The ancestral gene for transcribed, low-copy repeats in the Prader–Willi/Angelman region encodes a large protein implicated in protein trafficking, which is deficient in mice with neuromuscular and spermiogenic abnormalities

Yonggang Ji, Mitchell J. Walkowicz, Karin Buiting, Dabney K. Johnson, Rocio E. Tarvin, Eugene M. Rinchik, Bernhard Horsthemke, Lisa Stubbs and Robert D. Nicholls

Department of Genetics, Case Western Reserve University School of Medicine and Center for Human Genetics, University Hospitals of Cleveland, 10900 Euclid Avenue, Cleveland, OH 44106-4955, USA; 1Life Sciences Division, Oak Ridge National Laboratory, PO Box 2009, Oak Ridge, TN 37831-8077, USA and 2Institut für Humangenetik, Universitätsklinikum Essen, D-45122 Essen, Germany

Received December 4, 1998; Revised and Accepted December 21, 1998

Transcribed, low-copy repeat elements are associated with the breakpoint regions of common deletions in Prader–Willi and Angelman syndromes. We report here the identification of the ancestral gene (HERC2) and a family of duplicated, truncated copies that comprise these low-copy repeats. This gene encodes a highly conserved giant protein, HERC2, that is distantly related to p532 (HERC1), a guanine nucleotide exchange factor (GEF) implicated in vesicular trafficking. The mouse genome contains a single Herc2 locus, located in the jdf2 (juvenile development and fertility-2) interval of chromosome 7C. We have identified single nucleotide splice junction mutations in Herc2 in three independent N-ethyl-N-nitrosourea-induced jdf2 mutant alleles, each leading to exon skipping with premature termination of translation and/or deletion of conserved amino acids. Therefore, mutations in Herc2 lead to the neuromuscular secretory vesicle and sperm acrosome defects, other developmental abnormalities and juvenile lethality of jdf2 mice. Combined, these findings suggest that HERC2 is an important gene encoding a GEF involved in protein trafficking and degradation pathways in the cell.

INTRODUCTION

Human chromosome 15q11–q13 is characterized by several unusual genetic properties. Prader–Willi and Angelman syndromes (PWS and AS, respectively) are clinically distinct neurobehavioral disorders that result from a different parental origin for similar genetic abnormalities in 15q11–q13 (1). These two syndromes result from the loss of function of oppositely imprinted genes located within the proximal 2 Mb of the 15q11–q13 region. The majority (70–75%) of cases of PWS and AS are caused by ∼4 Mb deletions that include the imprinted domain and an ∼1–2 Mb non-imprinted domain (1). The vast majority (>95%) of these deletions in PWS and AS are indistinguishable in extent, with breakpoints clustered within defined regions. There are two common proximal and one distal deletion breakpoint regions that appear to reflect hotspots for recombination (Fig. 1a; 2–5).

Low-copy repeat elements have been identified in the vicinity of the three deletion breakpoint hotspots using molecular and cytological methods (Fig. 1a; 5). These repeat sequences (termed D15F37) were first identified by the microdissected MN7 clone (6). Subsequent YAC mapping experiments located at least four D15F37 copies in the vicinity of the three 15q11–q13 deletion breakpoint regions and at least two copies in 16p11.2 (5,6). The D15F37 repeat is expressed, with predominant 6–7 kb transcripts and a minor ∼15 kb transcript, in every tissue tested (6). In another study, DNA fragments positionedally cloned from the PWS/AS deletion breakpoint regions were also found to represent a transcribed low-copy repeat, present in similar locations but without sequence overlap to D15F37 (J.M. Amos-Landgraf et al., in preparation). However, the similar map positions suggest that these two classes of sequences may be part of a larger low-copy repeat, which may play a role in the mechanism of the common deletion in PWS and AS.
Human chromosome 15q11–q13 is homologous to mouse chromosome 7C (1). Non-imprinted and imprinted genes in these regions map in the same respective order in human and mouse, but lie in opposite centromere–telomere orientations (Fig. 1). In contrast to human, the mouse has a single locus homologous to D15F37, which maps close to and centromeric to the pink-eyed dilution (p) locus (7–10). The region homologous to the human distal breakpoint region was therefore predicted to lie in the same region as a previously defined locus, termed jdf2 (juvenile development and fertility) (11) or rjs (genty jerky gterle) (10). The jdf2 locus was first defined by the neurological and spermatogenesis defects observed in homozygous mice with the p<sup>6H</sup> and p<sup>25H</sup> radiation-induced mutations (9,12–15). The phenotypic features of jdf2 mice include running, a nervous jerky gait and tremor, male sterility, female semi-sterility and a reduced lifespan with juvenile lethality. This complex, pleiotropic phenotype was suggested on the basis of N-ethyl-N-nitrosourea (ENU) mutagenesis and complementation analyses to be due to dysfunction of a single gene (11). Positional cloning of a candidate gene has recently been reported (10,16), including the identification of a single intragenic deletion (10).

