nature genetics

Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome

Andrew J Sharp[1], Sierra Hansen[1], Rebecca R Selzer[2], Ze Cheng[1], Regina Regan[3], Jane A Hurst[4], Helen Stewart[4], Sue M Price[4], Edward Blair[4], Raoul C Hennekam[5,6], Carrie A Fitzpatrick[7], Rick Segraves[8], Todd A Richmond[2], Cheryl Guiver[3], Donna G Albertson[8,9], Daniel Pinkel[8], Peggy S Eis[2], Stuart Schwartz[7], Samantha J L Knight[3] & Evan E Eichler[1]

Genomic disorders are characterized by the presence of flanking segmental duplications that predispose these regions to recurrent rearrangement. Based on the duplication architecture of the genome, we investigated 130 regions that we hypothesized as candidates for previously undescribed genomic disorders[1]. We tested 290 individuals with mental retardation by BAC array comparative genomic hybridization and identified 16 pathogenic rearrangements, including *de novo* microdeletions of 17q21.31 found in four individuals. Using oligonucleotide arrays, we refined the breakpoints of this microdeletion, defining a 478-kb critical region containing six genes that were deleted in all four individuals. We mapped the breakpoints of this deletion and of four other pathogenic rearrangements in 1q21.1, 15q13, 15q24 and 17q12 to flanking segmental duplications, suggesting that these are also sites of recurrent rearrangement. In common with the 17q21.31 deletion, these breakpoint regions are sites of copy number polymorphism in controls, indicating that these may be inherently unstable genomic regions.

Approximately 5% of the human genome is composed of segmental duplications that are >1 kb in length and that show >90% sequence identity, the majority of which have an interspersed, rather than tandem, distribution[1,2]. These duplication blocks act as substrates for nonallelic homologous recombination (NAHR), leading to the deletion, duplication or inversion of the intervening sequence[3]. Segmental duplications are frequently associated with structural polymorphisms[4–8] and an increasing number of recurrent pathogenic rearrangements, termed genomic disorders[9]. Typically, these recurrent rearrangements are mediated by paired intrachromosomal duplications (or low-copy repeats) that are ≥10 kb in length, show ≥95% sequence identity and are separated by 50 kb–10 Mb of intervening sequence[3]. Based on these criteria, we created a map of potential 'rearrangement hotspots' in the human genome that, because of their flanking duplication architecture, we hypothesized would be liable to genomic instability[1]. We identified 130 such sites spanning ~274 Mb of nonredundant sequence (**Supplementary Table 1** online), and we constructed a segmental duplication BAC microarray targeted specifically to these regions[7].

In order to determine a baseline level of polymorphism, we first used our segmental duplication BAC array to investigate copy number polymorphism (CNP) at these rearrangement hotspots in a control population of 316 individuals, identifying 384 putative sites of copy number polymorphism[7,10] (**Supplementary Table 2** online). We then used the segmental duplication array to analyze a test population from the UK comprising 290 children and young adults with idiopathic mental retardation with or without associated dysmorphism or congenital anomalies. Approximately 50% of individuals were referred from clinical genetics centers, ~35% from community learning disability teams and ~15% from other sources, such as by referral from hospital neuropediatricians. All individuals had been reported to have normal G-banded karyotypes at 550-band resolution, and most had been reported negative for *FRAXA* mutations. All except 35 had been tested for cryptic subtelomeric rearrangements using FISH and reported as normal[11], and in some individuals, other specific genetic abnormalities had also been excluded. None had been previously analyzed by array comparative genomic hybridization (CGH). Parents or guardians of all subjects provided informed consent, and the protocol was reviewed and approved by the institutional review boards of Case Western Reserve University and the University of Washington as well as by ethical review committees in the UK, where the samples originated.

