ARTICLE

Copy-Number Gains of *HUWE1* Due to Replication- and Recombination-Based Rearrangements

Guy Froyen,[1,2,*] Stefanie Belet,[1,2] Francisco Martinez,[3] Cíntia Barros Santos-Rebouças,[4] Matthias Declercq,[1,2] Jelle Verbeeck,[1,2] Lene Donckers,[1,2] Siren Berland,[5,6] Sonia Mayo,[3] Monica Rosello,[3] Márcia Mattos Gonçalves Pimentel,[4] Natalia Fintelman-Rodrigues,[4] Randi Hovland,[6] Suely Rodrigues dos Santos,[7] F. Lucy Raymond,[8] Tulika Bose,[9] Mark A. Corbett,[9] Leslie Sheffield,[9] Conny M.A. van Ravenswaaij-Arts,[10] Trijnie Dijkhuizen,[10] Charles Coutton,[11,12,13] Veronique Satre,[11,12,13] Victoria Siu,[14] and Peter Marynen[2]

We previously reported on nonrecurrent overlapping duplications at Xp11.22 in individuals with nonsyndromic intellectual disability (ID) harboring *HSD17B10*, *HUWE1*, and the microRNAs *miR-98* and *let-7f-2* in the smallest region of overlap. Here, we describe six additional individuals with nonsyndromic ID and overlapping microduplications that segregate in the families. High-resolution mapping of the 12 copy-number gains reduced the minimal duplicated region to the *HUWE1* locus only. Consequently, increased mRNA levels were detected for *HUWE1*, but not *HSD17B10*. Marker and SNP analysis, together with identification of two de novo events, suggested a paternally derived intrachromosomal duplication event. In four independent families, we report on a polymorphic 70 kb recurrent copy-number gain, which harbors part of *HUWE1* (exon 28 to 3′ untranslated region), including *miR-98* and *let-7f-2*. Our findings thus demonstrate that *HUWE1* is the only remaining dosage-sensitive gene associated with the ID phenotype. Junction and in silico analysis of breakpoint regions demonstrated simple microhomology-mediated rearrangements suggestive of replication-based duplication events. Intriguingly, in a single family, the duplication was generated through nonallelic homologous recombination (NAHR) with the use of *HUWE1*-flanking imperfect low-copy repeats, which drive this infrequent NAHR event. The recurrent partial *HUWE1* copy-number gain was also generated through NAHR, but here, the homologous sequences used were identified as TcMAR-Tigger DNA elements, a template that has not yet been reported for NAHR. In summary, we showed that an increased dosage of *HUWE1* causes nonsyndromic ID and demonstrated that the Xp11.22 region is prone to recombination- and replication-based rearrangements.

Introduction

Loss-of-function mutation and gene deletion are commonly known in the etiology of intellectual disability (ID), a condition that affects about 2% of the population and results in a significant loss of adaptive behavior.[1,2] Most of the loss-of-function mutations have been found in genes located on the X chromosome (X-linked ID, or XLID). However, an increased dosage of genes on the X chromosome has been implicated in XLID as well. The most prevalent copy-number gain is known as the *MECP2* duplication syndrome (MIM 300260), found as a nonrecurrent microduplication at Xq28,[3,4] with the smallest region of overlap of 130 kb, which includes only *IRAK1* (MIM 300283) and *MECP2* (MIM 300005).[5] The 2-fold increase in expression of *MECP2* was demonstrated as the cause of this severe syndrome observed in affected males, whereas carrier females are protected through preferential inactivation of the mutated X chromosome.[3,4] A second well-known X-linked microduplication syndrome is located at Xq22 and includes *PLP1* (MIM 300401); this syndrome leads to Pelizaeus-Merzbacher disease (MIM 312080). Interestingly, loss-of-function mutations in *MECP2* as well as *PLP1* in males can result in clinical characteristics that resemble those caused by their respective copy-number gains. In two other reported ID-associated copy-number gains on the X chromosome, the recurrent *FLNA-GDI1* amplifications (MIM 300815) and the nonrecurrent *HSD17B10-HUWE1* duplications (MIM 300706), the actual causes have not been firmly demonstrated.[6,7] However, candidate genes have been proposed for each copy-number variation (CNV) on the basis of reported functional and disease-related data. For the nonrecurrent microduplication at Xp11.22, we described a subtle copy-number gain in six families with mild to moderate nonsyndromic ID, for which we proposed two candidate genes, the 17-beta-hydroxysteroid dehydrogenase (*HSD17B10* [MIM 300256]) and the HECT, UBA, and

[1]Human Genome Laboratory, VIB Center for the Biology of Disease, KU Leuven, 3000 Leuven, Belgium; [2]Human Genome Laboratory, Center for Human Genetics, KU Leuven, 3000 Leuven, Belgium; [3]Genetics Unit, Hospital Universitario La Fe, 46009 Valencia, Spain; [4]Departement of Genetics, State University of Rio de Janeiro, Rio de Janeiro 20550-013, Brazil; [5]Department of Clinical Genetics, St. Olav's Hospital, 7018 Trondheim, Norway; [6]Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Helse Bergen, 5021 Bergen, Norway; [7]Gaffrée and Guinle University Hospital, Federal University of Rio de Janeiro State, Rio de Janeiro 20550-013, Brazil; [8]Cambridge Institute for Medical Research, Addenbrooke's Hospital, Cambridge CB2 0XY, UK; [9]Genetics and Molecular Pathology, SA Pathology, Women's and Children's Hospital, Level 9, Clarence Rieger Building, North Adelaide, South Australia 5006, Australia; [10]Department Genetics, University Medical Center Groningen, 9700 RB Groningen, The Netherlands; [11]Laboratoire de Génétique Chromosomique, Département de Génétique et Procréation, Hôpital Couple Enfant, Centre Hospitalier Universitaire de Grenoble, 38700 Grenoble, France; [12]Equipe Génétique, Infertilité et Thérapeutique, Laboratoire AGIM, FRE3405, Centre National de la Recherche Scientifique, 38706 Grenoble, France; [13]Université Joseph Fourier, Domaine de la Merci, 38706 Grenoble, France; [14]Department of Pediatrics, University of Western Ontario, London, Ontario N6C 2V5, Canada
*Correspondence: guy.froyen@cme.vib-kuleuven.be
http://dx.doi.org/10.1016/j.ajhg.2012.06.010.

WWE domain-containing protein-1 (*HUWE1* [MIM 300697]). Missense mutations in *HSD17B10* had been related to neurodegeneration (MIM 300438),[8] and a silent mutation resulting in a splicing defect was identified in an XLID family.[9] For *HUWE1*, missense mutations of highly conserved amino acids cosegregated with disease in three unrelated XLID families.[7] In addition to these two genes, the smallest region of overlap of the microduplication also contains the microRNAs *miR-98* and *let-7f-2*, for which no clear function has been described yet.

Whereas recurrent rearrangements of equal length are mediated by nonallelic homologous recombination (NAHR) between flanking low copy repeats (LCRs),[10] NAHR between Alu repeat elements or long interspersed nuclear elements (LINEs) have also been reported. Nonrecurrent rearrangements, on the other hand, have scattered breakpoints and are variable in length. Thus far, junction analyses of nonrecurrent microduplications have revealed nonhomologous end joining (NHEJ), fork stalling and template switching (FoSTeS), and microhomology-mediated break-induced replication (MMBIR) as the mechanisms of rearrangement, of which the latter two are replication-based.[11] Often, a complex genomic architecture containing LCRs and other repeat structures, as well as a high guanine-cytosine (GC) content, seems to render the region unstable and, therefore, more susceptible to rearrangements.[12–15]

Here, we present six additional unrelated families (for a total of 12) with individuals suffering from mild to moderate ID, all of whom harbor overlapping duplications at Xp11.22. Mapping by oligonucleotide (oligo) array and quantitative PCR (qPCR) revealed that *HUWE1* was the only gene within the minimal duplicated region. Its involvement in the disease was strengthened by expression analysis in lymphocyte RNA samples. Furthermore, a recurrent polymorphic copy-number gain including both microRNAs localized in intron 59 of *HUWE1* does not seem to affect cognition. Breakpoint mapping and in silico analysis of the disease-associated copy-number gains revealed replication-induced as well as NAHR-based rearrangements. Interestingly, the polymorphic recurrent microduplication within *HUWE1* was generated by NAHR between DNA elements. Our study thus demonstrates that increased dosages of *HUWE1*, located in a region prone to genomic rearrangements, result in nonsyndromic ID.

Subjects and Methods

Subject Samples

The protocol was approved by the appropriate institutional review board of the different genetic institutes or hospitals, and informed consent was obtained from the parents of the affected individuals and their healthy family members. Genomic DNA from affected individuals as well as from healthy controls was isolated from peripheral blood in accordance with standard procedures. Fragile X-chromosome-negative unrelated males with ID were screened for copy-number alterations on the X chromosome with a bacterial artificial chromosome (BAC) or oligo array comparative genomic hybridization (array CGH), or specifically in the Xp11.22 region with locus-specific qPCR or multiplex ligation-dependent probe amplification (MLPA). The screens were performed in the different institutes involved in this study and were made up of about 1,000 affected individuals.