Here we demonstrate that the low-copy repeat sequences previously identified in human chromosome 15q11–q13 represent components of the same gene family, with an ancestral locus, HERC2, and duplicated, truncated copies. The large HERC2 protein has domains that allow prediction of function as a guanine nucleotide exchange factor and an E3 ubiquitin ligase potentially involved in protein trafficking and degradation pathways in the cell. Our identification of single base pair splice site mutations in three independent ENU-induced jdf2 mouse mutants provides definitive proof that Herc2 mutations underlie the complex jdf2 multisystem disorder.

RESULTS

A novel, highly conserved gene from the low-copy repeats flanking 15q11–q13

Partial sequence analyses of genomic clones isolated by positional cloning from low-copy repeats flanking 15q11–q13 identified a 162 bp sequence from λ6A1 (J.M. Amos-Landgraf et al., in preparation) with 91% identity to an expressed human sequence (EST 05046), flanked by potential splice sites. A 138 bp PCR probe (Fig. 2a, probe A) from the putative exon detected a single large transcript of ~15 kb in all human tissues tested, and, by Southern blot analysis, was shown to be evolutionarily conserved in mouse (data not shown). This probe was used to isolate a partial cDNA clone encoding a 2 kb open reading frame (ORF) with 90% identity to both the initial exon probe and EST 05046, identifying a gene family of related sequences. Overlapping clones comprising 6.9 kb of sequence, including the 2 kb partial human cDNA, were then isolated. The 5′-end of this 6.9 kb cDNA sequence contained a 61 bp 5′-untranslated region (5′-UTR), a putative translational initiation codon in the context of an adequate Kozak consensus sequence (17) and an ORF through the remainder of the cDNA sequence (Fig. 2a). The 5′ cDNA sequence likely represents the bona fide 5′-end, based on the sequence of seven 5′-RACE (rapid amplification of cDNA ends) clones (data not shown). Furthermore, the 5′-GC-rich 30 bp sequence is homologous to a CpG-island that is duplicated in genomic copies of the 15q11–q13 low-copy repeats and is flanked by a consensus splice donor motif (J.M. Amos-Landgraf et al., in preparation), consistent with the idea that it represents exon 1.

The 3′-portion of the 6.9 kb sequence, however, was 92% identical to a D15F37 cDNA (GenBank accession no. X69636), previously shown to identify a strong band of transcripts at 6–7 kb and a weakly detected transcript at ~15 kb (6). We show below that the 6.9 kb partial cDNA is part of the ~15 kb transcript, termed HERC2 (HGNC-approved nomenclature, for HERC2 ortholog (Fig. 2a) to provide a guideline for isolation of 3′ human HERC2 cDNA sequences. The mouse Herc2 gene was cloned by a combination of cDNA library screening, subtraction cloning from wild-type versus radiation-induced jdf2 mutant mRNAs, as well as 5′- and 3′-RACE (Materials and Methods), and was also independently isolated by Lehman et al. (10). The full-length mouse Herc2 transcript (15 247 bp in length) has a short 5′-UTR of 71 bp, a single ORF of 14 511 bp and a 665 bp 3′-UTR with two polyadenylation signals (Fig. 2b).

The mouse Herc2 sequence was used to search the EST databases for human sequences that showed significant homo-
Figure 2. The human and mouse HERC2 transcripts and encoded polypeptides. (a) Schematic of the human and mouse HERC2 transcripts. Shaded boxes represent the open reading frame, arrows the 5'→3' orientation and (A)_n the polyadenylated tails. Numbers above the transcripts represent the coordinates for the 5'-UTR, translation initiation and stop codons and 3'-UTR, respectively. A, PCR probe derived from a HERC2-homologous sequence in λ6A1; B–E, HERC2 cDNA probes. A minimal contig of cDNA clones identified for human HERC2 (GenBank accession no. AF071172) and mouse Herc2 (AF071173) are shown by thick lines with tissue sources (b, brain; fb, fetal brain; p, pancreas; t, testis; c, cerebellum) and respective nucleotide numbers. (b) Schematic of the human HERC2 and HERC1 (p532; 22) proteins and a putative protein from a full-length human EST, HERC3 (D25215). Over the C-terminal ~850 amino acids, HERC2 (3952–4785) is 40.4% identical to HERC1 (3997–4834) and 22.9% identical to HERC3 (103–1041), whereas the latter two proteins are 25.8% identical. Over the N-terminal 2229 amino acids of HERC2, HERC1 shows 22 blocks of 21–58 amino acids, each of 22–58% identity and 51–79% similarity and each in the same respective location in the two proteins. Shaded boxes represent putative functional motifs in the proteins. The locations of HERC2 structural changes identified in the ENU-induced jdf2 mutants are also shown (Fig. 5). Open triangles represent in-frame exon skipping events (jdf2^932SJ and jdf2^322SJ), while closed triangles represent frameshift mutations (jdf2^1971SJ and jdf2^322SJ).