[1]Department of Genome Sciences and The Howard Hughes Medical Institute, University of Washington School of Medicine, 1705 NE Pacific St., Seattle, Washington 98195, USA. [2]NimbleGen Systems, Inc., Madison, Wisconsin 53711, USA. [3]Oxford Genetics Knowledge Park, The Wellcome Trust Centre for Human Genetics, Churchill Hospital, Oxford OX3 7BN, UK. [4]Department of Clinical Genetics, Oxford Radcliffe Hospitals National Health Service (NHS) Trust, Churchill Hospital, Oxford OX3 7LJ, UK. [5]Clinical and Molecular Genetics Unit, Institute of Child Health, University College London, London, UK. [6]Department of Pediatrics, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands. [7]Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. [8]Comprehensive Cancer Center, University of California San Francisco (UCSF), San Francisco, California 94143, USA. [9]Cancer Research Institute, UCSF, San Francisco, California 94143, USA. Correspondence should be addressed to E.E.E. (eee@gs.washington.edu).

Received 15 May; accepted 20 July; published online 13 August 2006; doi:10.1038/ng1862

Table 1 Summary of abnormalities identified using the segmental duplication BAC array in 290 individuals with idiopathic mental retardation

Sample	Rearrangement	Size	Number of variant BACs	Inheritance	Secondary verification
IMR103[a]	del 17q21.31	740 kb*	4	*De novo*	FISH, microsatellites, oligo array
IMR253[a]	del 17q21.31	740 kb*	4	NA	FISH, oligo array
IMR255[a]	del 17q21.31	740 kb*	4	NA	FISH, oligo array
IMR376[a]	del 17q21.31	740 kb*	4	NA	FISH, oligo array
IMR379[b]	dup 17q12	1.46 Mb*	6	Duplication present in affected sibling, absent in normal mother	FISH, oligo array
IMR349[c]	del 15q24.1–q24.3	3.90 Mb*	9	*De novo*	FISH, oligo array
IMR43[d]	del 1q21.1	1.47 Mb*	9	*De novo*	Oligo array
IMR338[e]	del 15q13.1–13.3	3.95 Mb*	23	Deletion present in affected mother and sibling, absent in normal sibling	FISH, oligo array
IMR371[f]	del 15q23–24.2	3.81 Mb*	9	*De novo*	FISH, oligo array
IMR26[g]	del 22q11.2	2.4 Mb	9	*De novo*	FISH
IMR192[g]	dup 22q11.2	3.0 Mb	13	Duplication present in affected father	FISH
IMR146[h]	dup 17p11.2	3.8 Mb	17	*De novo*	FISH
IMR149[i]	del 1p36.3	1.4 Mb	9	NA	Not done
IMR184	dup 16p13.11–p13.2	2.1 Mb	9	NA	FISH

Rearrangement sizes are taken from the maximum extent of variant BACs on the segmental duplication array, except in cases in which further refinement was performed using oligonucleotide (oligo) arrays (*). Inheritance of each abnormality was performed by testing parental and sibling DNA using the segmental duplication array, except in cases in which no additional family members were available (NA). FISH validation data are shown in **Supplementary Figure 1**. We detected two additional individuals carrying probable unbalanced translocations (**Supplementary Table 3**). All individuals presented with mental retardation with or without dysmorphism or congenital abnormalities. Additional phenotype data are shown in **Supplementary Table 4**.

[a]Four identical deletions with paired segmental duplications (38 kb, 98% similarity) at breakpoints. [b]Paired segmental duplications (66 kb, 99.7% similarity) at breakpoints. [c]Paired segmental duplications (51 kb, 95% similarity) at breakpoints; deletion has similar proximal breakpoint to case 15 in ref. 16. [d]Both breakpoints embedded in large clusters of segmental duplication; deletion is similar to case 11 in ref. 16. [e]Both breakpoints embedded in large clusters of segmental duplication; proximal breakpoint occurs in common Prader-Willi syndrome/Angelman syndrome (PWS/AS) BP3 region[17]. [f]Both breakpoints located in unique sequence. [g]DiGeorge/velocardiofacial syndrome region[12]. [h]Smith-Magenis syndrome region[13]. [i]Monosomy 1p36 region[14].

