

# The human HERC family of ubiquitin ligases: novel members, genomic organization, expression profiling, and evolutionary aspects<sup>☆</sup>

Karin Hochrainer, Herbert Mayer, Ulrike Baranyi<sup>1</sup>, Bernd R. Binder,  
Joachim Lipp, Renate Kroismayr<sup>\*</sup>

*Department of Vascular Biology and Thrombosis Research, Medical University of Vienna, and Biomolecular Therapeutics, Brunnerstrasse 59,  
1235 Vienna, Austria*

Received 30 July 2004; accepted 13 October 2004  
Available online 2 December 2004

## Abstract

The HERC family of ubiquitin ligases is characterized by the presence of a HECT domain and one or more RCC1-like domains. We report the identification of two novel members, *HERC4* and *HERC6*, and subdivide the family into one group of two large and one group of four small members according to protein size and domain structure. The small members share a similar genomic organization, three of them mapping to chromosomal region 4q22, indicating strong evolutionary cohesions. Phylogenetic analysis reveals that the *HERC* ancestor emerged in nematodes and that the family expanded throughout evolution. The mRNA expression pattern of the small human members was found to be diverse in selected tissues and cells; overexpressed proteins display a similar cytosolic distribution. These data indicate that the HERC family members exhibit similarities in many aspects, but also sufficient differences indicating functional diversity.

© 2004 Elsevier Inc. All rights reserved.

**Keywords:** Ubiquitin ligase; HERC; HECT; RCC1; Alternative splicing; Evolution

Ubiquitination of proteins is implicated in many cellular processes, including protein degradation, receptor internalization and trafficking, signal transduction, cell cycle, apoptosis, and DNA repair [1–6]. Different types of modification are distinguished; the best known are polyubiquitination serving as a signal for proteasomal degradation [7] and monoubiquitination regulating endocytosis and lysosomal degradation [4,8]. Ubiquitination requires the concerted action of an enzymatic cascade involving an E1

ubiquitin-activating enzyme, an E2 ubiquitin conjugase, and an E3 ubiquitin ligase. Whereas there is only one E1 enzyme and a restricted number of E2 enzymes, ubiquitin ligases are abundant to guarantee specific substrate recognition. Depending on their mode of action and composition of the catalytic site, three types of E3 enzymes are known: HECT (homologous to E6-AP (E6-associated protein) C-terminus), RING, and U-box proteins. Whereas RING and U-box proteins transfer ubiquitin directly from an E2 to a target [9,10], HECT proteins form a thioester with ubiquitin by an active cysteine residue before transferring it to a substrate [11].

The HECT domain consists of 350 amino acids (aa) and was first defined in E6-AP-mediated ubiquitination of p53 in cells infected by high risk papillomaviruses [12]. Proteins containing RCC1 (regulator of chromosome condensation 1)-like domains (RLD) in addition to a HECT domain are denominated as HERC according to HUGO [13]. RCC1 is a GEF (guanine nucleotide exchange factor)

<sup>☆</sup> Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. AY221963, AF336798, AY650032, AY650033, AY650034, AY653201, AY653202, AY653203, BK005506, BK005507, BK005531, BK005532, BK005524, BK005525, BK005526, BK005527, BK005528, BK005529, and BK005530.

<sup>\*</sup> Corresponding author. Fax: +43 1 4277 62550.

E-mail address: [renate.kroismayr@univie.ac.at](mailto:renate.kroismayr@univie.ac.at) (R. Kroismayr).

<sup>1</sup> Current address: Department of Surgery, Medical University of Vienna, Währinger Gürtel 18–20, 1090 Vienna, Austria.

for Ran and participates in nucleocytoplasmic transport and mitotic spindle formation [14]. Therefore a role as both GEF and ubiquitin ligase is assumed for HERC proteins.