Array CGH and qPCR Analysis

Full-coverage X chromosome BAC array CGH was performed on the DNA of ID-affected individual EX469 essentially as described elsewhere.[16] Full-genome oligo array CGH was done on individual SB1 on Affymetrix 6.0 arrays and on individual FTD on the Agilent Human Genome CGH 44K, and individual F538 was analyzed with the Nimblegen 385K array, as described.[17] The duplication in individual AU88848 was identified in a qPCR screen with SYBR-green with the use of the *HUWE1* qPCR primers, and MLPA detected the copy-number gain in ON1.

Fine mapping of the copy-number gains at Xp11.22 and copy-number analysis were performed with a custom-designed 4 × 44K oligo array (Agilent Technologies) that covers the repeat-masked region 52.50 Mb to 54.50 Mb at tiling resolution. All positions in this study are based on the UCSC Genome Browser, National Center for Biotechnology Information (NCBI) build 36.1, hg18. Each individual sample was hybridized versus a single male control sample, as described previously.[6] Hybridization and data analysis were performed by the VIB Nucleomics Core (Leuven, Belgium). Mapping of the CNVs in ON1, ON2, and HF was done by a custom-designed X-chromosome-specific 244K oligo array (Agilent), which contains the X chromosome exome as well as its flanking 5′ and 3′ untranslated region (UTR) sequences. Confirmation of copy-number gain and fine mapping of duplications were done by qPCR with the SYBRgreen relative quantitation method on a LC480 apparatus (Roche, Basel, Switzerland) as described previously.[6] qPCR primers were designed with the LightCycler Probe Design2 software (Roche), and sequences are provided in Table S1 (available online) or can be obtained upon request.

Breakpoint Mapping and In Silico Analysis

Genomic and cDNA sequences were loaded into the Vector NTI program for easy analysis and visualization (Life Technologies). For mapping the exact positions of the duplications in the 12 individuals with *HUWE1*-duplications, we tested, in an iterative way, the proximal as well as distal duplicated locations via qPCR, on the basis of the tiling Xp11.22-targeted oligo-array data. A detailed description of the procedure can be found elsewhere.[12] Expand long template (ELT)-PCR (Roche) was used to amplify the presumed junction fragments, and PCR products were directly sequenced with BigDye v3.1 on an ABI3130xl sequencer (Life Technologies).

Bioinformatic analysis was then performed on the 300 bp sequences located at the proximal and distal side of each breakpoint of the five junctions that we were able to sequence (two breakpoint regions in individuals FAM3, P083, A057, EX469, and F538), yielding a total of ten regions of 600 bp. For control regions, we randomly selected 50 regions of 600 bp spread over the entire X chromosome, as described previously.[12] Each control and breakpoint region was subsequently analyzed for repetitive DNA with RepeatMasker, Tandem repeats with ETANDEM, fractional GC content with GEECEE, and non-B-DNA structures with Non-B DB. The alternative tools ZHUNT, QGRS, and

PALINDROME were also used for separate detection of such structures. Recombination-associated motifs were searched for with FUZZNUC. Additionally, a 2 Mb region surrounding *HUWE1* (52.6–54.6 Mb) was checked for the presence of segmental duplications in the Human Genome Segmental Duplication Database, and the self-chain tool of the UCSC Genome Browser was used to find highly similar DNA stretches elsewhere in the genome. The breakpoint regions of the remaining individuals in whom we could not define the junctions were analyzed with RepeatMasker.

cDNA Expression Analysis

Total human RNA was extracted from white blood cells or Epstein-Barr virus-transformed peripheral blood lymphocytes (EBV-PBLs) from individuals and controls, and a qPCR was performed via the SYBRgreen method with cDNA-specific primers (Table S1), as described previously.[6] The housekeeping genes *GUSB* and *HPRT* were used for normalization. For expression analysis in the mouse, total RNA was extracted from several mouse tissues, including different brain regions, with TRIzol (Life Technologies) as described elsewhere.[6] A qPCR for *Huwe1* expression was then performed with *Gusb* and *Hprt* used as normalizers.

Analysis of a *HUWE1-HUWE1* fusion gene in an individual with the partial *HUWE1* duplication was checked through regular PCR with the use of primers in the penultimate exon (exon 82) of *HUWE1* and the first exon (exon 29) within the duplicated *HUWE1* fragment (Table S1).

X-Inactivation, Marker, and SNP Analyses

Lymphocyte-derived genomic DNA was subjected to the androgen receptor (*AR*) methylation assay for assessment of the methylation status.[18] DXS988 (53.35 Mb) and DXS1199 (53.70 Mb) marker analysis was performed on 50 ng of DNA from affected male individuals and their female relatives for whom we had DNA. A ROX-labeled genotyping marker 100–500 was added to the samples, which were separated on an ABI3130xl automated DNA sequencer and analyzed with the GeneMapper analysis software (Life Technologies) for peak-position and area-intensity calculations. Data were further processed in Excel (Microsoft).

For SNP analysis, primers were designed to amplify rs1264014 (53.43 Mb; Centre d'Étude du Polymorphisme Humain [CEPH] allele frequency of C:66 and A:34), and rs266786 (53.62 Mb; CEPH allele frequency of G:37 and C:63). A PCR was performed on 50 ng of DNA with Platinum Taq (Life Technologies) and directly sequenced. Sequencing samples were analyzed on an ABI3130xl apparatus, and the nucleotide present at each SNP position was scored for each sample. Primer sequences can be found in Table S1.

Results

Subjects

Families with Xp11.22 Duplications that Include HUWE1

The clinical description and family pedigrees of individuals FAM3, A009, A049, A057, A119, and P083 have been described previously.[7] The pedigrees of the new families are shown in Figure 1A. All affected individuals presented with mild to moderate ID.

Family F538

Affected males from this South African family suffer from moderate nonsyndromic ID. Individual II.4 has a long face, long ears, and unequal pupils. Male III.5 has a large head circumference (orbitofrontal cortex [OFC)] > 97th percentile) and speaks with a stutter. Individuals IV.1 and IV.2, who could not be tested for the presence of the duplication, are overweight and have limited speech, an abnormal gait, and incurved fifth fingers. The healthy obligate carrier grandmother (II.2) has skewed X inactivation (98/2).

Family EX469

This family has been reported earlier as case 7 in an X chromosome array CGH screen.[16] The proband and his brother suffered from moderate nonsyndromic ID with minor dysmorphic features, including urogenital anomalies and gastroesophageal reflux with chronic vomiting and seizures. They also present with café-au-lait spots on the skin and have poor language abilities but learned to read and write. Brain computed tomography and MRI results were normal. Electroencephalogram (EEG) results in the proband showed epileptiform abnormalities of focal expression in deep structures of the left temporal lobe. Haplotyping revealed that the duplication arose de novo on the grand paternal X chromosome.

Family FTD

The proband (III.3) was born at 38 weeks' gestation with a weight, height, and OFC well within the normal ranges. On physical examination at 6 years, his height and weight were 113 cm (50th centile) and 19 kg (25th centile), respectively. Congenital abnormalities include functional heart murmur, chronic vomiting and diarrhea, urolithiasis, bilateral inguinal hernia, and cryptorchidism. Facial dysmorphism is seen in the form of low-set and dysmorphic ears, a bulbous nose, and the absence of some teeth. He has mild ID (intelligence quotient = 67) and global developmental delay, sitting without support at 12 months and walking without support at 19 months. He spoke his first word at 24 months and spoke simple sentences at 42 months. He suffered from paroxysmal sleep episodes, attention deficit, and hyperactivity. Currently at the age of 11, he attends a regular school with curricular adaptation; he has learned to read and write. Brain MRI results were normal, but EEG results showed some cortico-subcortical dysfunction with sporadic potentials of acute morphology in the vertex. Interictal EEG results showed acute abnormalities in sleep in the vertex and frontal regions of both hemispheres. His carrier mother (II.2) and older sister (III.2) are both healthy. Haplotyping demonstrated that the duplication arose de novo on the paternal X chromosome of the mother.

Family AU88848

In this small Australian family, both affected brothers (II.1 and II.2) suffer from mild nonsyndromic ID. Their sisters (II.3 and II.4) as well as mother (I.2) are phenotypically normal. No additional information is available.