In order to minimize false positives, we focused on copy number variations not observed in the control population that were defined by two or more adjacent BAC clones on our segmental duplication array. A complete list of all BACs identified as potentially copy-number variant in the 290 individuals is shown in **Supplementary Table 2**. After excluding copy number changes that either (i) had been observed as polymorphisms in control individuals, (ii) did not segregate with the phenotype in each pedigree or (iii) could not be verified by an independent technique such as FISH, we identified 16 individuals (5.5%) with microdeletions or duplications that are likely to be pathogenic (**Table 1**). A list of seven additional variations of uncertain significance is shown in **Supplementary Table 3** online. The changes considered to be pathogenic included two unbalanced translocations and several corresponding to known genomic disorders[12–14]. In addition, we identified six other rearrangements that might represent previously unidentified genomic disorders in the human population.

Notably, four individuals in our test population showed deletion of the same four contiguous BACs spanning ~500 kb in 17q21.31, a region of known polymorphic inversion[15] (**Fig. 1a**). We confirmed a hemizygous deletion in all four individuals by FISH (**Fig. 1b** and **Supplementary Fig. 1** online), and in the one individual for whom parental DNA was available, BAC and oligonucleotide array analysis

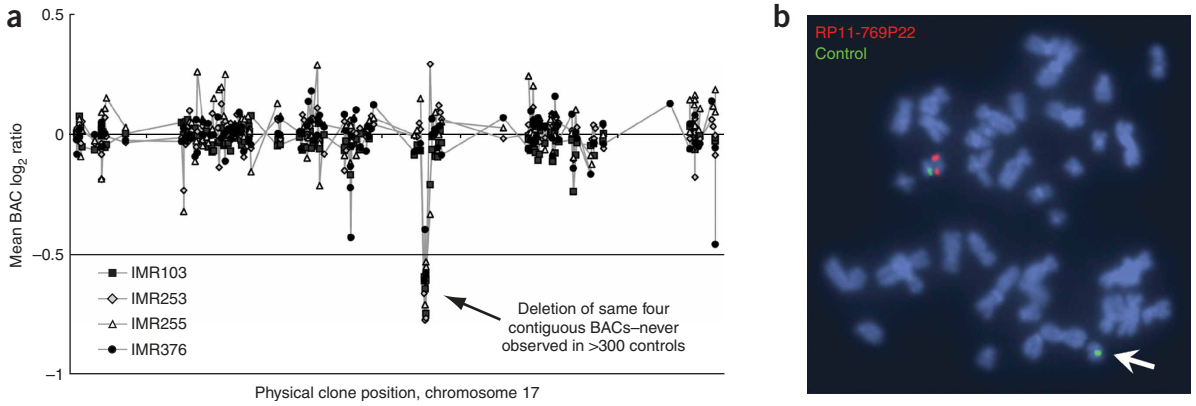

Figure 1 Detection of a previously unidentified recurrent microdeletion in 17q21.31 by targeted array CGH. (**a**) Chromosome 17 data from the segmental duplication BAC array. Four unrelated individuals with idiopathic mental retardation all show consistently reduced $\log_2$ ratios of the same four contiguous BACs spanning ~500 kb in 17q21.31, which we did not observe as a polymorphism in control individuals (**Supplementary Table 2**). Other hybridization signals in these four cases correspond to known copy number polymorphisms (**Supplementary Table 2**). (**b**) FISH using one of the four BACs that showed a deletion on the segmental duplication array (RP11-769P22, red) and using a control probe located on 17p (RP11-646F1, green) confirms a heterozygous 17q21.31 deletion in IMR376. We obtained similar results in the three other cases identified. In IMR103, ~10% of peripheral blood lymphocytes analyzed by FISH showed no evidence for deletion of RP11-769P22, suggesting somatic mosaicism in this case. Similar somatic mosaicism is occasionally observed in other genomic disorders[27,28].