Four human HERC family members are described so far. HERC1 localizes to inner cell membranes and is an active GEF for ARF and Rab family GTPases [15]. In the mouse *Herc2* is located in the *jdk2/rjs* chromosomal region involved in a genetic disorder resulting in neuromuscular secretory vesicle defects, sperm acrosome defects, and juvenile lethality [13,16]. In the human genome both genes map to chromosome 15 and the encoded proteins of approximately 5000 aa are assumed to function in vesicular trafficking [17]. *HERC3* and *HERC5* map to chromosome 4 and the encoded proteins have a molecular weight of approx 120 kDa each [18,19]. The subcellular distribution of HERC3 is similar to that of HERC1, suggesting a role for HERC3 in intracellular trafficking [20]. For HERC5, identified as cyclin E binding protein-1, a role in the cell cycle was suggested [19]. Furthermore we have shown that the cellular protein level of HERC5 is tightly regulated in endothelial cells [21]. The ubiquitin ligase activity has hitherto been demonstrated only for HERC3 and HERC5 [18,21,22].

Here we report the identification of two novel human *HERC* members, *HERC4* and *HERC6*. Apart from the major transcripts we identify several alternative splice products of both genes. By means of chromosomal localizations, genomic organizations, and phylogenetic comparisons of the family members we reveal evolu-

tionary cohesions and identify *HERC4* as the common ancestor.

## Results

### Identification and splicing of *HERC6*

During cloning of *HERC5* [21], we identified a cDNA fragment of 1137 bp encoding a partial HECT domain with approx 50% nucleotide sequence identity to the corresponding region of *HERC5* in an endothelial cell (EC) cDNA library screen. We applied 5' and 3' RACE methods and obtained a cDNA of 3.89 kb, containing an open reading frame (ORF) of 1022 aa, a 5'UTR of 183 bp, and a 3'UTR of 636 bp. In addition to the HECT domain at the C-terminus, a RLD is encoded in the N-terminal part of the protein spanning aa 22–304. The derived sequence was submitted to GenBank (AF336798) and approved as *HERC6* by HUGO.

Cloning of full-length *HERC6* cDNA from Jurkat cells revealed that 40% of all clones contained deletions and insertions with respect to the finally designated major transcript (Fig. 1A). All sequence gaps and inserts were found to represent exons flanked by typical exon–intron boundaries of the *HERC6* gene locus. Alternative splicing of exons 7, 10, 14, and 17 was observed and representative transcripts were submitted to GenBank (AY653203, AY-653202). We used RNA from different sources to compare