Structural features of the HERC2 protein

The human and mouse HERC2 genes encode huge proteins of 4834 and 4836 amino acids, respectively, each with a predicted molecular mass of 528 kDa. Amino acid sequence comparison of the two proteins shows 95% identity and 99% similarity overall (data not shown). This striking and unusual level of human–mouse conservation clearly attests to the likely functional importance of the HERC2 gene product in both species. Further analysis identified several protein motifs (Fig. 2b; 10,16). Three RCC1-like domains (RLDs) are identified in HERC2, spanning amino acid residues 423–783 (RLD-1), 2959–3331 (RLD-2) and 3952–4323 (RLD-3) in human. Each RLD in HERC2 contains seven conserved monomeric repeats of ~60 residues each, as in RCC1 (18). HERC2 also contains a ZZ-type putative zinc finger motif (19), with six conserved cysteine residues and two outlying histidine residues that might contribute to binding of Zn^{2+}. Other motifs found in HERC2 include a C-terminal HECT or E3 ubiquitin ligase domain (10,16,20,21) and several potential phosphorylation sites for tyrosine kinase and cAMP- and cGMP-dependent protein kinases. The overall structure of HERC2 resembles that of the p532 protein (Fig. 2b; 22) (HGNC-approved nomenclature: HERC1). While HERC1 (p532) lacks RLD-2 and the zinc finger motif contained in HERC2, it contains seven Gβ-(WD40) repeats (β subunit of heterotrimeric G-proteins) which are believed to play a role in protein–protein interactions. The overall similarity of HERC2 and HERC1 (p532) is more extensive, spanning two large segments, and the C-terminal segment of each also resembles a third putative protein, HERC3 (HGNC-approved nomenclature; GenBank accession no. D25215) (Fig. 2b). Combined, the structural and sequence data suggest that HERC2, HERC1 (p532)
and HERC3 share an ancient evolutionary origin from a common ancestral gene.

Expression analyses of the human and mouse HERC2 genes

Use of a 1.1 kb human HERC2 cDNA probe (Fig. 2a, probe C) on northern blots identified a single 15.5 kb transcript in all human and mouse tissues tested, with high levels in fetal tissues and adult skeletal muscle, heart, ovary, testis and brain. Occasionally, smaller transcripts were detected in some human tissues at a very low level, but were not observed in mouse. The MN7 (D15F37) microdissection clone (6) has 99% sequence identity with HERC2 (nucleotide sequence 5526–5695) and detected multiple transcripts of 6–7 and 15.5 kb in length in all human tissues, as previously seen (6), with a low level of the 6–7 kb transcripts in fetal brain. In contrast, a single 15.5 kb transcript was detected by this probe in mouse (7–9), identical in size to that detected by the HERC2-specific probe. Combined, these data suggest that the MN7-related 6–7 kb transcripts are homologous to just part of a larger HERC2 gene.

Since HERC2 is located adjacent to an imprinted domain, we also tested imprinting of the human gene using somatic cell hybrids containing a chromosome 15 of maternal or paternal origin (23). Expression was equal from the maternal or paternal chromosome 15 (data not shown), which is consistent with a recessive mouse phenotype (see below; 10).

The ancestral HERC2 gene maps distal to the P gene in 15q13

Southern blot analysis using probes derived from the 5′-portion of the HERC2 cDNA (Fig. 2a, probes B–D) detected multiple fragments in YACs spanning the 15q11–q13 proximal and distal low-copy D15F37 repeat regions, as well as in human genomic and chromosome 15 hybrid cell line DNA (data not shown). The fragments detected by the HERC2 probes represent a combination of multiple exons and the duplicated HERC2-related loci at both ends of the PWS and AS common deletion breakpoint region (5). However, the extreme 3′ HERC2 cDNA probe (Fig. 2a, probe E) detected only a single, unique restriction fragment in human DNA (Fig. 3a), indicating that the 3′-portion of HERC2 is not duplicated. We hybridized this probe to Southern blots containing DNA from PWS and AS patients with the common deletion and normal controls. The signal produced by this probe in eight PWS patients (Fig. 3a) and two AS patients (data not shown) was only 50% as intense as that produced in controls, verifying that at least the 3′-end of HERC2 is within the common PWS/AS 15q11–q13 deletion. Furthermore, probe E (3′ HERC2) detected a common restriction fragment on two YACs containing the human P gene, but did not detect sequences in D15F37 repeat YACs from proximal or distal 15q11–q13 that did not contain P (Fig. 3b). A human genomic BAC clone positive for the HERC2 3′-UTR also contains the P promoter (Fig. 3b). Combined, these data demonstrate that HERC2 maps very close to P in human 15q13.