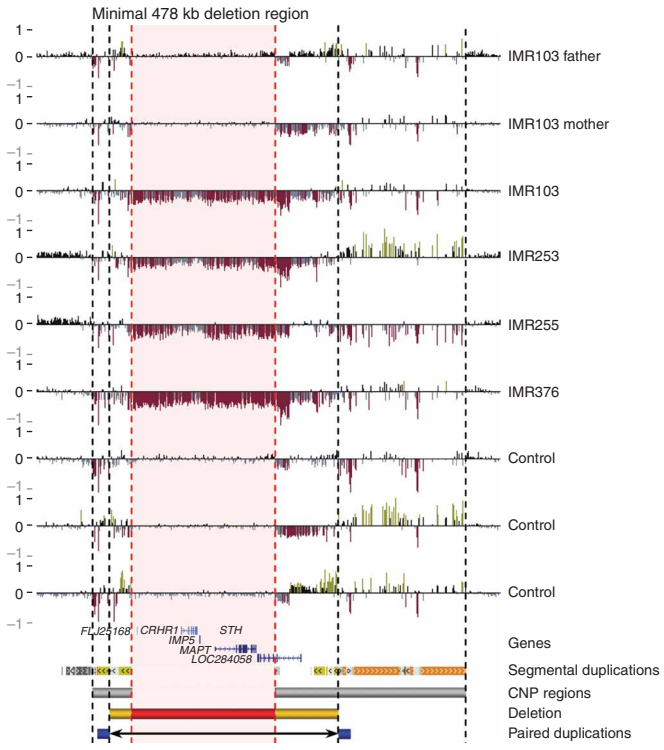

Figure 2 Structural resolution of a 1.5-Mb region of 17q21.31 (human genome May 2004 assembly (hg17), chr17:40,750,000–42,250,000) using high-density oligonucleotide arrays. Plots show the $\log_2$ ratio (*y* axis) for 6,818 probes (*x* axis) in the initial four microdeletion cases ascertained by the segmental duplication BAC array and five control individuals. Consistent with our BAC array data and previous reports[7,10,15], this region contains multiple copy number polymorphisms flanking the microdeletion region (gray bars), with at least six different CNP regions discernable flanking the deletion region. Although these complicate definition of the deletion breakpoints, a pair of segmental duplications (length 38 kb, similarity 98%, separation 740 kb) (blue bars) flank a region that is deleted in all four cases (gold and red bar). We hypothesize that these two segmental duplications are likely to mediate this recurrent rearrangement. Examination of this deletion region in normal individuals defines a 478-kb critical region that is deleted in all four affected individuals but that is invariant in controls (hg17, chr17:41042028–41520694) (red bar and pink shading). For each individual, deviations of probe $\log_2$ ratios from 0 are depicted by gray and black lines; those exceeding a threshold of 1.5 s.d. from the mean probe ratio are colored green and red to represent relative gains and losses, respectively. For clarity, only RefSeq genes within the minimal deletion region are shown.

showed that the proband's deletion was *de novo*. The confirmation of a *de novo* event, together with marked phenotypic similarities in these cases (**Supplementary Table 4** online), strongly suggests that this deletion is pathogenic and that this represents a previously undescribed recurrent microdeletion syndrome.

We characterized these four rearrangements in more detail using a high-density oligonucleotide array (mean density, 1 probe per 131 bp). This demonstrated that the deletion was clearly a recurrent event; in all cases, the breakpoints mapped to large clusters of flanking segmental duplication (**Fig. 2**). In addition to the pathogenic deletion, considerable structural polymorphism was also present flanking this locus, with at least six separate sites of copy number variation identified in controls. Although this complicated the precise determination of the deletion breakpoints, in all four cases the boundaries of the deleted segment seemed to coincide with a pair of directly oriented segmental duplications 38 kb in length with 98% similarity and separated by 740 kb, which we hypothesize mediated these rearrangements by NAHR. By comparison with parental and other unaffected samples, we defined a minimal 478-kb critical region that was recurrently deleted in all four test individuals but invariant in controls. This region contains six known genes, including *CRHR1* (OMIM #122561) and *MAPT* (OMIM #157140). Both of these are highly expressed in brain and have been implicated in several neurodegenerative and behavioral phenotypes and therefore represent excellent candidates for dosage-sensitive genes underlying this 17q21.31 deletion syndrome. On the basis of shared phenotypic features to the initial four deletion carriers (a full phenotype checklist is shown in **Supplementary Table 4**), we subsequently identified a fifth individual (**Fig. 3**) by array CGH (Case 338H5) who also showed a *de novo* deletion of the same four BAC clones in 17q21.31, confirming this as an identifiable syndrome.

