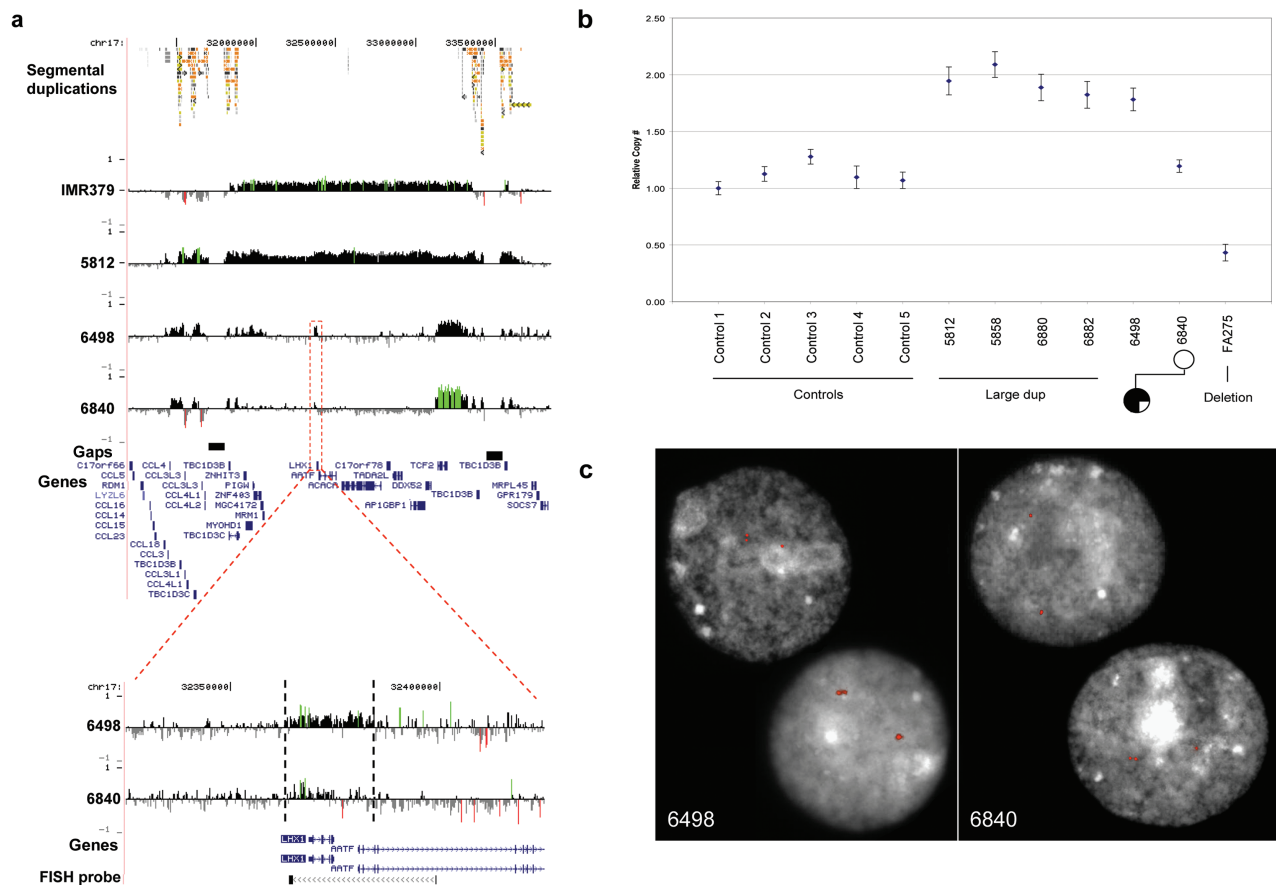
Figure 5. Reciprocal 17q12 duplication. *a*, Structural resolution of the reciprocal 1.5-Mb microduplication region of 17q12 by use of high-density oligonucleotide microarrays. Plots show the $\log_2$ ratio (*Y*-axis) for 48,256 probes (*X*-axis) in this region of 17q12 (hg17 coordinates chr17:31200000–33850000). Three affected individuals with mental retardation and epilepsy and the unaffected mother of 6498 are shown. An expanded view of the *LHX1* region (hg17 coordinates chr17:32325000–32345000) is also shown for individuals 6498 and 6840. The average $\log_2$ ratio for the 765 probes in the 27-kb region (*dotted lines*) is 0.176 for individual 6498 and 0.030 for individual 6840 (the mother). *c*, Quantitative PCR results for exon 1 of *LHX1* gene, confirming four patients with duplication (large dup) and patient FA-275 with deletion, when compared with five unaffected controls. Patient 6498 with epilepsy shows a duplication by quantitative PCR when compared with the mother (6840). *d*, Interphase FISH with *LHX1* genomic probe (WIBR2-6022g13; see panel b for location). Of the lymphocyte nuclei from affected patient 6498, 98% show evidence of tandem duplication (three signals), compared with 28% of lymphoblast cells from the mother (6840). Two nuclei from 6840 show discordant results, with duplication present in only one of the two cells shown, suggesting that the mother is a mosaic for the duplication (see the slight signal-intensity increase observed in panel a).