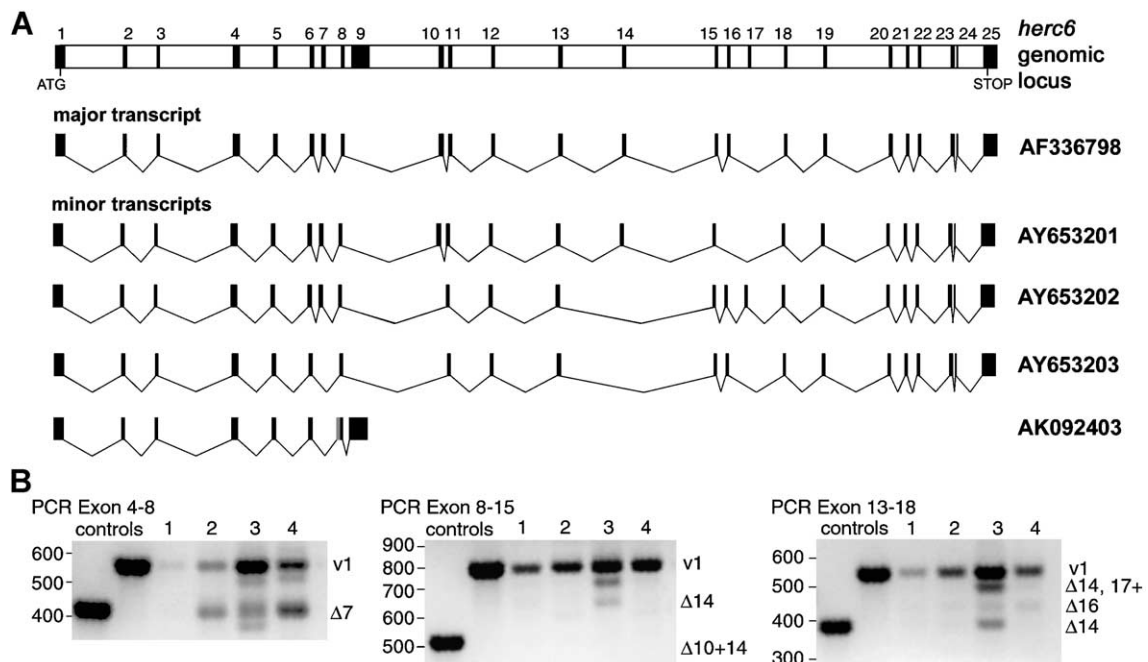


Fig. 1. Genomic organization and splicing of human *HERC6*. (A) Diagram of the *HERC6* locus. Exons, designated 1–25, are shown as filled bars; the 34 nt upstream of exon 8 in AK092403 are shown as a gray box. (B) Presence of *HERC6* splice products analyzed by RT-PCR. Plasmids encoding the transcript variants shown in (A) were used as positive controls: lanes 1, placenta; lanes 2, HeLa cells; lanes 3, Jurkat cells; lanes 4, human skin microvascular endothelial cells; v1, major transcript; DNA marker is indicated at the left.

the result from Jurkat to other cell types and tissues (Fig. 1B). Whereas transcripts lacking exon 7 are common in HeLa, Jurkat, and human skin microvascular endothelial cells (HSMEC), transcripts lacking exons 10 and 14 are less abundant. All these splicing incidents cause a frameshift in the ORF resulting in truncated proteins encoding RLD only (322 and 364 aa). Insertion of exon 17 was detectable in Jurkat only in transcripts missing exon 14. Additionally we have identified a transcript missing exon 16 in HSMEC, by PCR also detectable in Jurkat (AY653201). Deletion of exon 16 does not prematurely terminate translation and results in a protein of 986 aa. AK092403, isolated by others from placenta, lacks exon 7, utilizes an intraintronic splice site 34 nt upstream of exon 8, and terminates in exon 9. We were not able to detect this splice variant in any cell or tissue examined by RT-PCR using various primers amplifying different fragments between exons 4 and 9 (data not shown); hence, this splice product is probably very rare.

#### *Cloning and alternative splicing of HERC4*

By browsing human genome databases we identified a human cDNA derived from fetal brain, KIAA1593/AB046813 [23], due to its striking similarity to *HERC6*. Sequence comparison of the KIAA1593 cDNA with the mouse ortholog implicated an extended 5' end of the human gene. Alignment of the RLDs of all HERC family members supported this observation. Elongation of the sequence by PCR using primers designed according to public EST clones resulted in a cDNA of 4.45 kb (GenBank AY221963) with the HUGO-approved gene symbol *HERC4*, encoding a protein of 1057 aa, a 5' UTR of 249 bp, and a 3' UTR of 1023 bp.

Several EST sequences map to the *HERC4* genomic locus, indicating the existence of alternative splice products (Fig. 2A). Clones were obtained from the IMAGE Consortium (<http://image.llnl.gov>) and after resequencing submitted to GenBank as AY650034, AY650033, and AY650032. The sequence of AK026808, obtained from the NEDO human cDNA sequencing project (<http://www.nedo.go.jp/bio-e/>), was confirmed. AY221963 and BC039600, submitted recently [24], represent the longest cDNAs, consisting of 26 and 25 exons, respectively, differing only by exon 20 encoding 8 aa. An alternative start codon is used in AB046813, encoding a protein of 947 aa, which lacks part of the RLD. Exons 24 and 25 are missing in AY650033, altering the HECT domain. BX400886, submitted by others, utilizes a cryptic splice site within exon 6 maintaining the ORF. Absence of exons 5, 9, and 20 in AY650032 causes a frameshift resulting in a truncated protein of 118 aa, which lacks most of the RLD and the HECT domain. AK026808 as well as AY650034 terminates in exon 4; the corresponding 110-aa proteins include neither a RLD nor a HECT domain. Existence of exon 12 is demonstrated by BX537563, an ORF that stops

in the accessory part of exon 13 (13\*) generated by intraintronic splicing.