The 15q11–q13 low-copy repeats comprise duplicated, truncated copies of HERC2

We performed sequence comparisons of the human HERC2 15.5 kb cDNA with all the currently isolated 6–7 kb D15F37 and D16F37 transcripts (5,6) and a chromosome 16p11.2 genomic sequence (D16F37; Fig. 4). For the most part, the 6–7 kb transcripts are closely related to HERC2 (95–97% identical; see Materials and Methods), although the extreme 3′-ends of the 6–7 kb transcripts do not recognize a counterpart in the HERC2 cDNA sequence and some 5′ HERC2 sequences are not present in the 6–7 kb transcripts. Therefore, we rename the D15F37 gene family, beginning with HERC2 for the ancestral locus (see above). By analysis of diagnostic nucleotides, HERC2 corresponds to D15F37S1 and the other transcripts correspond to paralogous loci D15F37S2–S4 (Fig. 4; 5). HERC2P1–P3 therefore replace D15F37S2–S4 (Fig. 4; P for likely pseudogene; see below). HERC2P4 corresponds to the chromosome 16p11.2 BAC clone (AC002041, D16F37S5) and HERC2P5 to the other chromosome 16 locus (D16F37S6) (5). The 6–7 kb HERC2P1–P3 transcripts initiate from duplicated copies of the putative HERC2 CpG-island promoter (data not shown). However, the ORFs potentially encoded by the HERC2P1–P3 transcripts do not contain any of the known functional polypeptide motifs found in HERC2 (Fig. 2b, 4). In the case of the HERC2P4 genomic sequence, the absence of these motifs appears to have arisen by five genomic deletions within a duplicated copy of the ancestral HERC2 locus (Fig. 4). The members of the HERC2P1–P5 subfamily also contain premature stop codons that would produce significantly truncated proteins compared with that encoded by HERC2 (Fig. 4). These transcripts therefore most likely represent transcribed pseudogenes that have evolved from HERC2 or, alternatively, they may have acquired new functions.

In this study, we also identified several genomic sequences (λ6A1 and λ11A1) which show ~90% identity to portions of HERC2 not present in the D15F37 transcripts. EST clone 05046 (GenBank accession no. AF071178), described above, contains a 309 bp sequence with 91% nucleotide identity to a portion of the HERC2 cDNA (nt 2625–2933). The 3′-end of EST 05046 contains a 210 bp Alu sequence. This 1131 bp EST showed 100% identity to two exons in the human genomic λ11A1 clone (J.M. Amos-Landgraf et al., in preparation), proving the genomic origin of this truncated, and presumably non-functional, family member. The 5′ loci represented by 6A1 and 11A1 have been defined as HERC2P6 and HERC2P7, respectively. These data show that the HERC2 ancestral gene and related low-copy repeat sequences have undergone several independent duplications and truncations to form the gene family.

Identification of Herc2 point mutations in ENU-induced jdf2 mutants

In contrast to human, the mouse haploid genome contains a single copy of the Herc2 gene. Analyses of large radiation-induced deletions associated with the jdf2 (rjs) phenotype have identified Herc2 gene rearrangements in most mutants examined (10; M.J. Walkowicz et al., in preparation), including a single published interstitial gene deletion (10). While the latter result suggests a role for this gene in the rjs (jdf2) phenotype (10), proof of an etiological role for Herc2 is still necessary. Therefore, mutation analyses were performed to compare the Herc2 gene from ENU-induced jdf2 mutants with that of the control parental strain (BJR), since the chemical mutagen ENU exclusively induces point mutations under the protocol used (24). These studies identified Herc2 point mutations in the three ENU-induced jdf2 mutants analyzed to date.
Figure 3. The ancestral HERC2 gene maps adjacent to the P gene in 15q13. (a) HERC2 is deleted in the common PWS 15q11–q13 deletion. DNA from cell lines was digested with HindIII and hybridized with probe E (Fig. 2) or a single-copy control probe from chromosome 5q (I.A. Gray and R.D. Nicholls, unpublished data). The signal intensity was quantitated by scanning and normalization to the average of lanes 1 and 2 and the normalized ratios are shown. A decrease in intensity from 0.82–1.39 for control, UPD and non-deletion samples to 0.39–0.63 for probe E is seen in all PWS deletion patients. DNAs are: N, normal (lanes 1 and 2, cell lines 12C and 12B, respectively); del, PWS patients with a 15q11–q13 deletion (lanes 3–5, 8–10, 12 and 13, cell lines 20A, 19A, 17A, GM09819, GM11382, GM11385, GM06024 and PWS109, respectively); UPD, PWS patients with uniparental disomy (lanes 6 and 7, cell lines 8A and 7A, respectively); non-del, PWS-like non-deletion patient (lane 11, cell line GM04297). (b) Mapping of HERC2 to 15q13. A minimal YAC contig and a BAC clone are shown. Circles represent loci identified by STS and hybridization mapping.

Figure 4. The ancestral HERC2 gene is the precursor of a family of duplicated, truncated transcripts. Schematic comparison of HERC2, the cloned transcripts from duplicated loci HERC2P1–P3 and a chromosome 16p11.2 genomic copy (HERC2P4). Boxes represent homologous sequences in each cDNA clone, while lines represent sequences not present in other clones. Open arrows are potential translation initiation codons, filled arrows potential stop codons and arrowheads polyadenylation signals. Over 2270 nt of shared sequence in the MN7 region, HERC2 is 97.1, 97.3, 97.2 and 95.0% identical to HERC2P1, HERC2P2, HERC2P3 and HERC2P4, respectively. HERC2P1–P3 are 99.3–99.6% identical to each other (95% to HERC2P4). All but one of the duplicated loci have stop codons in the immediate 5′ region. In c17.6, the first stop codon is at nt 2964 within the first copy of a 62 bp sequence repeated five times. This repeat is also present in three other cDNA clones.