Our results show that four of the five pediatric patients and all three of the patients with MODY5 have deletions that are nearly identical to that of our fetal case, with breakpoints in all cases mapping to flanking segmental-duplication blocks (fig. 1*c*). Case FR09 provides the greatest breakpoint resolution, with a proximal breakpoint at 31,835,000 ± 1 kb and a distal breakpoint at 33,357,000 ± 1 kb defining the common minimal deletion region. The extensive and variable copy-number polymorphism in the flanking regions (figs. 3 and 4) and the presence of a sequence assembly gap make it difficult to define the breakpoints more precisely in the other three cases, but they all occur within the large segmental duplication blocks (proximal: 31,500,000–31,900,000 bp; distal:

33,300,000–33,600,000 bp). One case (FR35) is atypical, with a larger de novo deletion at least 2.1 Mb in size. The distal breakpoint is in unique sequence ~33 kb distal to the *TCF2* gene (hg17 coordinate ~33212500), suggesting this deletion arose from a mechanism other than NAHR. These results show that the 17q12 deletion is a recurrent genomic disorder with breakpoints in flanking segmental duplications that results in renal disease and diabetes.

Reciprocal 17q12 Duplication

We previously identified a duplication of the 17q12 region in a patient with idiopathic mental retardation (IMR379) and an affected sibling.[7] We have identified two additional
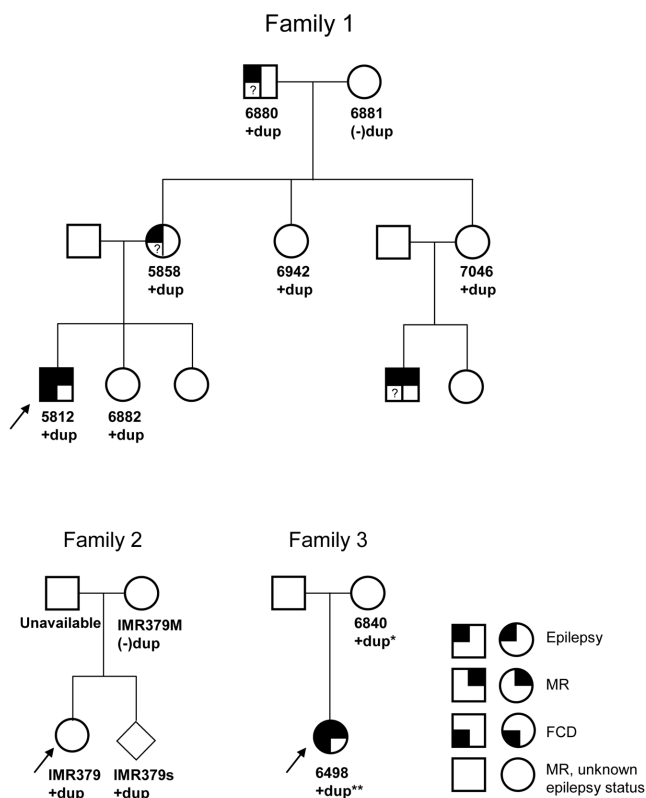
Figure 6. Pedigrees of three families with duplication in the 17q12 region. +dup = duplication is reciprocal to 17q12 deletion; +dup* = duplication of *TCF2,* mosaic duplication of *LHX1*; +dup** = duplication of *TCF2* and *LHX1* (see text); FCD = focal cortical dysplasia; MR = mental retardation. Only individuals 5812 and 6498 have had magnetic resonance imaging to evaluate for FCD.