Family SB1

In this family, all three boys suffered from ID. The unaffected mother (II.2) is the carrier of the duplication, which most likely is a de novo event, because the duplication was

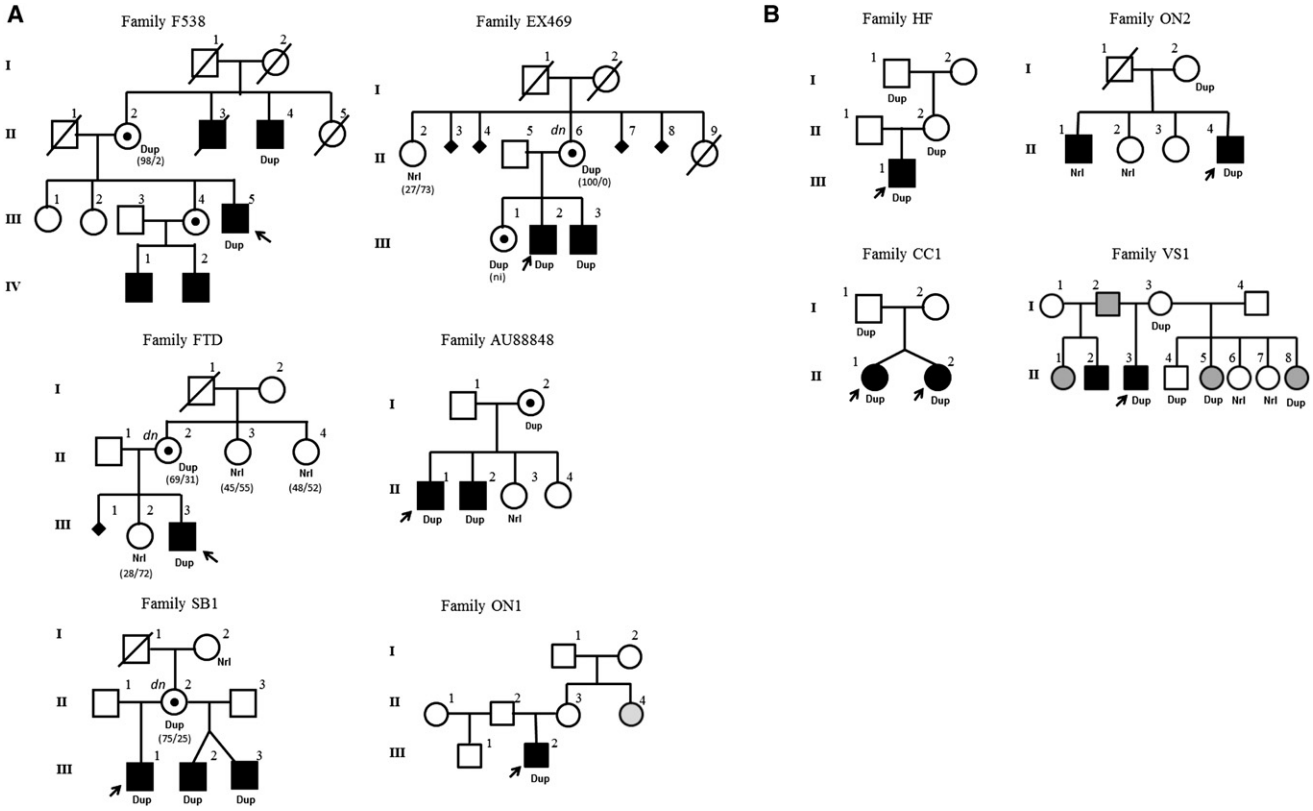

Figure 1. Pedigrees of Families with Xp11.22 Copy-Number Gains
(A) Pedigrees of six unreported families with a nonrecurrent microduplication at Xp11.22, including *HUWE1*. All tested individuals are marked with Nrl if the duplication was not present or Dup if the duplication was found in this individual. X-inactivation ratios in females are provided between brackets. Ni, not informative; dn, de novo event.
(B) Pedigrees of the four families with the partial *HUWE1* copy-number gain. Segregation of the aberrations was analyzed by qPCR in family members from whom DNA was available. Family members with weak cognitive levels are indicated in gray.

absent in their maternal grandmother (I.2) and the deceased grandfather did not present with ID. The elder brother presents with mild ID and remarkably slow speech. He walked at 16 months of age with some balance problems and talked at 4 years of age. At age 16, he developed kyphosis, later diagnosed as Bechterew syndrome (MIM 106300), and gynecomastia, which could however be due to a significant weight loss. His testosterone levels and gonadotropin levels (luteinizing hormone and follicle-stimulating hormone) were also low to subnormal. The twin brothers walked at 22 months, and talking was only problematic at that time. They are both diagnosed with mild ID, somewhat worse than their brother. None of the brothers have behavioral or social problems. All three boys have downslanting palpebral fissures and a prominent supraorbital ridge.

Family ON1

The proband (III.2) is a 12-year-old boy who was born from an uneventful delivery, at 3,650 g and 53 cm. He grew with developmental delay, especially in speech, walking at 24 months, speaking a few words at 36 months, and controlling his sphincter at 48 months. Hyperactivity and attention problems were also observed, and at school, he showed hyperkinetic behavior. He was treated with haloperidol and carbamazepine. At 7 years of age, he presented with normal growth parameters, partial lack of speech, mild to moderate ID, hyperactivity, and self-destructive behavior. He had anterior down-sweep scalp, microphtalmos, broad nasal root, bulbous nose, high-arched palate, square and small teeth, micrognathism, low-set ears, malformed auricles, large thoracic cage, hypoplasia of nipples, and brachydactyly and clinodactyly of the fifth fingers.

Families with the Partial HUWE1 Copy-Number Gain

The pedigrees of these four families are shown in Figure 1B.

Family HF

The proband was a 4-year-old boy who was born after an uneventful pregnancy with a low birth weight (2,720 g, 3^{rd}–10^{th} centile), club feet, and flexion contractures of the knees and elbows. He was able to walk without support at the age of 17 months but showed a speech delay. He developed epilepsy at the age of 1 year and suffered from mild to moderate ID, requiring special education.

Family CC1

Both monozygotic female twins from unrelated healthy parents presented with ID and dysmorphic features. However, their clinical phenotype was attributed to a de novo 7.5 Mb autosomal deletion at 18q (unpublished data).

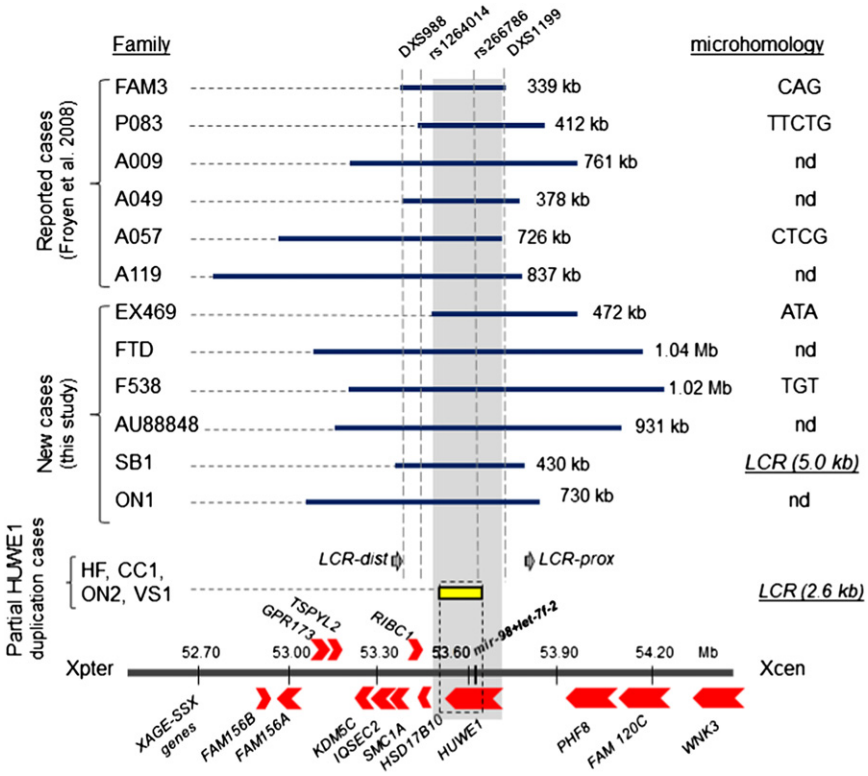

Figure 2. Overview of the 12 Nonrecurrent and 4 Recurrent Microduplications at Xp11.22

The locations and sizes of each nonrecurrent duplication, as defined by high-resolution oligo array and iterative qPCR mapping, are illustrated by horizontal blue bars. The sizes of each copy-number gain are indicated, as well as the sequence used as microhomology substrates for rearrangement, if defined. Where LCR was involved, the minimal size of the LCR is provided between brackets. The positions of the SNPs and markers that we analyzed are indicated at the top. The genes present within this region are shown at the bottom, in red. The shortest region of overlap is indicated in gray and harbors *HUWE1* and the intronically located microRNAs *miR-98* and *let-7f-2*. The recurrent polymorphic duplication identified in the four unrelated families is indicated by the yellow box shown above the genes. Positions are according to UCSC hg18. Nd, not determined due to insufficient DNA, repeat-rich, or nonreference breakpoint regions.

Family ON2

Both affected boys of this Brazilian family present with nonsyndromic ID, which was more severe in II.4 compared to II.1. No additional information is available on this family.