jdf2^122SJ and jdf2^1971SJ are ENU mutations that result in a significant reduction in Herc2 mRNA levels (M.J. Walkowicz et al., in preparation). RT–PCR of the Herc2 cDNA sequence from 12 278 to 14 082 bp (Fig. 5a) amplifies two fragments in the jdf2^122SJ mutant instead of one wild-type fragment. Sequencing of the RT–PCR products from this mutant identified 7 bp missing in the apparently ‘normal’ fragment, representing a frameshift and premature stop of protein translation, and 84 bp missing in the smaller one, which is predicted to encode a protein missing 28 amino acids (Figs 2b and 5b). Sequencing of genomic PCR products identified a single A→T transversion in the conserved splice acceptor site of the 84 bp exon (Fig. 5c). Pre-mRNA in the
Figure 5. ENU-induced Herc2 gene mutations in jdf2 mice. (a) RT–PCR of mutant alleles. The hemizygous jdf2^322SJ allele shows two bands compared with one band in other alleles and wild-type BJR. A smaller band is produced by the jdf2^322SJ allele in hemizygous or compound heterozygous mice, compared with other alleles (850SJ or wild-type control BJR). A band shift also occurs for jdf2^932SJ compared with other alleles. 46D denotes p.46DFiOD, which is a large radiation-induced deletion including both p and the Herc2 locus (11,48), while 850SJ (11) represents an ENU-induced allele, jdf2^850SJ, whose molecular basis is currently unknown. (b) Exon skipping of mutant alleles. The locations are indicated for Herc2 cDNA (nt) or for HERC2 protein in amino acid position. Amino acid sequences in a skipped exon are marked by a horizontal bracket, underlined amino acids are in-frame after the skipped exon and amino acids in bold, after the skipped exon, are out-of-frame and lead to a premature stop codon. For jdf2^322SJ, two abnormal mRNA products are present, one from an 84 bp exon skip which leads to an in-frame deletion of 28 amino acids, and the second due to the use of a cryptic splice site and a partial exon deletion of 7 bp, leading to a translation frameshift and premature stop. A 109 or a 159 bp exon is skipped in the jdf2^1971SJ or jdf2^932SJ mutants, leading to a frameshift and premature stop of protein translation or in-frame deletion of 53 amino acids, respectively. (c) Splice sequence mutations (arrowheads) in the Herc2 gene in ENU-induced mutants (GenBank accession nos AF071174–AF071177). An A→T point mutation occured in jdf2^322SJ in the consensus splice acceptor of the skipped exon. A cryptic splice site utilized in jdf2^322SJ is shown in bold, lowercase letters. A transition mutation (A→G) occured in jdf2^1971SJ and a transversion (T→G) in jdf2^932SJ, both in the consensus splice donor of the skipped exon.

jdf2^322SJ mutant therefore uses a cryptic splice site located 7 bp into the exon to generate the 7 bp deletion or skips the whole exon to form the 84 bp cDNA deletion (Fig. 5b). A single, smaller sized RT–PCR product was identified in the jdf2^1971SJ mutant compared with the normal control (Fig. 5a; Herc2 nucleotides 7547–8320). Sequencing of RT–PCR products revealed a 109 nt deletion in the mRNA produced by jdf2^1971SJ and sequence analysis of genomic PCR products showed that the deleted sequence is a single exon (Fig. 5b). In jdf2^1971SJ an A→G transition was identified in the consensus 5’ splice donor site (25) compared with the wild-type sequence (Fig. 5c). The absence of the 109 nt exon leads to a frameshift and premature stop of protein translation.
translation (Figs 2b and 5b) so that the Herc2 mRNA in the jdf2\textsuperscript{1971S} mutant encodes a putative protein of only 2733 amino acids and lacks several of the known polypeptide motifs (Fig. 2b). jdf2\textsuperscript{2972S} is an ENU mutation consistently producing 140% of wild-type Herc2 mRNA levels (M.J. Walkowicz et al., in preparation). A single smaller sized RT–PCR product was identified compared with the wild-type control (Fig. 5a; nt 9899–11 637 of Herc2 cDNA). Sequencing of RT–PCR and genomic PCR products identified a 159 bp in-frame cDNA deletion corresponding to omission of a single exon (Figs 2b and 5b). Genomic PCR and sequence analysis identified a point mutation (T→G) in the conserved splice donor site (Fig. 5c) of the skipped exon.