To verify the transcript variants we performed a series of PCR amplifications on cDNAs derived from human fetal and adult brain, as well as brain EC (HBEC), smooth muscle cells (HSMC), and HeLa cells using specific primers for exons 4, 5, 6, 9, 12, 13, 20, 24, and 25 (Fig. 2B). The selection of tissue and cell types was based on the expression profile for *HERC4* provided by Kazusa DNA Research Institute and our own experiments (see below). Presence of exon 4, 6, 12, or 13\* is hardly detectable in all samples tested, whereas a splice form lacking exon 9 is observed only in HSMC. Absence of exons 24 and 25, represented by AY650032, is infrequent, as their presence is easily detectable by RT-PCR. All these observations are supported by the fact that for each case only a few EST clones are annotated to public databases. According to our analysis AY221963 and BC039600 represent the two major transcripts, with exon 20 present in brain tissue and HeLa cells, absent in HBEC and HSMC. Moreover numerous EST clones and annotated cDNAs support both variants.

Based on our data, a transcriptional regulation by alternative splicing is conceivable for both novel genes; however, the functional role remains uncertain. In the following we refer to AY221963 and AF336798 as *HERC4* and *HERC6*, respectively.

#### *Subdivision of human HERC proteins into large and small members*

With the identification of HERCs 4 and 6 the human HERC family comprises six members (Fig. 3). Whereas *HERC1* and *HERC2* are large proteins of about 500 kDa, HERCs 3 to 6 have a molecular weight of approx 120 kDa. According to PFAM analysis [25] HERCs 3 to 6 are characterized by the presence of only one RLD at the N-terminus and a HECT domain at the C-terminus. In contrast, *HERC1* and *HERC2* contain up to three RLDs and several other protein domains. According to these differences we divide the family into a large and a small subfamily as also suggested by others [26].

Alignment of protein sequences of the small human HERCs underlines their high similarity (Fig. 4). Amino acid identities range from 52% between HERCs 3 and 4 to 34% between HERCs 4 and 5 (Supplementary Table S1). The HECT domains comprise the most conserved part of the proteins, with 61% identity between HERCs 3 and 4, 60% between HERCs 5 and 6. Within the RLD, HERCs 3 and 4 are most homologous with 55% aa identity. All comparisons indicate HERCs 3 and 4 as well as HERCs 5 and 6 to be most closely related to each other.

#### *Chromosomal localization and genomic organization*

*HERC3* and *HERC5* both have been reported to be localized on the long arm of chromosome 4 by FISH