Family VS1

The male proband (II.3) presented with severe ID, partial complex seizures, and moderate to profound sensorineural hearing impairment. Both of his parents, as well as three out of six children from other marriages of each parent, showed variable borderline degrees of learning disability.

Identification of Microduplications at Xp11.22

On the basis of our initial report on six male cases of ID with microduplications at Xp11.22,[7] we collected, in an international collaborative effort, six additional individuals (EX469, FTD, F538, AU88848, SB1, and ON1) with a copy-number gain in this region. The CNVs were identified with different array platforms as described in Subjects and Methods. On the basis of the array data, all six microduplications had overlapping but different locations, with sizes varying between 0.4 and 1.0 Mb. All duplications were confirmed through qPCR with the use of primer sets in *HSD17B10* (53.47 Mb) and *HUWE1* (53.72 Mb), and segregation analysis on available samples from each family was in line with its proposed disease-causing effect (Figure 1A). We confirmed that a total of five additional male individuals with ID in these six families harbored the Xp11.22 duplication, and two from family F538 probably had the duplication but could not be tested.

Through oligo X-chromosome-specific array CGH, we also detected a much smaller, apparently recurrent copy-number gain of about 100 kb in two unrelated families (HF and ON2) (Figure 2; Figure S1). We then obtained access to two other families (CC1 and VS1) with an apparent similar copy-number gain. qPCR data revealed four copies in family ON2, whereas two copies were confirmed in the other three families. Segregation analysis demonstrated that this CNV does not correlate with the ID phenotype, because in families HF and CC1, the healthy grandfather and father, respectively, also harbor this subtle copy-number gain (Figure 1B). Moreover, the equally affected brother of the proband in family ON2 did not harbor the aberration. In family VS1, no correlation could be made, because of the overall weak cognitive skills in these family members. This subtle duplication is not reported in the DECIPHER or ECARUCA databases and is present in three cases (nssv585241, nssv585242, and nssv581443) of two families in the ISCA database as a 63-kb-large gain. Therefore, it is categorized as an extremely rare polymorphic variant that we will call a "partial *HUWE1* copy-number gain" to discriminate it from the larger causal nonrecurrent Xp11.22 duplications that fully include *HUWE1*, among others.

Duplication Mapping

Duplication mapping was performed in the probands of all 12 families. For eight individuals for whom the duplication had been detected by low-resolution X-chromosome-specific BAC array (FAM3, A049, A057, A119, AU88848,

Table 1. Mapping the Microduplications via Oligo Array CGH and Junction Sequencing

N°	Fam	Last Normal	Start Dupl.	Stop Dupl.	First Normal	Estim. Size	Start Dupl.	Stop Dupl.	Exact Size	μhom
1	FAM3	53,399,094	53,399,112	53,739,965	53,739,977	340,853	53,401,082	53,739,998	338,916	CAG
2	P083	53,457,383	53,457,408	53,864,127	53,869,098	406,719	53,457,459	53,869,097	411,638	TTCTG
3	A057	53,004,397	53,004,414	53,729,683	53,729,701	725,269	53,004,378	53,729,969	725,591	CTCG
4	EX469	53,501,651	53,501,669	53,973,945	53,974,166	472,276	53,501,661	53,974,001	472,340	ATA
5	A049	53,409,616	53,409,631	53,787,587	53,787,659	377,956	nd	nd	nd	
6	A119	52,839,940	52,842,342	53,679,493	53,681,026	837,151	ni	ni	ni	
7	AU88848	53,182,062	53,186,632	54,117,977	54,119,631	931,345	nd	nd	nd	
8	FTD	53,210,426	53,215,720	54,254,252	54,262,967	1,038,532	nd	nd	nd	
9	A009	53,235,000	53,237,000	53,998,000	54,003,000	761,000	nd	nd	nd	
10	F538	53,232,358	53,233,028	54,256,395	54,256,395	1,023,367	53,232,548	54,256,847	1,024,299	TGT
11	SB1	53,384,501	53,387,143	53,807,385	53,813,962	420,242	53,380,181	53,809,779	429,598	LCR
12	ON1	52,994,153	52,999,509	53,729,676	53,979,838	730,167	nd	nd	nd	

N°, number; Fam, family; Dupl., duplication; Estim., estimated; μhom, microhomology; nd, not determined; ni, not investigated; LCR, low-copy repeat. Arrays for ID-affected individuals 1–8, based on custom Agilent targeted Xp11 oligo array data; individuals 9–10, based on Nimblegen X-chromosome-specific oligo array data; individual 11, based on Affy 6.0 oligo array hg18 data; individual 12, based on custom Agilent X-chromosome-specific oligo array data. All positions are based on UCSC Genome Browser hg18 (March 2006).

P083, EX469, and FTD), we performed hybridization onto the custom-designed 4 × 44K tiling oligo array (Figure S2). All positions are based on the UCSC Genome Browser (build 36.1, hg18). For six persons (not A049 and A119), both the proximal and distal breakpoint regions could be mapped to < 5 kb regions. Data of these arrays are deposited in the GEO database under accession number GSE32945. For the remaining four families, the microduplications were mapped with a custom 385K X-chromosome-specific oligo array CGH for families F538 and A009 as described earlier,[17] and those for SB1 were localized with the Genome-wide Affymetrix 6.0 SNP array. Finally, the aberration in ON2 was fine mapped with our custom-designed oligo X-chromosome-specific array. The breakpoint regions deduced from the array data are summarized in Table 1, and a schematic overview of the 12 microduplications is shown in Figure 2.

The partial *HUWE1* copy-number gain was identified in the probands of families HF and ON2 as an increased Cy5/Cy3 $\log_2$ ratio on our custom oligo X-chromosome-specific array. Probes spanning the region 53.58 to 53.64 Mb deviated from the normal $\log_2$ ratio (Figure S1). The proximal breakpoint of this partial *HUWE1* copy-number gain is located in intron 28, and the distal one maps downstream of *HUWE1*. In agreement with the qPCR data, the mean $\log_2$ ratio of the array obtained for HF was 0.54, suggesting a 2-fold increase, whereas ON2 had a $\log_2$ ratio of 0.78, confirming the higher copy-number gain. Fine mapping through qPCR revealed a maximal duplication size of 88 kb (53.55 to 53.64 Mb), which could not be refined any further by qPCR because of the highly repetitive nature of the remaining unmapped region. The minimal size was 64 kb (Figure 3). We obtained the same qPCR mapping results in samples CC1 and VS1, strongly suggesting it is a recurrent recombination event in the four families.

Investigation of Recombination Breakpoint Sites

Our mapping data of the nonrecurrent duplications indicated that all proximal and distal breakpoints differed; therefore, each copy-number gain is unique (Figure 2), which excludes NAHR as the common mechanism for Xp11.22 copy-number gains. Mapping was not continued for individual A119, because the distal breakpoint is located in the highly repetitive *MAGE-XAGE-SSX* cluster at Xp11.22. We designed PCR primers that span the presumed junction sites, and ELT-PCR generated PCR products of the expected size for five families (FAM3, P083, A057, EX469, and F538) not present in control samples. Sequencing of these products revealed the exact positions of these five junctions (Table 1), which showed microhomology of 3 to 5 nucleotides of the proximal and distal ends (Figure 2; Figure S3). In family P083, however, a tandem TTCTGCCTGGG sequence is present, pointing to DNA slippage and suggesting a replication-dependent rearrangement. The TTCTG sequence, present at both breakpoint sites, could have been used as a microhomology anchor point (Figure S3). In family SB1, an LCR was present at both breakpoint sites; the distal LCR was 5.6 kb (LCR-dist; chrX:53,380,181-53,385,785) and the proximal LCR was 5.0 kb (LCR-prox; chrX:53,809,782-53,814,750) (Figure 2; Figure S3). ELT-PCR with the LCR-prox-for and LCR-dist-rev primers revealed a PCR product of the expected size of 6 kb in SB1, but not in controls (Figure S3). Sequence analysis of both ends confirmed that the junction of the duplication in SB1 was generated

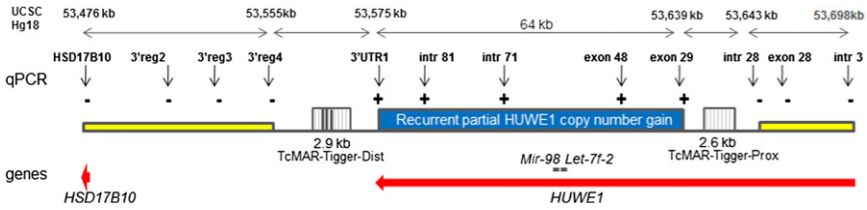

Figure 3. A Recurrent Partial *HUWE1* Copy-Number Gain Is Identified in Four Unrelated Families