**DISCUSSION**

We have described the isolation of a functional ‘ancestral’ gene, Herc2, from low-copy repeats flanking human chromosome 15q11–q13. Recent evolutionary genomic duplications of Herc2 have led to a family of adjacent (15q13) and dispersed (15q11 and 16p11.2) copies, many or all of which are transcribed but truncated relative to the ancestral gene and that contain internal stop codons. It is currently unknown whether the duplications contain only Herc2 or also include other adjacent genes. Clearly, Herc2 is the ancestral gene, given its evolutionarily conserved functions, its orthologous genetic map position relative to the unique mouse Herc2 gene (10,16) and the presence of intragenic deletions and premature translational termination codons in the HERC2P1–P7 cDNAs. Comparison of Herc2 and duplicated chromosome 15 genomic sequences (HERC2P6 and HERC2P7) suggests that these two duplicated sequences arose 14–20 million years ago, if we assume estimates of mutation rates for orthologous silent site substitutions and intronic sequences (26). [If paralogous site replacement rates were assumed (26), the estimates would be more recent, but such estimates would be counteracted by sequence homogenization mechanisms, which would lead to an underestimation of the age of the duplicated sequences. Primate studies may resolve these issues.] Similar analyses indicate that the chromosome 16p11.2 HERC2P4 gene diverged from other HERC2 sequences 7–10 million years ago, but that the HERC2P1–P3 expressed sequences diverged from HERC2 3–6 million years ago. However, the latter three sequences are more like each other than HERC2 and thus either they diverged from a common HERC2-related sequence 1–2 million years ago or an alternative scenario is that this may indicate that some of the HERC2 sequences undergo sequence homogenization by unequal crossover or gene conversion. Furthermore, the sequences defining HERC2P6 and HERC2P7 are absent from the HERC2P4 genomic BAC sequence and the HERC2P1–P3 cDNAs. Combined, these data indicate that HERC2-related subfamilies of sequence appear to have arisen several times in the evolution of the human genome. The simplest model is that the ancestral HERC2 gene first duplicated to an adjacent location in a position equivalent to human 15q13. Subsequently, additional duplications and divergence occurred at this location, followed by duplication and transposition of a block of HERC2 sequences to a position several megabases away in 15q11. Finally, pericentromeric duplication (26) resulted in the two additional copies located in human chromosome 16p11.2. A complete characterization of HERC2 and duplicated sequences in human and other primates will provide a better understanding of the evolution of these and similar low-copy, subchromosomal repeats in the human genome and their important role in human disease (5,27–30).

The giant protein encoded by Herc2 is unusually highly conserved in mouse and human (compare with average in ref. 31). Herc2-related sequences are also identified in marsupials, chicken, fish and fruit fly by moderate stringency hybridization (Y. Ji and R.D. Nicholls, unpublished data). Moreover, by database searches, we have identified a Drosophila EST (AA567486) and a chicken STS (X855533) highly homologous to human and mouse Herc2. The high degree of homology of Herc2 proteins from evolutionarily diverse species accordingly implies that the function of Herc2 has been well conserved throughout the animal kingdom.

We have shown that the Herc2, Herc1 (p532) and Herc3 genes evolved from a common, evolutionarily distant ancestral gene. Each encoded protein shares C-terminal RCC1-like and E3 ubiquitin ligase domains, in addition to more extensive sequence homology and an N-terminal RCC1-like domain for Herc2 and Herc1. Since the completely sequenced Saccharomyces cerevisiae and Caenorhabditis elegans genomes do not contain an ortholog of these genes, but do contain genes encoding proteins with RCC1, HERC or ZZ, zinc finger domains, it is likely that the ancestral Herc gene evolved by ancient gene duplication and gene fusion events. Alternatively, the ancestral gene may have been lost in yeast and nematodes. Subsequent to the origin of an ancestral gene, the three gene family likely arose as a consequence of the two genomic duplications thought to have occurred early in vertebrate evolution, followed by functional divergence (32). Intriguingly, the central part of Herc2 encoding the second RCC1-like domain is replaced in Herc1 (p532) by a Gg-(WD40) repeat (22). The RCC1 and Gg repeats each form a seven-bladed propeller structure from seven internal repeats (33–35). This observation suggests that a central seven-bladed propeller structure is essential for Herc2 and Herc1 function and raises a fascinating evolutionary question of how two genes independently gain different, but similarly folding and presumably similarly functioning, structural motifs.

Clues to possible functions of Herc2 come from studies of the Herc1 (p532) protein, which is located in the Golgi apparatus and cytoplasm (22,36). Biochemical studies have shown specific in vitro association of the Herc1 C-terminal RLD with ARF1 and that the N-terminal RLD acts as a guanine nucleotide exchange factor (GEF) for ARF1 and Rab (22), two small Ras superfamily GTPases involved in intracellular vesicular transport and membrane trafficking (37,38). Indeed, cytosolic Herc1 (p532) interacts with clathrin, in an ATP-dependent ternary complex with Hsp70 (36). Clathrin functions as the major structural component of coated vesicles involved in receptor-mediated endocytosis and exocytosis (39), suggesting that Herc1 functions in vesicular transport processes (36). Since the RLDs of Herc2 are as similar to RCC1 as yeast and human RCC1 are to each other (40) and given the homology of Herc2 and Herc1, we and others (10) propose that Herc2 may function as a GEF for an unknown small GTPase, most likely of the Rab or ARF subclasses.

Proteins containing a C-terminal HECT domain have been shown biochemically to function as E3 ubiquitin protein ligases, usually in the proteasome proteolytic pathway (20,21), although other roles in endocytosis of cell surface receptors (41) and...
modification of protein kinase activity (42) have been found. Similarly, the highly conserved HECT domain of HERC proteins likely also confers E3 ubiquitin ligase function. The ZZ-type zinc finger motif present in HERC2 probably mediates specific protein–protein interactions (19). Combined, these observations suggest that the large HERC2 protein has multiple macromolecular interactions within the cell, including likely functions as a GEF and E3 ubiquitin protein ligase, and which may be involved in protein trafficking and degradation pathways within the cell.