patients with mild-to-moderate mental retardation, epilepsy, and focal cortical dysplasia who have overlapping duplications on 17q12 by BAC array CGH (table 2). Oligonucleotide microarray CGH analysis of these three individuals shows that two individuals, IMR379 and 5812, have duplications of the entire 17q12 region, with breakpoints mapping to the same duplication blocks associated with renal disease and MODY5 deletions (fig. 5a). These duplication events, therefore, represent the expected reciprocal duplication to the common 17q12 deletion and provide further evidence that these rearrangements are mediated by NAHR.[22] Pedigree analysis of these families shows that dup17q12 can be stably transmitted meiotically (fig. 6). The duplication is present in all affected individuals in family 1 who we were able to test. However, apparently unaffected individuals in the family were also found to carry the duplication, so the pathological significance is unclear. It is possible that the duplication is a benign copy-number variant segregating with disease in this family. Alternatively, there may be incomplete penetrance of the seizure phenotype.

To assess the frequency of the 17q12 duplication in the general population, we analyzed a larger control group for copy-number variation of this locus ($n = 960$ unrelated white individuals who had been genotyped with a high-density SNP Genotyping BeadChip for variable efficacy of statin-drug response by cardiovascular disease [R. Krauss and D. Nickerson, personal communication]). We identified a single individual with a duplication corresponding to the 1.5-Mb region (fig. 7), whereas no individuals with the deletion were observed. Combined with our previous analysis of 270 HapMap individuals, this suggests a slight enrichment of the 17q12 duplication among patients with mental retardation and/or epilepsy (2 of 380 affected individuals; 1 of 1,276 control individuals) but does not reach statistical significance ($P = .13$, by Fisher's exact test).

Interestingly, the third patient (6498), who has mental retardation, focal cortical dysplasia, and focal seizures, has a smaller and apparently more complex duplication in which there is a 27-kb duplication encompassing the *LHX1* gene (MIM 601999; hg17 coordinates chr17: 32360000–32387000) and a 259-kb duplication corresponding to the *TCF2* gene (hg17 coordinates chr17: 33122000–33381000) (fig. 5a). The mother of patient 6498 is unaffected and also carries a duplication of the *TCF2* gene. However, both oligonucleotide array CGH and real-time quantitative PCR of the 27-kb region encompassing *LHX1* suggested that the mother does not carry the *LHX1* duplication (fig. 5). Interphase FISH with peripheral blood lymphocytes from both the mother and daughter showed that the mother is mosaic for the duplication (28% of 135 nuclei), whereas the daughter has the *LHX1* duplication in virtually all cells (98% of 138 nuclei) (fig. 5c).