Mapping of the recurrent copy-number gain was performed by iterative rounds of qPCR, for which each result is indicated as "−" for a normal copy number of 1.0 or "+" when the locus had a copy-number gain. The duplication starts at 53.57 Mb, which is upstream of *HUWE1*, and ends at 53.63 Mb within intron 28 of this gene. Analysis of both breakpoint regions by RepeatMasker identified the presence of 2.9 and 2.5 kb TcMAR-Tigger DNA elements (indicated as vertical-striped boxes) at the distal (TcMAR-Tigger-Dist) and proximal (TcMAR-Tigger-Prox) side, respectively. The locations of *HSD17B10*, part of *HUWE1*, and both microRNAs are at the bottom. Positions (in kb) are according to UCSC hg18.

within both LCRs most likely by the mechanism of NAHR. The percentage of identity between both LCRs is > 91%, but the matching sequences (3,920 bp in total) are interrupted by four short stretches of interspersed repeats, one in LCR-dist (465 bp) and three in LCR-prox (319, 446, and 460 bp). Therefore, we call those imperfect LCRs. For the five remaining families for which the generation of the junction was unsuccessful (A009, A049, FTD, AU88848, and ON1), iterative rounds of qPCR were employed to more precisely map the breakpoint sites. Then, we tried different ELT-PCR conditions, also taking into account potential inverse orientations, all without success. For A009, ELT-PCR was unsuccessful even on control regions, suggesting that the quality of DNA was insufficient. For ON2, the distal breakpoint region maps to the *FAM156A* and *FAM156B* inverse repeat, precluding efficient primer design. Finally, for FTD and AU88848, we could not amplify one or both (total of three) of the reference breakpoint regions. Two out of three of these breakpoint sequences could not be obtained in controls as well, strongly suggesting that the reference sequences at both positions deviate from the actual sequence, at least in our analyzed control samples.

Because this Xp11.22 region seems particularly prone to genomic rearrangements, we analyzed with bioinformatic tools the ten breakpoint regions for which we defined the breakpoint to the nucleotide level (families FAM3, P083, A057, EX469, and F538). For each proximal and distal breakpoint, we analyzed the flanking 600 bp (300 bp at each side) for GC content and the presence of repeats, non-B-DNA structures, and known recombination motifs. We then compared these results with those of 50 control regions. The GC content of the ten breakpoint regions was significantly higher (48% ± 10) compared to that of the controls (36% ± 6) (Table 2). We did not find differences in repeat content, but for the non-B-DNA structure analysis, the abundance of G-quadruplex-forming repeats and palindromes was significantly higher for the breakpoint regions compared to controls ($p = 0.029$ and $p = 0.007$, respectively). Similarly, of the 41 analyzed recombination motifs, a significantly higher percentage was observed in the breakpoint regions for DNA polymerase A frameshift hot spot 1 ($p = 0.0115$), DNA polymerase B frameshift hot spot 1 ($p = 0.0034$), translin binding site 2 ($p = 0.0004$), murine parvovirus recombination hot spot ($p = 0.0001$), and murine major histocompatibility complex recombination hot spot ($p = 0.0206$) (Table 2). We realize that the low number of breakpoint sites requires careful interpretation of the data.

For the partial *HUWE1* copy-number gain, the proximal breakpoint region was narrowed to a 3.7 kb sequence in intron 28, whereas the distal breakpoint region could only be reduced to a 20 kb sequence (Figure 3). The repeat content of the 20 kb sequence was 96.4%, as defined by RepeatMasker. In silico analysis of the distal and proximal breakpoint regions with Bl2SEQ (NCBI) and RepeatMasker revealed a 2.3-kb-long DNA element of the subtype TcMAR-Tigger2 in both regions. Therefore, we hypothesized that NAHR was the mechanism for this apparent recurrent CNV and designed unique primers flanking each DNA element (Figure 3). ELT-PCR with the primer pair Tigger-Prox-for and Tigger-Dist-rev yielded a PCR product of the expected 3.5 kb size in individuals HF, ON2, CC1, and VS1, but not in controls (Figure S4). We subsequently confirmed this recombination event in all the families by cloning and sequencing the PCR products, which mapped the recombination site within a 562 bp stretch of identical sequence (ChrX: 53,567,965–53, 568,527 for the distal element, and ChrX: 53,641,033–53, 641,595 for the proximal one) present within both TcMar-Tigger2 elements (Figure S5). Our data thus show that this DNA element was used as the substrate for NAHR, generating a recurrent direct tandem duplication with a 562 bp recombination hot spot.

Gene Content of the Duplications

The duplication sizes of the Xp11.22 gains vary from 339 kb (FAM3) to about 1.0 Mb (FTD and F538) with a common minimal overlap of 228 kb (start at 53,501,669 [EX469]; end at 53,729,682 [A057]) (positions based on UCSC hg18) (Figure 2 and Table 1). This common interval contains *HUWE1* as well as the microRNAs *miR-98* and *let-7f-2* located in intron 59 of *HUWE1*, strongly suggesting that a 2-fold increase in the dosage of (one of) these genes is causing the mild to moderate nonsyndromic ID in our affected individuals. Because of the variable location and size of each of these CNVs, other genes with mutations resulting in known diseases are involved as well. Duplication

Table 2. Bioinformatic Analysis of the 600 bp Regions of the Ten Breakpoints Identified in the Individuals with ID and Comparison with 50 Randomly Selected Regions on the X Chromosome

	10 Breakpoint Regions		50 Control Regions		
Motif	**N°**	**Freq.**	**N°**	**Freq.**	**p Value**
Two-Tailed t Test					
GC content		48% ± 10		36% ± 6	0.0019 **
Fisher's Exact Test					
Direct repeat (slipped motif)	2	20%	5	10%	0.33
Inverted repeat (cruciform motif)	2	20%	8	16%	0.53
Mirror repeat	1	10%	13	26%	0.94
A-phased repeat	1	10%	5	10%	0.68
G-quadruplex-forming repeat	3	30%	2	4%	0.029 *
Z-DNA motif	2	20%	6	12%	0.40
Palindromes	6	60%	8	16%	0.007 **
DNA polymerase A frameshift hot spot 1 TCCCCC	4	40%	3	6%	0.011 *
DNA polymerase B frameshift hot spot 1 ACCCWR	10	100%	25	50%	0.003 **
Translin binding site 2 GCCCWSSW	7	70%	6	12%	0.0003 ***
Consensus scaffold attachment region 4 TWWTDTTWWW	4	40%	41	82%	0.011 *
Murine parvovirus recombination hot spot CTWTTY	7	70%	4	8%	0.00008 ***
Murine MHC recombination hotspot GAGRCAGR	5	50%	7	14%	0.02 *

N°, number; Freq., frequency; GC content, guanine-cytosine content; MHC, major histocompatibility complex.
Asterisks represent significant difference (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

in six families (A057, A119, FTD, F538, AU88848, and ON1) also harbor *KDM5C* (MIM 314690) and *IQSEC2* (MIM 300522), and in three families (FTD, F538, and AU88848), *PHF8* (MIM 300560) is involved. In all families except one (EX469), *HSD17B10* (MIM 300256), implicated in a neurodegenerative disease (MIM 300438), is duplicated as well. Finally, *SMC1A* (MIM 300040), mutations of which result in Cornelia de Lange syndrome 2 (MIM 300590), is contained within the duplication of all but two families (P083 and EX469). Even though the ID phenotype is similar in all individuals described in this study, we cannot exclude the possibility that other clinical characteristics are due to subtle dosage effects of any of the other genes.

The recurrent partial *HUWE1* copy-number gain starts in intron 28 and ends downstream of *HUWE1*. Because this polymorphic CNV was shown to be a direct tandem event, one copy of *HUWE1* is still present, an explanation for toleration of this CNV in healthy persons. The presence of the two microRNAs *miR-98* and *let-7f-2* in intron 59 of *HUWE1* implies that both microRNAs are present at four copies in the male of family ON2, and at two copies in the males of families HF, CC1, and VS1, who harbor this CNV.

Quantitation of *HSD17B10* and *HUWE1* mRNAs and Analysis of X Inactivation

Expression of *Huwe1* in several mouse tissues (cortex, hippocampus, tongue, eye, kidney, liver, adrenal gland, and tail fibroblasts) as measured by qPCR revealed highly similar mRNA levels in all tissues investigated. The expression levels were moderately high because expression was only four times (delta cycle threshold = 2) lower compared to the housekeeping gene *Hprt* (data not shown). In human control EBV-PBL cell lines, we did not find any difference in *HUWE1* expression levels between males and females, strongly suggesting that this gene does not escape X inactivation. As demonstrated earlier, we found about 2-fold-increased *HUWE1* levels in cell lines from duplication individuals. Given that *HSD17B10* is duplicated in all families except EX469, we wanted to exclude the possibility that a position effect of the copy-number gain in this family resulted in increased *HSD17B10* levels as well. Therefore, we quantified the mRNA levels in blood lymphocytes from the proband of this family. As can be seen in Figure 4, expression of *HUWE1* was about 2-fold higher compared to controls, whereas the mRNA levels of *HSD17B10* were not enhanced, demonstrating that *HSD17B10* is unlikely to be the dosage-sensitive candidate gene that confers the ID phenotype. To investigate altered transcript levels in individuals with the partial *HUWE1* copy-number gain, we tested expression of *HUWE1* in lymphocytes of the proband (II.3) of family VS1 and his healthy half-brother (II.4), who also carries the duplication, using qPCR primer pairs in exon 23–24 (not duplicated) and in exon 29 (duplicated), but no increased mRNA levels were found as compared to controls (data