In addition to this recurrent 17q21.31 deletion, we also investigated five other abnormalities in more detail using oligonucleotide arrays (**Fig. 4** and **Table 1**). In four of these five cases, exactly as observed for the 17q21.31 deletion, the breakpoints localized to large clusters of flanking segmental duplications (clusters contain a mean of 181 pairwise alignments spanning 712 kb; **Supplementary Table 5** online), strongly suggesting that these rearrangements were catalyzed by the duplication architecture of the genome. In two cases, the sequences present at both breakpoints showed direct pairwise homology and orientation with one another, consistent with the classical model of recurrent genomic disorders caused by NAHR between flanking repeats[3]. Furthermore, the deletions we identified at 1q21.1, 15q13.1–13.3 and 15q24.1–q24.3 have one or both breakpoints that are apparently very similar to rearrangements described previously[16,17]. Thus, even though we observed each abnormality only once in our study population, we hypothesize that these four loci are also likely to be sites of recurrent rearrangement.

Of note, the breakpoints of the 15q24.1–q24.3 deletion were located within a pair of 51-kb segmental duplications with 95% similarity, and both breakpoints of the 17q12 duplication mapped to a pair of 66-kb duplications with 99.7% similarity. For both the 1q21.1 and 15q13.1–13.3 deletions, although breakpoints in each case were located within large blocks of intense segmental duplication (clusters contain a mean of 335 pairwise alignments, spanning 1.32 Mb), inspection of the reference sequence did not uncover obvious pairwise homology

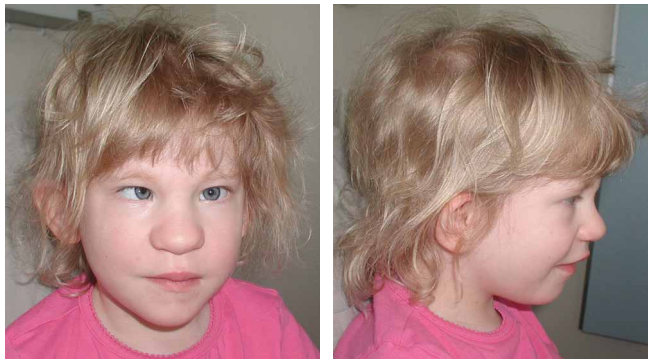

Figure 3 Photographs of individual 338H5 (del 17q21.31), age 4 years. This individual was ascertained on the basis of shared phenotypic features to the initial four 17q21.31 deletion carriers (**Supplementary Table 4**). Silvery depigmentation of the hair, blue eyes and a bulbous nose were consistent observations in all cases.