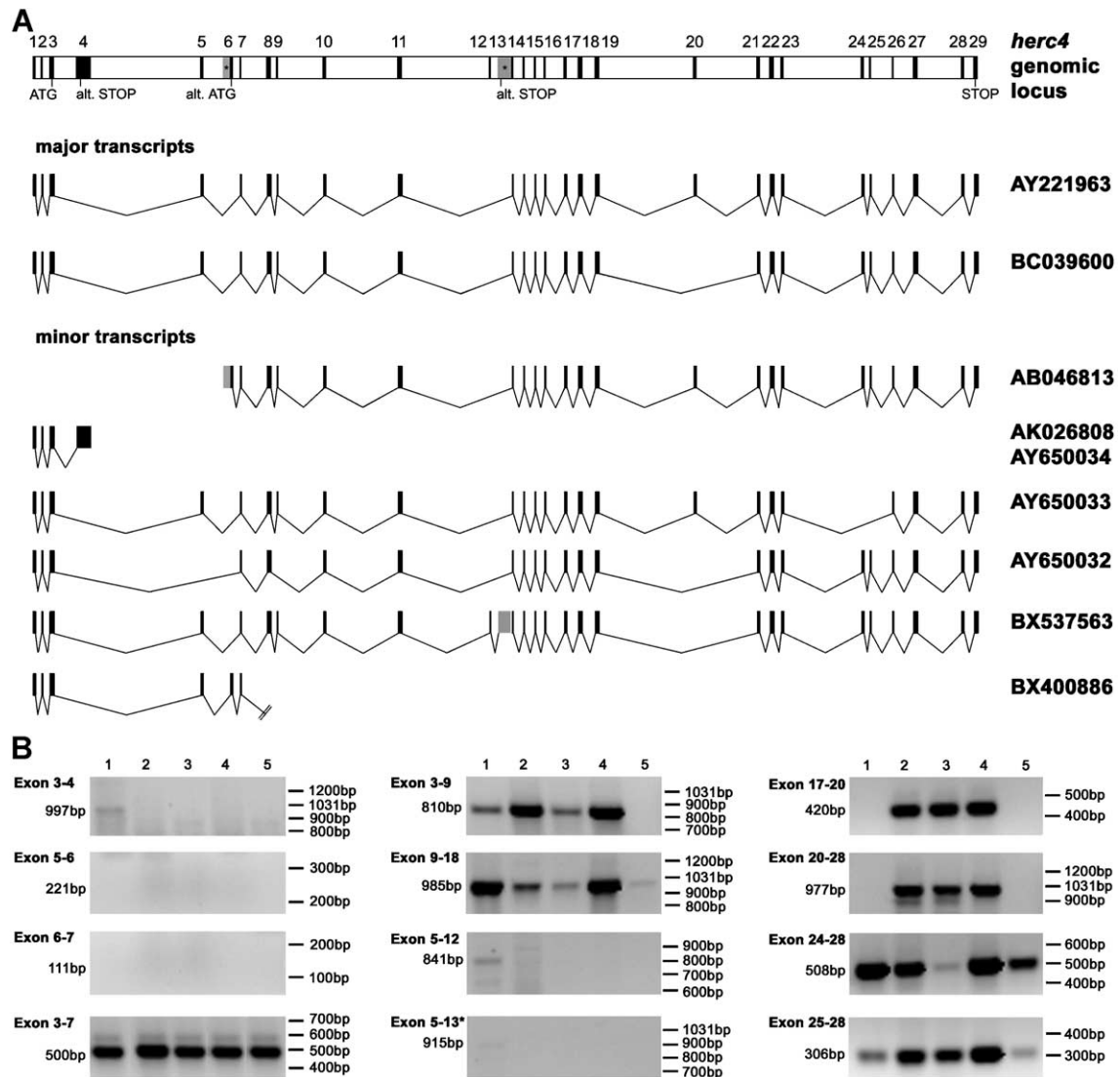


Fig. 2. Alternative splicing of human *HERC4*. (A) Diagram of the *HERC4* genomic locus and alternatively spliced mRNAs. Exons are shown as filled bars, critical exons are indicated by asterisks. GenBank accession numbers are listed at the right. (B) RT-PCR analysis of *HERC4* transcripts, primers are binding in the exons indicated at the left. A DNA size marker is indicated at the right. Lanes 1, human brain endothelial cells; lanes 2, fetal brain; lanes 3, adult brain; lanes 4, HeLa cells; lanes 5, human uterus smooth muscle cells.

experiments [18,19], *HERC1* and *HERC2* map to chromosome 15q [13,17]. By hybridization of a somatic cell line blot we mapped *HERC5* and *HERC6* to chromosome 4 (data not shown). Alignments of respective cDNAs to the human genome using the UCSC genome browser, July 2003 assembly, confirm this localization. *HERC4* localizes to chromosome 10q22, spanning a genomic region of 118 kb (Fig. 5A). *HERC3*, *HERC5*, and *HERC6* are clustered on chromosome 4q22 spanning a region of about 330 kb at ~90 Mb with a spacer of 14 kb between *HERC6* and *HERC5* and 86 kb between *HERC5* and *HERC3*.

Via genomic alignments we calculated both intron and exon sizes of the small *HERC* genes (Table 1, Supplementary Tables S2 S3 S4 S5). Both *HERC3* and *HERC4* are composed of 26 exons, with the ORF starting in exon 3. *HERC5* and *HERC6* contain 23 exons, the start ATG is

encoded in exon 1. Exons vary in length from 24 to 382 bp and show striking similarity or even identity in terms of size and arrangement, intron sizes are not conserved. Chromosomal localization as well as organization indicates a common evolutionary history of all *HERC* members.