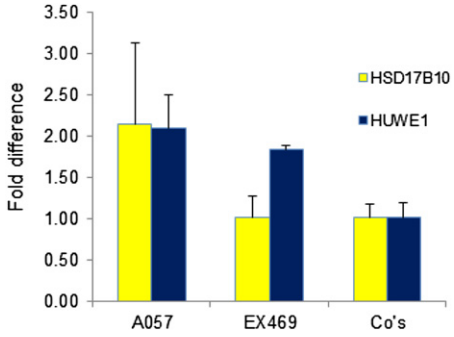

Figure 4. *HUWE1* but Not *HSD17B10* mRNA Expression Is Increased in EX469
The microduplication in EX469 is the only one that does not include *HSD17B10*, thereby limiting the minimal duplicated region to *HUWE1* only. The qPCR data are the mean of four independent experiments. Four controls were used in each experiment. SDs are provided.

not shown). Also, expression of *HSD17B10* was not affected. Because of a direct tandem duplication event, we checked for the formation of a potential fusion gene coupling the penultimate exon (82) or last exon (83) of *HUWE1* to exon 29, but no PCR products were obtained. The lack of a fusion gene is in agreement with our expression data.

X-inactivation analysis of four carrier females of the Xp11.22 CNVs revealed ratios of 69/31, 75/25, 98/2 and 100/0 (Figure 1A). The ratios in the females of HF1 and ON2 with the partial *HUWE1* copy-number gain were 70/30 and 56/44, respectively. DNA from other females was either not available to us or not informative at the *AR* locus.

Chromosomal Origin of Duplication Events

To assess the chromosomal origin of recombination, we searched for heterozygous calls in the duplicated regions. We first analyzed the repeat length of the highly polymorphic dinucleotide markers DXS988 (53.35 Mb) and DXS1199 (53.70 Mb) (indicated in Figure 2). Of the 12 families, DXS988 is located within the duplicated region of 9 families, and DXS1199 for 11 families. Both markers showed homozygous alleles in all probands. Measured repeat lengths were 128, 130, 132, 134, and 136 bp for DXS988, and 278, 280, 282, 284, and 286 bp for DXS1199. Note that allele lengths were independent of region of origin. Notably, haplotypes were shared in only three (out of 12) unrelated probands (Table S2). Two carrier females (A009 and P083) as well as six control females selected from the same countries as the affected individuals revealed heterozygous calls for at least one of both markers. Furthermore, all probands were homozygous for the highly polymorphic SNPs rs1264014 and rs266786, both located within the common duplicated region (Table S2). As haplotypes, we obtained seven C-A, four G-C, and one C-C combination. Although these data cannot prove the origin of the rearrangements, they are strongly in favor of intrachromosomal duplication events.

In all four families with the partial *HUWE1* copy-number gain, genotyping SNP rs266786, located within the aberration, revealed homozygous calls (C in HF and CC1; A in ON2 and VS1).

Discussion

We previously reported on overlapping microduplications at Xp11.22 in six families with mild to moderate nonsyndromic ID.[7] Here, we report on six additional families with overlapping microduplications, which reduced the smallest region of overlap to 228 kb, excluding *HSD17B10*. Moreover, four families with a recurrent subtle polymorphic copy-number gain within *HUWE1* were identified. We thus could exclude both microRNAs, *miR98* (*MIR98*) and *let7f-2* (*MIRLET7F2*), as well. Our data therefore show that *HUWE1* is the dosage-sensitive gene for which a 2-fold overexpression results in cognitive impairment in males. The ubiquitously expressed *HUWE1*, previously called *MULE1* and *ARF-BP1*, codes for an E3 ubiquitin ligase that was initially found to have a crucial role in cancer.[19,20] However, HUWE1 was later implicated in neuronal development as well. Neuronal and/or glial cell-specific *Huwe1* knockout mice showed abnormalities in the laminar patterning of the cerebral cortex, in the distribution of the external granule layer of the cerebellum, and in the organization of Bergmann glial cells, pointing to a important role of Huwe1 in the programming of neural progenitor cells.[21,22] Loss- and gain-of-function assays in the mouse cortex provide a critical function for Huwe1 in neuronal differentiation through ubiquitination of N-Myc, thereby affecting the N-Myc-DLL3 pathway.[22,23] Similarly, Huwe1 seems to contribute to the physiological process of neuronal turnover in the brain through ubiquitination of Tp53. Mice with an increased Tp53 activity show premature loss of neurogenic capacity.[24] Disturbance of the TP53 levels probably affects the rate of proliferation of neuronal progenitor cells, but not the rate of apoptosis of these progenitors. Therefore, the fine-tuning of this balanced proliferation and apoptosis mechanism, in order to establish a correct number of newly differentiated neurons, is disturbed. Finally, missense mutations of highly conserved amino acids previously identified in three ID families emphasize its role in physiological neuronal processes.[7]

We previously demonstrated via fluorescence in situ hybridization that the duplicated regions in A009, A049, and A057 are located at Xp11.22 and not inserted elsewhere in the genome.[7] Here, we demonstrated direct tandem duplications in FAM3, P083, A057, EX469, F538, and SB1 through junction analyses, showing that at least seven microduplications are due to local rearrangements. The junction sequences revealed direct tandem events having a microhomology of 2 to 5 bp with the absence of an information scar, or the insertion of sequences from distant regions between the proximal and distal

breakpoint regions. For the junction of the 412 kb duplication in P083, a stretch of 11 bp was repeated once, which apparently resulted in a 6 bp insertion. The replication-based mechanism of serial replication slippage (SRS) is not a probable explanation here because of the large size of the aberration. SRS-induced duplications are thought to be restricted to the Okazaki fragment, which is only a few hundred bp in men.[25]

Our attempts to generate junction PCR products in the other families were not successful due to insufficient or poor-quality DNA (A009), the presence of long-interspersed (A049) or low-copy repeats (A119 and ON1) at one of the breakpoints, or sequences at one breakpoint region that deviate from the reference sequence (FTD and AU88848). Even though repetitive or nonreference sequences could have precluded obtaining the junction fragments in the latter four families, a more complex structure of the rearrangements might also explain our inability to obtain the junctions. The relatively high occurrence of copy-number gains at Xp11.22,[7,26,27] including this study, indicates that this region might be more amenable to DNA breakage or replication-fork stalling, by which NHEJ or FoSTeS (or MMBIR) repair mechanisms, respectively, occur more frequently.[28] A role for the nearby LCR-rich *MAGE-XAGE-SSX* complex architectural region (51.6–52.8 Mb) can be hypothesized, predisposing the Xp11.22 region to nonrecurrent rearrangements.[29] For the five nonrecurrent simple junctions with microhomology, we expect MMBIR to be involved. First, at our sequenced junctions, we did not find an information scar reflecting the incorrect repair that usually occurs after rejoining broken DNA fragments, which is typical for NHEJ.[30] Second, we also did not detect any inserted sequences from neighboring locations, making FoSTeS a less probable explanation also, given that these rearrangements are generally more complex.[31] Surprisingly, from the 12 identified nonrecurrent *HUWE1* microduplications, only the one in SB1 resulted from NAHR between directly oriented LCRs (5.0 and 5.6 kb in size), indicating that NAHR is not expected to be a frequent cause for generating a recurrent copy-number gain at Xp11.22. This rare occurrence could be explained by the relatively imperfect structural similarity of both. Indeed, a minimal efficient processing segment is required for NAHR to take place most efficiently.[32,33]

Reciprocal deletions at this genomic interval including *HUWE1* have never been reported, most likely due to prenatal lethality in males, which is in line with the high rate of perinatal lethality in *Huwe1* knockout mice.[21,22] For the same reason, we speculate that *HUWE1* nonsense or severe functionally damaging mutations will not be viable in male subjects.

In silico analysis of the flanking 600 bp sequences at the ten breakpoints revealed a significantly higher GC content compared to the 50 random control regions (48% versus 36%). A similar result was observed for other chromosomal rearrangements.[12,34,35] Second, an increased presence of DNA polymerase A/B frameshift hot spots and translin binding 2 sites has been reported in translocation breakpoints as well.[36–38] The significantly enriched murine parvovirus recombination hot spot CTWTTY is striking, but for now its relevance is still unknown. Third, our study showed a significant increase of palindromes and, to a lesser extent, G-quartet structures at the breakpoint sites compared to the control regions. Both non-B-DNA structures have already been implicated in several rearrangements,[39,40] and it is now well accepted that these structures often coincide with breakpoint sites.[41] Importantly, the resulting non-B-DNA-driven rearrangements can be generated by replication-based as well as replication-independent mechanisms.[42] In view of the replication-driven rearrangement, it was recently demonstrated that aphidicolin-induced DNA replication stress-induced CNVs in fibroblasts can result in head-to-tail tandem duplications with microhomology at the junctions,[43] as has been detected in our families. At least in family SB1, the involvement of LCRs at both breakpoints demonstrates that NAHR is the mechanism for this genomic aberration.