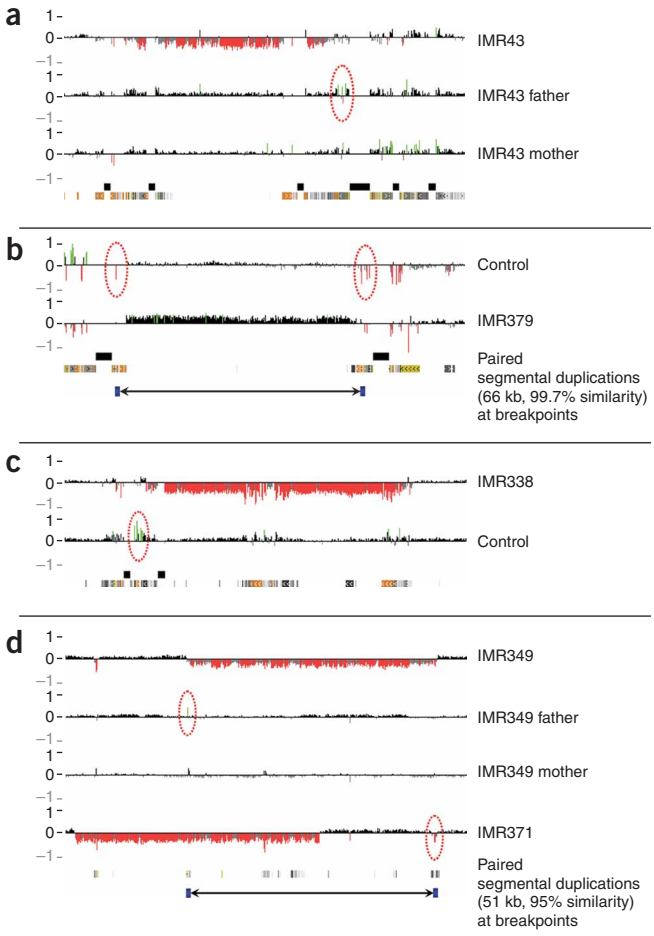

Figure 4 Structural resolution of five pathogenic rearrangements using high-density oligonucleotide arrays. Each plot shows the $\log_2$ ratio (*y* axis) for probes within each region of hg17 (*x* axis). (**a**) A 3-Mb region of 1q21.1 (chr1:143,000,000–146,000,000) showing a 1.47-Mb deletion in IMR43. (**b**) A 2.5-Mb region of 17q12 (chr17:31,500,000–34,000,000) showing a 1.46-Mb duplication in IMR379. (**c**) A 6-Mb region of 15q12–q14 (chr15:25,500,000–31,500,000) showing a 3.95-Mb deletion in IMR338. (**d**) A 6.5-Mb region of 15q23–q25.1 (chr15:70,000,000–76,500,000) showing overlapping 3.90-Mb and 3.81-Mb deletions in IMR349 and IMR371. For four of these rearrangements, the breakpoints map to clusters of flanking segmental duplications. Note that in each case these breakpoints coincide with the location of copy number variations in control individuals (dotted red ellipses), suggesting that these may be inherently unstable regions of the genome prone to chromosomal breakage[22]. For each individual, deviations of probe $\log_2$ ratios from 0 are depicted by gray and black lines, with those exceeding a threshold of 1.5 s.d. from the mean probe ratio colored green and red to represent relative gains and losses, respectively. Tracks below each plot indicate genome assembly gaps (black bars) and segmental duplications (gray/yellow/orange bars). In **b** and **d**, paired segmental duplications located at both rearrangement breakpoints are shown (blue bars joined by an arrow). For each individual except IMR43, rearrangements were also confirmed by FISH (**Supplementary Fig. 1**).

between the flanking sequences. In contrast, both breakpoints of the final case that we examined, a 15q23–24.2 deletion, occurred within unique sequence, suggesting this was not mediated by homology-based mechanisms and is therefore less likely to represent a recurrent event.

Notably, when we examined the breakpoint regions of the five pathogenic rearrangements that were embedded in large segmental duplication clusters, we noted extensive copy number variation of the flanking segmental duplications among normal individuals that coincided precisely with each breakpoint. This was most obvious in the case of the 17q21.31 microdeletion, where several structural polymorphisms occur at both the proximal and distal breakpoints (**Fig. 2**). In each of the other four cases, either one or both breakpoints also occurred at a site of copy number polymorphism in control individuals (**Fig. 4**). We note that similar copy number polymorphisms[7,10] (A.J.S. and E.E.E., unpublished data) and inversions[18–21] of flanking duplications are also evident for several other genomic disorders. Based on these findings, we propose that segmental duplications showing evidence of structural polymorphism may be another predictive feature of regions prone to recurrent rearrangement, perhaps corresponding to inherently unstable regions of the genome[22]. Given that we observed copy number variation at these breakpoint regions, it is also possible that in some individuals these flanking duplication structures exist in an alternative configuration (compared with the reference assembly) that predisposes them to altered frequencies of pathogenic rearrangement[23].