Discussion

We have used a targeted approach to identify novel genomic disorders in individuals with congenital anomalies. Our results emphasize the importance of evaluating de novo structural-variation events in pediatric diseases other than mental retardation and the importance of duplication architecture as a predisposing factor for disease. We show that recurrent microdeletions of chromosome 17q12, which cause pediatric renal disease and MODY5, are mediated by flanking segmental duplications. The mechanism of recurrent deletion by NAHR explains the high rate of de novo *TCF2* deletions detected in previous studies of pediatric renal disease and MODY5. Indeed, it has been estimated that de novo structural rearrangements may occur at a frequency of between 10^{-6} and 10^{-4} (100–10,000 times more frequently than de novo point mutations).[23] The 17q12 deletion is the first genomic disorder that results in diabetes and is the third genomic disorder with significant renal involvement (the other two are juvenile nephronophthisis and polycystic kidney disease). This recurrent microdeletion will have a significant impact on diagnosis, prognosis, and management of renal
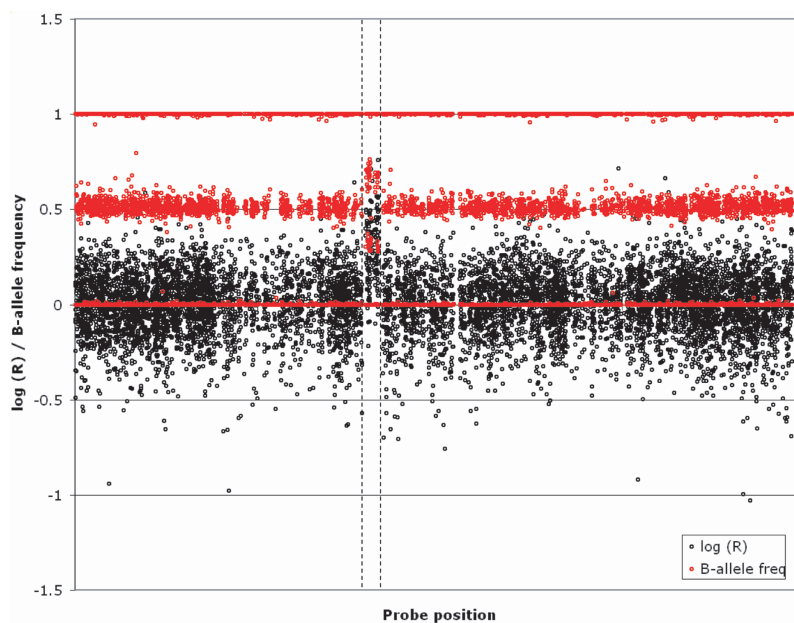
Figure 7. A 17q12 duplication identified in 1 of 960 unaffected control samples by use of the HumanHap300 Genotyping BeadChip (Illumina). Data show probe position (*X*-axis) against the log(R) intensity ratio (*black*) and the B-allele frequency (*red*) for probes on chromosome 17. Within the limits of resolution of these data, this duplication appears to be identical to that observed in patients IMR379 and 5812.

disease and early-onset type II diabetes in children, and we recommend that evaluation of this microdeletion be considered early in the diagnostic workup for children with renal pathology.

The 17q12 deletion results in a wide range of phenotypes with considerable variability in expressivity. Individuals with nearly identical deletions can have severe renal disease detectable prenatally, renal disease detected in childhood, or MODY5 (usually diagnosed before the 4th decade). Previous studies suggest that, although the deletion is found in individuals at both ends of the phenotypic spectrum, those with earlier-onset disease are more likely to have a large genomic deletion than a point mutation in *TCF2*. Of those individuals with *TCF2* alterations in three different series, 36% of individuals who presented with a MODY5 phenotype,[12] 64% of patients ascertained in childhood with renal abnormalities,[11] and 89% of patients with renal disease detected by prenatal ultrasound (and who have not yet developed diabetes)[10] had genomic deletions versus point mutations of *TCF2*. These results suggest that the microdeletion is associated with earlier onset of renal pathology. One of the genes in the 17q12 deleted region is *LHX1,* a limb homeodomain gene important for renal development in the mouse.[24,25] It may be that haploinsufficiency of both *TCF2* and *LHX1*—two genes important for renal development—influence earlier onset of renal disease than do point mutations in *TCF2.* However, some individuals with the deletion do have a milder phenotype, and individuals with a point mutation in *TCF2* may have early-onset disease;

this suggests that *TCF2* is the major gene contributing to the phenotype in patients with deletion. In addition, there may be genetic modifiers outside the 17q12 deleted region that contribute to the phenotype.