We also detected an unreported copy-number gain of about 73 kb in four unrelated ID families. The "partial *HUWE1* copy-number gain" includes *HUWE1* exon 29 to the 3′ region of this gene as well as *miR-98* and *let-7f-2*. It does not segregate with the ID phenotype in at least three families, because it was also present in unaffected males (I.1 in family HF; I.1 in family CC1; II.4 in VS1) and was not found in the equally affected brother (II.1) of family ON2. Therefore, this partial *HUWE1* copy-number gain does not seem to affect the function of HUWE1, most likely because it leaves one copy of the gene intact. We did not find evidence for increased expression of the duplicated part of *HUWE1* (exon 29 to 3′ UTR) nor for the generation of a *HUWE1* fusion gene.

The two duplicated microRNAs are located in intron 59 of *HUWE1*. They belong to the same microRNA family and are highly expressed in melanocytes and melanomas, as found in miRBase. Although some reports propose a function for this class of microRNAs in immunity[44,45] or tumorigenesis,[46,47] it is clear from our study that an increased copy number of one or both microRNAs does not affect normal cognitive development. Lack of segregation in the families points to the fact that the actual cause of the ID phenotype has to be located elsewhere. Indeed, in the twins of family CC1, a 7 Mb causal deletion was detected. However, careful analysis of cancer- or immune-related endophenotypes or disorders later in life due to duplication of both microRNAs is warranted. Thus far, in our cohort, we cannot yet relate this apparent polymorphic CNV with any disease condition. It is a very rare CNV, in that it is not yet reported in the database of Genomic Variants (DGV), in DECIPHER or ECARUCA. However, ISCA reports three similar cases in two families for which its significance with disease is unknown. Finally, this subtle copy-number gain is not a founder mutation, because we find A and C calls for SNP rs266786, located within the duplication, in the families

with geographical locations in three different continents (Table S2).

The recurrent partial *HUWE1* copy-number gain is caused by NAHR between two adjacent DNA elements of the TcMAR-Tigger2 subtype. Tiggers are DNA transposons that are flanked by terminal inverted repeats and encode a transposase that binds to the TIRs and induces cutting and pasting of the element.[48] NAHR has been reported between LINEs, short interspersed nuclear elements (SINEs), and long-terminal-repeat elements, which are all retrotransposons, but the use of DNA elements as the apparent homologous templates for NAHR has not been reported. It will be interesting to search for CNVs that are flanked by similar DNA elements in order to establish the use of DNA elements as an unreported recombination driver.

Taken together, this study identifies *HUWE1* as the dosage-sensitive gene for which a 2-fold increase in expression results in mild to moderate ID in males. In concordance with homozygosity within the duplications, the copy-number gains suggest an intrachromosomal meiotic repair event. Our data provide evidence for recombination- as well as replication-based processes, most likely through NAHR and MMBIR, respectively. In the two de novo events that we could determine, the duplication has arisen on the paternal X chromosome, suggesting these events did happen during spermatogenesis. Finally, we report on NAHR between DNA elements, particularly of the TcMAR-Tigger2 subfamily, which resulted in a recurrent polymorphic copy-number gain within *HUWE1* that does not affect its expression. As a result, this polymorphic CNV excludes both microRNAs for a role in ID. We postulate the occurrence of additional DNA-element-driven copy-number changes resulting in disease or being involved in evolution.

Supplemental Data

Supplemental Data include five figures and two tables and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

We thank the families for their willingness to participate in this study. The microarray analyses were performed by the VIB Nucleomics Core. This study is supported by grants from the Fondo de Investigaciones Sanitarias (Spanish Ministry of Health)/FEDER (Fondo Europeo de Desarrollo Regional) (PI080648), from the CNPq (National Council for Scientific and Technological Development) (565705/2008-3) and FAPERJ (Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro) (E-26/110.765/2010 and E-26/103.215/2011) in Brazil, and from the Women's and Children's Hospital Foundation in Australia. S.M. is supported by an IIS Fundación La Fe-Bancaja fellowship (Spain).

Received: April 24, 2012
Revised: May 21, 2012
Accepted: June 21, 2012
Published online: July 26, 2012

Web Resources

The URLs for data presented herein are as follows:

BL2SEQ, http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::bl2seq
DECIPHER, http://decipher.sanger.ac.uk/
DGV, http://projects.tcag.ca/variation/
ECARUCA, http://umcecaruca01.extern.umcn.nl:8080/ecaruca/ecaruca.jsp
ETANDEM, http://bioweb.pasteur.fr/seqanal/interfaces/etandem.html
FUZZNUC, http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::fuzznuc
GEECEE, http://bioweb.pasteur.fr/seqanal/interfaces/geecee.html
NCBI GEO database, http://www.ncbi.nlm.nih.gov/geo/
ISCA, https://www.iscaconsortium.org/
miRBase, http://www.mirbase.org/
Non-B DB, http://nonb.abcc.ncifcrf.gov
Nucleomics Core, http://www.nucleomics.be/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
PALINDROME, http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::palindrome
QGRS, http://bioinformatics.ramapo.edu/QGRS/analyze.php
RepeatMasker, http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker
UCSC Genome Browser, http://genome.ucsc.edu/
Unigene, http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene
ZHUNT, http://gac-web.cgrb.oregonstate.edu/zDNA/

Accession Numbers

The GEO database accession number for the data of the eight hybridizations on the custom oligo arrays reported in this paper is GSE32945.

References

1. Ropers, H.H. (2008). Genetics of intellectual disability. Curr. Opin. Genet. Dev. *18*, 241–250.
2. Gécz, J., Shoubridge, C., and Corbett, M. (2009). The genetic landscape of intellectual disability arising from chromosome X. Trends Genet. *25*, 308–316.
3. Van Esch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K., Lugtenberg, D., Bienvenu, T., Jensen, L.R., Gécz, J., et al. (2005). Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. Am. J. Hum. Genet. *77*, 442–453.
4. Meins, M., Lehmann, J., Gerresheim, F., Herchenbach, J., Hagedorn, M., Hameister, K., and Epplen, J.T. (2005). Submicroscopic duplication in Xq28 causes increased expression of the MECP2 gene in a boy with severe mental retardation and features of Rett syndrome. J. Med. Genet. *42*, e12.
5. Carvalho, C.M., Zhang, F., Liu, P., Patel, A., Sahoo, T., Bacino, C.A., Shaw, C., Peacock, S., Pursley, A., Tavyev, Y.J., et al. (2009). Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching. Hum. Mol. Genet. *18*, 2188–2203.
6. Vandewalle, J., Van Esch, H., Govaerts, K., Verbeeck, J., Zweier, C., Madrigal, I., Mila, M., Pijkels, E., Fernandez, I., Kohlhase, J.,