The recurrent 17q21.31 deletion that we describe occurs within a common ~900-kb inversion present in ~20% of northern Europeans[15,24]. Observations in three other genomic disorders suggest that microdeletions arise preferentially from chromosomes carrying an inversion of that same region[18–21]. Consistent with this, we observed that in both cases in which DNA samples from parents were available, the transmitting parents of the proband were carriers of the inversion-specific haplotype (**Supplementary Fig. 2** online). Given that the 17q21.31 inversion shows a geographical bias towards Europeans[15,24], we propose that the recurrent deletion of this region may also show similar population specificity, although additional data are required to test this.

Our study suggests a frequency of ~1% for this deletion in individuals with mental retardation, but testing of non-European populations should be performed to determine the true frequency worldwide. Furthermore, we note that the 17q21.31 inversion has undergone positive selection during recent human evolution, suggesting it provides a selective advantage to carriers[15]. Given the negative phenotypic consequences of the deletion, to our knowledge this is the first example in which segmental duplication architecture predisposes to both positive and negative selective pressures within a species. It will be interesting to study the evolutionary history of other complex regions of the genome in the context of counter-balancing positive and negative selective forces.

In conclusion, we demonstrate that by using an informed approach based on the duplication architecture of the genome, we have efficiently identified and refined sites of recurrent chromosomal rearrangement that underlie previously undescribed genomic disorders. This was exemplified by the discovery and description of a new 17q21.31 microdeletion syndrome as well as the identification of four additional strong candidate regions for recurrent rearrangements. We suggest that further analysis of the duplication architecture of the genome will be a rich source for the identification of structural variation, leading to improved understanding of the molecular basis of genomic rearrangements in genetic disease.

METHODS

BAC array CGH. A description of the segmental duplication BAC array and hybridization protocol have been published previously[7,10]. Briefly, the

segmental duplication array consists of 2,007 BACs, each spotted in triplicate, targeted to 130 complex regions of the genome flanked by paired intrachromosomal segmental duplications ≥10 kb in length and ≥95% sequence identity and separated by 50 kb–10 Mb of intervening sequence. All hybridizations were dye-swapped replicate experiments using a single male reference (GM15724, Coriell Institute). A BAC was considered copy number variant if the $\log_2$ ratio of fluorescence measurements exceeded twice the standard deviation of the autosomal clones in both dye-swapped experiments[10]. The control population was assayed using an identical protocol and consisted of 269 individuals of European ($n = 90$), sub-Saharan African ($n = 90$), Chinese ($n = 45$) and Japanese ($n = 44$) ancestry used in the International HapMap Project[25], and a further 47 individuals of diverse ethnic origin[7]. Copy number changes consisting of two or more adjacent BACs that were observed in the test population, but not in controls, were considered as potential pathogenic rearrangements.

A subset of individuals were further investigated by hybridization to a whole-genome tiling BAC array consisting of 32,989 BACs (unpublished data) in order to confirm and refine specific rearrangements.

Oligonucleotide array CGH. A custom duplex oligonucleotide array (NimbleGen Systems), consisting of 166,000 isothermal 45- to 70-bp probes, was targeted to seven microdeletion regions identified by the segmental duplication BAC array, covering a total of 23.17 Mb of sequence (mean density, one probe per 131 bp). Hybridizations were performed as described previously[26] and used the same reference individual as the BAC array hybridizations (GM15724).

Primer sequences. A list of primer sequences is provided in **Supplementary Table 6** online.