Interestingly, we have also identified the reciprocal duplication of 17q12, which may be associated with epilepsy—a strikingly different phenotype than that associated with the deletion. A smaller, more complex duplication involving the *TCF2* and *LHX1* genes was found in another patient with focal cortical dysplasia and epilepsy, which suggests that one or both of these genes in the 17q12 region may be important for epilepsy and brain development. Although the known biology of *LHX1* makes it a good candidate gene for abnormal brain development (i.e., the gene is expressed in human brain, it is required for migration of motor axons to the limb,[26] and knockout mice have anencephaly[25]), the association of the duplication with disease is less clear than that of the deletion phenotype. Screening of 1,276 unaffected control subjects (see the "Material and Methods" section) found a single occurrence of this 1.5-Mb duplication (but not the reciprocal deletion), suggesting that this variant is rare in the unaffected population and is enriched in patients with mental retardation and epilepsy. One possible explanation is incomplete penetrance of the duplication-associated phenotype. Incomplete penetrance and variable expressivity have been described for several genomic disorders, including 22q11 deletion[27] and duplication[28,29] syndromes. Additional studies will be required before a definitive association of the variant with disease can be stated, as op-

posed to a rare copy-number variant that fortuitously segregates in some families with affected individuals.

Finally, we find that the segmental duplications located at the breakpoints are polymorphic in copy number and structure among unaffected individuals. Similar variation has been noted in previous studies of genomic disorders. We propose that this variation may contribute to the susceptibility of individual genomes to rearrangement, wherein increased copy number in the flanking segmental duplications increases the likelihood of NAHR. Evaluation of parents in whom the deletion or duplication arose will help determine the contribution of flanking polymorphism to genomic rearrangement.

In summary, we have found novel microdeletions and duplications in a series of fetal samples with congenital anomalies other than mental retardation. Combined with recent studies of similar genetic variation in children in mental retardation and autism,[7,30,31] we put forward the structural-variant disease hypothesis: large (>100-kb) structural genetic changes are not uncommon in germ cells and are more likely to be *recurrent,* under strong *selective constraint,* and distributed *nonrandomly* in the genome. Therefore, de novo events will have a significant impact on rare and common diseases in the population that cannot be tracked by traditional linkage-disequilibrium association mapping approaches. We advocate generalized screening of genomic "hotspot" regions of both parents and offspring for other pediatric diseases for which the genetic etiology is not well understood (including schizophrenia, asthma, and cardiovascular disease).

Web Resources

The URLs for data presented herein are as follows:

Human Genome Structural Variation Project, http://humanparalogy.gs.washington.edu/structuralvariation/ (for oligonucleotide array data)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for MODY5, *CCL3L1, CCL4L1, TBC1D3, TCF2,* and *LHX1*)