Further clues to the functional role of the Herc2 gene are provided by analyses of jdf2 mutations. All three ENU-induced mutations are Herc2 splice site mutations, leading to skipping of the adjacent exon. The exon skipping in jdf2′H23083′ and the 7 bp deletion in jdf2′H2322S′ arising from use of a cryptic 5′ splice site each cause a frameshift and premature stop of protein translation. The latter events are usually associated with a reduced mRNA half-life (25), as found for Herc2 mRNA levels in these two mutants but not in the third mutant (jdf2′H22325′) with an in-frame exon skip (M.J. Walkowicz et al., in preparation). While our paper was in preparation, Lehman et al. (10) independently identified the same mouse gene and suggested that it is responsible for the jdf (i.e., jdf2) phenotype. However, two of the three alleles studied (10) are very large deletions and provide no direct evidence toward a role for a gene at the deletion end-point in the underlying phenotype. While the third allele was an intragenic deletion, a smaller gene or genes lying within introns cannot be excluded as contributing to the phenotype. Therefore, our finding of three independent point mutations now conclusively demonstrates that Herc2 gene defects do lead to the complex, pleiotropic phenotypic abnormalities of the jdf2 syndrome. While the basis of running and neuromuscular tremors/weakness is unknown, sterility or reduced fertility in females results from an immature ovary phenotype, which may indicate an endocrinological (43) and/or an intrinsic ovary defect. The spermatogenesis defect is autonomous to the germ cell (14), with spermatid head defects characterized by the secretion and attachment of multiple acrosomal vesicles to the nuclear membrane, in addition to frequent binucleated spermatids with the nuclei conjoined by a single acrosome (11). The acrosome is a specialized structure formed from the Golgi during spermatogenesis and is involved in secretion of proteins into the egg upon fertilization (44). Intriguingly, pituitary neurosecretory vesicle defects with degeneration of neurosecretory axons of the pars nervosa were seen in the p23H radiation-induced mutant (13). These observations including defects in secretory pathways in sperm and the nervous system are consistent with the proposed role for the HERC2 protein in vesicular trafficking as suggested by structural relationships between HERC2 and other proteins. Identification of the human disease equivalent of jdf2, of mutations in other members of this and similar biochemical pathways, and of the specific functions of HERC proteins will likely shed significant light on the role of protein trafficking and degradation pathways in neuromuscular function, gametogenesis and cell development.

**MATERIALS AND METHODS**

**Molecular cloning of human HERC2 and mouse Herc2 cDNAs**

Human HERC2 cDNA (15.3 kb) was isolated by a combination of several techniques. A 138 bp PCR probe was amplified from bacteriophage clone λ6A1 (J.M. Amos-Landgraf et al., in preparation) using primers RN304 (5′-CGCTTCTCTGGAAGAGGCTG-3′) and RN305 (5′-GCTCTTCTTATCGCCGTTG-3′) and PCR conditions of 94°C for 30 s, 60°C for 30 s, 72°C for 30 s for 35 cycles and used to isolate clones from a cDNA library. Human EST 3′ of the 6.9 kb partial HERC2 cDNA were identified by BLAST search (http://www.ncbi.nlm.nih.gov/blast/blast+) and used with the mouse Herc2 cDNA sequence. Three ESTs were sequenced completely (EST H05966, highly similar to positions 9.5–10.8 kb of the Herc2 cDNA sequence; EST AA158176, 13.5–14.6 kb; EST H23083, 14.1–15.3 kb). Primers from the 6.9 kb Herc2 partial cDNA (RN384, 5′-AAACAATGCACATAACTCCAGCT-3′) and EST H05966 (RN651, 5′-GACCTGTCTTATT- TTGGCTTTC-3′) or EST H05966 (RN650, 5′-CTGCTGATGCGGATGACACACTCGT-3′) and EST AA158176 (RN653, 5′-AGCAAGTGAGCAGTCGTTG-3′) were used for long-range PCR to amplify cDNA clones spanning the remainder of HERC2 from a human fetal brain Marathon-Ready cDNA library with the Advantage cDNA PCR kit (Clontech, Palo Alto, CA). The same cDNA was used for 5′-RACE.