Accession numbers and microarray data. BAC array data have been deposited in the NCBI Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE5373. Oligonucleotide array data are available at http://humanparalogy.gs.washington.edu/structuralvariation/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The authors would like to thank all participating families and clinicians, particularly J. Flint, P. Bolton, A. Clarke, C. Fairhurst, T. Wolff, S. Mansour, S. Holder, R. Gibbons, L. Brueton, P. Day, F. Stewart, S. Keane, N. Meston, A. Seller, P. Clouston and K. Smith. This work was supported by grants from the US National Institutes of Health (NIH) (HD043569; E.E.E.), Merck Research Laboratories (A.J.S.), The Health Foundation (S.J.L.K.) and the Oxford Genetics Knowledge Park (S.J.L.K., R.R., C.G.). E.E.E. is an Investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

This study was coordinated by A.J.S., P.S.E., S.S., S.J.L. and E.E.E.; the manuscript was written by A.J.S. and E.E.E.; experimental work was performed by A.J.S., S.H., R.R.S., R.R., C.A.F., R.S. and C.G.; clinical work was performed by J.A.H., H.S., S.M.P., E.B. and R.C.H.; computational analysis was performed by Z.C.; and array production was performed by T.A.R., D.G.A. and D.P.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Bailey, J.A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
2. Cheung, V.G. *et al.* Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409**, 953–958 (2001).
3. Stankiewicz, P. & Lupski, J.R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
4. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
5. Iafrate, J.A. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
6. Fredman, D. *et al.* Complex SNP-related sequence variation in segmental genome duplications. *Nat. Genet.* **36**, 861–866 (2004).
7. Sharp, A.J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
8. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nat. Genet.* **37**, 727–732 (2005).
9. Lupski, J.R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14**, 417–422 (1998).
10. Locke, D.P. *et al.* Linkage disequilibrium and heritability of CNPs within duplicated regions of the human genome. *Am. J. Hum. Genet.* **79**, 275–290 (2006).
11. Knight, S.J. *et al.* Subtle chromosomal rearrangements in children with unexplained mental retardation. *Lancet* **354**, 1676–1681 (1999).
12. Edelmann, L. *et al.* A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum. Mol. Genet.* **8**, 1157–1167 (1999).
13. Potocki, L. *et al.* Molecular mechanism for duplication 17p11.2 – the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat. Genet.* **24**, 84–87 (2000).
14. Heilstedt, H.A. *et al.* Physical map of 1p36, placement of breakpoints in monosomy 1p36, and clinical characterization of the syndrome. *Am. J. Hum. Genet.* **72**, 1200–1212 (2003).
15. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).
16. de Vries, B.B. *et al.* Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.* **77**, 606–616 (2005).
17. Amos-Landgraf, J.M. *et al.* Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am. J. Hum. Genet.* **65**, 370–386 (1999).
18. Gimelli, G. *et al.* Genomic inversions of human chromosome 15q11-q13 in mothers of Angelman syndrome patients with class II (BP2/3) deletions. *Hum. Mol. Genet.* **12**, 849–858 (2003).
19. Osborne, L.R. *et al.* A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. *Nat. Genet.* **29**, 321–325 (2001).
20. Visser, R. *et al.* Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion. *Am. J. Hum. Genet.* **76**, 52–67 (2005).
21. Kurotaki, N., Stankiewicz, P., Wakui, K., Niikawa, N. & Lupski, J.R. Sotos syndrome common deletion is mediated by directly oriented subunits within inverted Sos-REP low-copy repeats. *Hum. Mol. Genet.* **14**, 535–542 (2005).
22. Zhou, Y. & Mishra, B. Quantifying the mechanisms for segmental duplications in mammalian genomes by statistical analysis and modeling. *Proc. Natl. Acad. Sci. USA* **102**, 4051–4056 (2005).
23. Kato, T. *et al.* Genetic variation affects *de novo* translocation frequency. *Science* **311**, 971 (2006).
24. Evans, W. *et al.* The tau H2 haplotype is almost exclusively Caucasian in origin. *Neurosci. Lett.* **369**, 183–185 (2004).
25. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
26. Selzer, R.R. *et al.* Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosom. Cancer* **44**, 305–319 (2005).
27. Liehr, T. *et al.* Mosaicism for the Charcot-Marie-Tooth disease type 1A duplication suggests somatic reversion. *Hum. Genet.* **98**, 22–28 (1996).
28. Juyal, R.C. *et al.* Mosaicism for del(17)(p11.2p11.2) underlying the Smith-Magenis syndrome. *Am. J. Med. Genet.* **66**, 193–196 (1996).