et al. (2009). Dosage-dependent severity of the phenotype in patients with mental retardation due to a recurrent copy-number gain at Xq28 mediated by an unusual recombination. Am. J. Hum. Genet. *85*, 809–822.
7. Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C., Bauters, M., Govaerts, K., Vandeleur, L., Van Esch, H., et al. (2008). Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. Am. J. Hum. Genet. *82*, 432–443.
8. Ofman, R., Ruiter, J.P., Feenstra, M., Duran, M., Poll-The, B.T., Zschocke, J., Ensenauer, R., Lehnert, W., Sass, J.O., Sperl, W., and Wanders, R.J. (2003). 2-Methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency is caused by mutations in the HADH2 gene. Am. J. Hum. Genet. *72*, 1300–1307.
9. Lenski, C., Kooy, R.F., Reyniers, E., Loessner, D., Wanders, R.J., Winnepenninckx, B., Hellebrand, H., Engert, S., Schwartz, C.E., Meindl, A., and Ramser, J. (2007). The reduced expression of the HADH2 protein causes X-linked mental retardation, choreoathetosis, and abnormal behavior. Am. J. Hum. Genet. *80*, 372–377.
10. Shaw, C.J., and Lupski, J.R. (2004). Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum. Mol. Genet. *13* (Spec No 1), R57–R64.
11. Simmons, A.D., Carvalho, C.M., and Lupski, J.R. (2012). What have studies of genomic disorders taught us about our genome? Methods Mol. Biol. *838*, 1–27.
12. Bauters, M., Van Esch, H., Friez, M.J., Boespflug-Tanguy, O., Zenker, M., Vianna-Morgante, A.M., Rosenberg, C., Ignatius, J., Raynaud, M., Hollanders, K., et al. (2008). Nonrecurrent MECP2 duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair. Genome Res. *18*, 847–858.
13. Lee, J.A., Inoue, K., Cheung, S.W., Shaw, C.A., Stankiewicz, P., and Lupski, J.R. (2006). Role of genomic architecture in PLP1 duplication causing Pelizaeus-Merzbacher disease. Hum. Mol. Genet. *15*, 2250–2265.
14. del Gaudio, D., Fang, P., Scaglia, F., Ward, P.A., Craigen, W.J., Glaze, D.G., Neul, J.L., Patel, A., Lee, J.A., Irons, M., et al. (2006). Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. Genet. Med. *8*, 784–792.
15. Stallings, R.L., Yoon, K., Kwek, S., and Ko, D. (2007). The origin of chromosome imbalances in neuroblastoma. Cancer Genet. Cytogenet. *176*, 28–34.
16. Madrigal, I., Rodríguez-Revenga, L., Armengol, L., González, E., Rodriguez, B., Badenas, C., Sánchez, A., Martínez, F., Guitart, M., Fernández, I., et al. (2007). X-chromosome tiling path array detection of copy number variants in patients with chromosome X-linked mental retardation. BMC Genomics *8*, 443.
17. Whibley, A.C., Plagnol, V., Tarpey, P.S., Abidi, F., Fullston, T., Choma, M.K., Boucher, C.A., Shepherd, L., Willatt, L., Parkin, G., et al. (2010). Fine-scale survey of X chromosome copy number variants and indels underlying intellectual disability. Am. J. Hum. Genet. *87*, 173–188.
18. Allen, R.C., Zoghbi, H.Y., Moseley, A.B., Rosenblatt, H.M., and Belmont, J.W. (1992). Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. Am. J. Hum. Genet. *51*, 1229–1239.
19. Zhong, Q., Gao, W., Du, F., and Wang, X. (2005). Mule/ARF-BP1, a BH3-only E3 ubiquitin ligase, catalyzes the polyubiquitination of Mcl-1 and regulates apoptosis. Cell *121*, 1085–1095.
20. Chen, D., Kon, N., Li, M., Zhang, W., Qin, J., and Gu, W. (2005). ARF-BP1/Mule is a critical mediator of the ARF tumor suppressor. Cell *121*, 1071–1083.
21. D'Arca, D., Zhao, X., Xu, W., Ramirez-Martinez, N.C., Iavarone, A., and Lasorella, A. (2010). Huwe1 ubiquitin ligase is essential to synchronize neuronal and glial differentiation in the developing cerebellum. Proc. Natl. Acad. Sci. USA *107*, 5875–5880.
22. Zhao, X., D' Arca, D., Lim, W.K., Brahmachary, M., Carro, M.S., Ludwig, T., Cardo, C.C., Guillemot, F., Aldape, K., Califano, A., et al. (2009). The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. Dev. Cell *17*, 210–221.
23. Zhao, X., Heng, J.I., Guardavaccaro, D., Jiang, R., Pagano, M., Guillemot, F., Iavarone, A., and Lasorella, A. (2008). The HECT-domain ubiquitin ligase Huwe1 controls neural differentiation and proliferation by destabilizing the N-Myc oncoprotein. Nat. Cell Biol. *10*, 643–653.
24. Medrano, S., and Scrable, H. (2005). Maintaining appearances—the role of p53 in adult neurogenesis. Biochem. Biophys. Res. Commun. *331*, 828–833.
25. Lovett, S.T. (2004). Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. Mol. Microbiol. *52*, 1243–1253.
26. Giorda, R., Bonaglia, M.C., Beri, S., Fichera, M., Novara, F., Magini, P., Urquhart, J., Sharkey, F.H., Zucca, C., Grasso, R., et al. (2009). Complex segmental duplications mediate a recurrent dup(X)(p11.22-p11.23) associated with mental retardation, speech delay, and EEG anomalies in males and females. Am. J. Hum. Genet. *85*, 394–400.
27. Bonnet, C., Grégoire, M.J., Brochet, K., Raffo, E., Leheup, B., and Jonveaux, P. (2006). Pure de-novo 5 Mb duplication at Xp11.22-p11.23 in a male: phenotypic and molecular characterization. J. Hum. Genet. *51*, 815–821.
28. Hastings, P.J., Lupski, J.R., Rosenberg, S.M., and Ira, G. (2009). Mechanisms of change in gene copy number. Nat. Rev. Genet. *10*, 551–564.
29. Stankiewicz, P., Shaw, C.J., Dapper, J.D., Wakui, K., Shaffer, L.G., Withers, M., Elizondo, L., Park, S.S., and Lupski, J.R. (2003). Genome architecture catalyzes nonrecurrent chromosomal rearrangements. Am. J. Hum. Genet. *72*, 1101–1116.
30. Lieber, M.R. (2008). The mechanism of human nonhomologous DNA end joining. J. Biol. Chem. *283*, 1–5.
31. Lee, J.A., Carvalho, C.M., and Lupski, J.R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell *131*, 1235–1247.
32. Waldman, A.S., and Liskay, R.M. (1988). Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol. Cell. Biol. *8*, 5350–5357.
33. Rubnitz, J., and Subramani, S. (1984). The minimum amount of homology required for homologous recombination in mammalian cells. Mol. Cell. Biol. *4*, 2253–2258.
34. Hannes, F., Drozniewska, M., Vermeesch, J.R., and Haus, O. (2010). Duplication of the Wolf-Hirschhorn syndrome critical region causes neurodevelopmental delay. Eur. J. Med. Genet. *53*, 136–140.

35. Albano, F., Anelli, L., Zagaria, A., Coccaro, N., Casieri, P., Rossi, A.R., Vicari, L., Liso, V., Rocchi, M., and Specchia, G. (2010). Non random distribution of genomic features in breakpoint regions involved in chronic myeloid leukemia cases with variant t(9;22) or additional chromosomal rearrangements. Mol. Cancer *9*, 120.
36. Abeysinghe, S.S., Chuzhanova, N., Krawczak, M., Ball, E.V., and Cooper, D.N. (2003). Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. Hum. Mutat. *22*, 229–244.
37. Gajecka, M., Gentles, A.J., Tsai, A., Chitayat, D., Mackay, K.L., Glotzbach, C.D., Lieber, M.R., and Shaffer, L.G. (2008). Unexpected complexity at breakpoint junctions in phenotypically normal individuals and mechanisms involved in generating balanced translocations t(1;22)(p36;q13). Genome Res. *18*, 1733–1742.
38. Kasai, M., Matsuzaki, T., Katayanagi, K., Omori, A., Maziarz, R.T., Strominger, J.L., Aoki, K., and Suzuki, K. (1997). The translin ring specifically recognizes DNA ends at recombination hot spots in the human genome. J. Biol. Chem. *272*, 11402–11407.
39. Kurahashi, H., Inagaki, H., Ohye, T., Kogo, H., Kato, T., and Emanuel, B.S. (2006). Palindrome-mediated chromosomal translocations in humans. DNA Repair (Amst.) *5*, 1136–1145.
40. Wu, Y., and Brosh, R.M., Jr. (2010). G-quadruplex nucleic acids and human disease. FEBS J. *277*, 3470–3488.
41. Bacolla, A., Jaworski, A., Larson, J.E., Jakupciak, J.P., Chuzhanova, N., Abeysinghe, S.S., O'Connell, C.D., Cooper, D.N., and Wells, R.D. (2004). Breakpoints of gross deletions coincide with non-B DNA conformations. Proc. Natl. Acad. Sci. USA *101*, 14162–14167.
42. Zhao, J., Bacolla, A., Wang, G., and Vasquez, K.M. (2010). Non-B DNA structure-induced genetic instability and evolution. Cell. Mol. Life Sci. *67*, 43–62.
43. Arlt, M.F., Mulle, J.G., Schaibley, V.M., Ragland, R.L., Durkin, S.G., Warren, S.T., and Glover, T.W. (2009). Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. Am. J. Hum. Genet. *84*, 339–350.
44. Liu, Y., Chen, Q., Song, Y., Lai, L., Wang, J., Yu, H., Cao, X., and Wang, Q. (2011). MicroRNA-98 negatively regulates IL-10 production and endotoxin tolerance in macrophages after LPS stimulation. FEBS Lett. *585*, 1963–1968.
45. Wang, S., Tang, Y., Cui, H., Zhao, X., Luo, X., Pan, W., Huang, X., and Shen, N. (2011). Let-7/miR-98 regulate Fas and Fas-mediated apoptosis. Genes Immun. *12*, 149–154.
46. Wendler, A., Keller, D., Albrecht, C., Peluso, J.J., and Wehling, M. (2011). Involvement of let-7/miR-98 microRNAs in the regulation of progesterone receptor membrane component 1 expression in ovarian cancer cells. Oncol. Rep. *25*, 273–279.
47. Legesse-Miller, A., Elemento, O., Pfau, S.J., Forman, J.J., Tavazoie, S., and Coller, H.A. (2009). let-7 Overexpression leads to an increased fraction of cells in G2/M, direct down-regulation of Cdc34, and stabilization of Wee1 kinase in primary fibroblasts. J. Biol. Chem. *284*, 6605–6609.
48. Smit, A.F., and Riggs, A.D. (1996). Tiggers and DNA transposon fossils in the human genome. Proc. Natl. Acad. Sci. USA *93*, 1443–1448.