Mouse Herc2 cDNA clones were isolated by subtraction cloning (M.J. Walkowicz et al., in preparation) and standard cDNA library screening, using a testis library (Clontech) and a size-selected (4–10 kb) cerebellum cDNA library (45), as well as 5′- and 3′-RACE. Herc2 gene-specific primers were used to amplify the 5′-end (RN545, 5′-AAGAGAAGCGAAGGGAAGGAGT-3′, and nested primer RN544, 5′-AACACCAGCAGAAGACAC- TCT-3′) or two fragments from the 3′-end (RN555, 5′-AGAACATTTGAGCCTGTCTTG-3′, and nested primer RN556, 5′-ATCTGTTTTGCTGCTTGGAAAG-3′; RN568, 5′-AATCTGTGACCATTTGATTTGTA-3′, and nested primer RN569, 5′-CCAGGCTACATTTGGCCAGATTA-3′). 3′-RACE using Marathon-Ready cDNA (Clontech) with primers RN568 and RN569 yielded two similarly sized 7 kb products containing poly(A) tails derived from alternative poly(A) site utilization. Human HERC2 and mouse Herc2 cDNA sequences, generated using an Applied Biosystems 377 PRISM automated sequencer (SequiNet, Fort Collins, CO), were analyzed using BLAST. Amino acid motifs were identified by BLAST and PROSITE (http://expasy.hcuge.ch/sprot/prosite.html) and aligned using MegAlign (DNASTar, Madison, WI).

**Sequence comparisons of HERC2 and D15F37 cDNA clones**

Multiple cDNA clones representing the 6–7 kb D15F37 transcripts seen on human northern blots were recently isolated (5). Two cDNA clones, c17.6 (6.306 kb, HERC2P3) and c17.66 (6.071 kb, HERC2P2), were completely sequenced (GenBank accession nos AF041081 and AF041080, respectively): c17.6 lacks 3505 nt of the HERC2 sequence from bp 134 to 3638, whereas c17.66 shares sequence with HERC2 until bp 248 of HERC2 and contains HERC2 sequence from bp 1145 to 1318 (Fig. 4). Partial sequence analysis of two other D15F37 cDNA clones, c17.28 and c17.30, suggests that they are alternatively spliced forms of c17.66 and c17.6, respectively (Fig. 4). From bp 3638 to 8321 of the HERC2 transcript, c17.6 and c17.66 are highly homologous to the HERC2 sequence with only minor gaps (Fig. 4). Downstream of this region, the two D15F37 cDNA sequences diverge completely from the HERC2 sequence. Two additional sequences were also analyzed. The sequence in GenBank accession no.
AB002391 is truncated at the 5′-end (bp 170) of HERC2, but otherwise is very similar to cl7.28 and cl7.30. A chromosome 16p11.2 genomic clone (GenBank accession no. AC002401) contains additional exons in three segments homologous to 5′ HERC2 cDNA sequences compared with the other loci identified. Partial sequence of an additional six cDNA clones and a GDB search identified >60 ESTs related to the 3′ D15F37 sequence (not present in the HERC2 cDNA sequence).

**Cell lines, Southern and northern hybridizations**

Lymphoblastoid cell lines from PWS and AS patients as well as normal controls (46; NIGMS, Coriell Institute for Medical Research, Camden, NJ) were cultured by standard methods. Human chromosome 15–rodent hybrid cell lines were as described (23). Genomic DNA was isolated by phenol–chloroform extraction (47) and Y AC DNA by standard protocols (NIH, Bethesda, MD). Multi-tissue northern blot filters (Clontech) and Southern blots were prepared and hybridized using standard methods (47).

**Mutation detection in ENU-induced jdf2 mutants**

Total brain RNA was reverse transcribed into cDNA with MMLV reverse transcriptase (Gibco BRL, Gaithersburg, MD) and PCR amplified in 1.0–2.0 kb segments. Primers used for RT–PCR to detect the Herc2 cDNA deletion in each mutant were: jdf2197S, RN556 (5′-ATCGTGTTGCTGGTTGGAAGG-3′) and RN570 (5′-AAAGATGCTGGTTGGAAGG-3′); jdf2223S, RN667 (5′-GGGGAGATGGGAAGGTTG-3′) and RN668 (5′-ATTGTCTCGGTTTCTGTGA-3′); jdf2922S, RN665 (5′-AGACTTGGCTGTTTCC-3′) and RN666 (5′-TCCAGGGTCAAAAGCAAGGC-3′). Genomic fragments from wild-type BJR and mutant DNA were amplified using the Advantage Genomic PCR kit (Clontech) with primers designed from cDNA sequences flanking the apparent cDNA deletions: jdf2197S, RN597 (5′-ACTCATCAAAATGGCCCGTTAG-3′) and RN600 (5′-CA-AAAATCAAGATCATCATACAGTTTC-3′); jdf2223S, RN691 (5′-GCTCCTGGTTGAAACTGACTTG-3′) and RN693 (5′-ACAGGCCCATTGGCGATCTCG-3′); jdf2922S, RN687 (5′-AATGGCTGACTGTTGGAAGA-3′) and RN690 (5′-TCTCCCACTATTCTCCCAGACG-3′). Amplified products were analyzed by 0.8–1.5% agarose gels and PCR products cloned into pCR2.1 (Invitrogen, Carlsbad, CA) for DNA sequence analysis.

**ACKNOWLEDGEMENTS**


**REFERENCES**


45. Doyle, J., Ren, X., Lennon, G. and Stubbs, L. (1997) Mutations in the CauMil1α4 calcium channel gene are associated with seizures, cerebellar degeneration and ataxia in tottering and leaner mutant mice. Mamm. Genome, 8, 113–120.