PharmGKB: PARC Profile, http://www.pharmgkb.org/network/members/parc.jsp#team

References

1. Shaw CJ, Lupski JR (2004) Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet Spec 1 13:R57–R64
2. de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM, Reijmersdal S, Nillesen WM, Huys EH, Leeuw N, et al (2005) Diagnostic genome profiling in mental retardation. Am J Hum Genet 77:606–616
3. Friedman JM, Baross A, Delaney AD, Ally A, Arbour L, Armstrong L, Asano J, Bailey DK, Barber S, Birch P, et al (2006) Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. Am J Hum Genet 79:500–513
4. Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R, Kooy RF, Reyniers E, Romano C, Fichera M, et al (2006) A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nat Genet 38:999–1001
5. Menten B, Maas N, Thienpont B, Buysse K, Vandesompele J, Melotte C, de Ravel T, Van Vooren S, Balikova I, Backx L, et al (2006) Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports. J Med Genet 43:625–633
6. Rosenberg C, Knijnenburg J, Bakker E, Vianna-Morgante AM, Sloos W, Otto PA, Kriek M, Hansson K, Krepischi-Santos AC, Fiegler H, et al (2006) 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 43:180–186
7. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC, et al (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet 38:1038–1042
8. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM, et al (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am J Hum Genet 79:275–290
9. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, et al (2005) Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77:78–88
10. Decramer S, Parant O, Beaufils S, Clauin S, Guillou C, Kessler S, Aziza J, Bandin F, Schanstra JP, Bellanne-Chantelot C (2007) Anomalies of the TCF2 gene are the main cause of fetal bilateral hyperechogenic kidneys. J Am Soc Nephrol 18:923–933
11. Ulinski T, Lescure S, Beaufils S, Guigonis V, Decramer S, Morin D, Clauin S, Deschenes G, Bouissou F, Bensman A, et al (2006) Renal phenotypes related to hepatocyte nuclear factor-1β (TCF2) mutations in a pediatric cohort. J Am Soc Nephrol 17:497–503
12. Bellanne-Chantelot C, Clauin S, Chauveau D, Collin P, Daumont M, Douillard C, Dubois-Laforgue D, Dusselier L, Gautier JF, Jadoul M, et al (2005) Large genomic rearrangements in the hepatocyte nuclear factor-1β (TCF2) gene are the most frequent cause of maturity-onset diabetes of the young type 5. Diabetes 54:3126–3132
13. Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia

F, Haden K, Li J, Shaw CA, Belmont J, et al (2006) High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Res 16: 1136–1148

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

15. Selzer RR, Richmond TA, Pofahl NJ, Green RD, Eis PS, Nair P, Brothman AR, Stallings RL (2005) Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. Genes Chromosomes Cancer 44:305–319

16. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365–386

17. Lichter P, Tang CJ, Call K, Hermanson G, Evans GA, Housman D, Ward DC (1990) High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. Science 247:64–69

18. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al (2006) Global variation in copy number in the human genome. Nature 444:444–454

19. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, et al (2007) A comprehensive analysis of common copy-number variations in the human genome. Am J Hum Genet 80:91–104

20. Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, Cockburn BN, Lindner T, Yamagata K, Ogata M, Tomonaga O, et al (1997) Mutation in hepatocyte nuclear factor-1β gene (TCF2) associated with MODY. Nat Genet 17:384–385

21. Muller D, Klopocki E, Neumann LM, Mundlos S, Taupitz M, Schulze I, Ropers HH, Querfeld U, Ullmann R (2006) A complex phenotype with cystic renal disease. Kidney Int 70:1656–1660

22. Stankiewicz P, Lupski JR (2002) Genome architecture, rearrangements and genomic disorders. Trends Genet 18:74–82

23. Lupski JR (2007) Genomic rearrangements and sporadic disease. Nat Genet 39:S43–S47

24. Pedersen A, Skjong C, Shawlot W (2005) Lim1 is required for nephric duct extension and ureteric bud morphogenesis. Dev Biol 288:571–581

25. Shawlot W, Behringer RR (1995) Requirement for Lim1 in head-organizer function. Nature 374:425–430

26. Kania A, Johnson RL, Jessell TM (2000) Coordinate roles for LIM homeobox genes in directing the dorsoventral trajectory of motor axons in the vertebrate limb. Cell 102:161–173

27. Robin NH, Shprintzen RJ (2005) Defining the clinical spectrum of deletion 22q11.2. J Pediatr 147:90–96

28. Ensenauer RE, Adeyinka A, Flynn HC, Michels VV, Lindor NM, Dawson DB, Thorland EC, Lorentz CP, Goldstein JL, McDonald MT, et al (2003) Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients. Am J Hum Genet 73:1027–1040

29. Yobb TM, Somerville MJ, Willatt L, Firth HV, Harrison K, MacKenzie J, Gallo N, Morrow BE, Shaffer LG, Babcock M, et al (2005) Microduplication and triplication of 22q11.2: a highly variable syndrome. Am J Hum Genet 76:865–876

30. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, et al (2007) Strong association of de novo copy number mutations with autism. Science 316:445–449

31. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L, et al (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 39:319–328