#### Evolutionary relationship of the *HERC* gene family

By using BLASTN and BLAT we searched for *HERC* orthologs in mouse, rat, pufferfish, zebrafish, *Drosophila*, and *Caenorhabditis elegans*. Whereas for many orthologs full-length cDNAs are annotated to public databases, only partial cDNAs or predictions were available for mouse, rat, and *Fugu HERC1*, rat and *Fugu HERC2*, *Fugu HERC3*, *Fugu HERC4*, and mouse and rat *HERC6*. Therefore, we generated the appropriate full-length sequences from

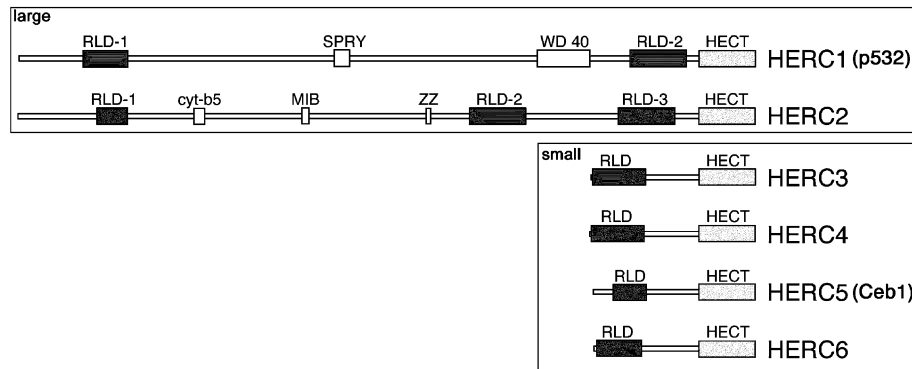


Fig. 3. Structural characterization of human HERC members. Names are according to HUGO, prior designations are put in parentheses. RLD and HECT domain are shown as gray boxes, others are displayed as white boxes: SPRY, Spia and ryanodine receptor domain; WD40, WD40 domain and G- $\beta$  repeat; cyt-b5, cytochrome *b5*-like heme/steroid binding domain; MIB, mind bomb herc2 domain; ZZ, zinc finger, ZZ type.

available genomic data upon homology to known *HERC* genes. Where applicable, confirmation by EST clones was performed. Derived sequence data were annotated to the Third Party Annotation (TPA) database, accession numbers used for further studies are listed in Table 2.

The evolutionary relationships among the *HERC* orthologs are shown in Fig. 5B. In agreement with domain architecture, the phylogenetic tree contains two distinct clusters of closely related sequences representing the *HERC* subfamilies. The *C. elegans* ortholog of *HERC4* emerges directly out of the basal line, linking the two clusters. Concomitantly, *Drosophila* *HERC4* interconnects the sub-clusters of the small family members. Phylogenetic comparison predicts *HERC4* as the most ancient family member, with the *C. elegans* ortholog most closely related to a common ancestor.

Based on genomic analysis of small *HERC* orthologs, a speculative evolutionary scheme can be drawn (Fig. 5C). *herc4* is the only *HERC* existing in nematodes, in insects two *HERCs* are present *herc2* and *herc4*. In teleosts *herc3* and *herc1*, in rodents *Herc6*, and in primates *HERC5* are emerging. All *HERC* genes map to regions of conserved synteny between human, mouse, and rat. In the human genome, *HERC3*, *HERC5*, and *HERC6* map within one synteny region of chromosome 4q22, in contrast to *HERC1* and *HERC2* localizing to different synteny regions on chromosome 15.

*HERC3* and *HERC4* are the most conserved small *HERC* genes; their exon structures are nearly identical across species (data not shown). Amino acid sequence comparison reveals that *HERC3* and *HERC4* are best conserved throughout evolution (Supplementary Table S1). Human *HERCs* 3 and 4 show 92 and 93% identity to their mouse and rat orthologs, whereas human and mouse *HERC6* are only 66% identical.

All these data indicate that the *HERC* family evolved from a common ancestral gene, pointing to *HERC4* as the most ancient member. We hypothesize that *HERC3* evolved out of *HERC4* by duplication and chromosomal rearrangement, whereas for the emergence of *HERC5* and *HERC6* gene duplication was sufficient.

#### RNA expression pattern of human small *HERC* genes

We examined the expression profiles by quantitative RT-PCR to evaluate *HERC3* to *HERC6* levels in selected tissues and cells. Expression levels were determined for each target gene and for tissues and cells by using the mean value of all samples as the reference in the equation for relative expression.  $\beta$ 2-Microglobulin ( $\beta$ 2m) was used for normalization in all experiments. As shown in Fig. 6A *HERC3* is highly expressed in fetal brain and reasonably detectable in adult brain, whereas in placenta, heart, and testis the mRNA level is fairly low. *HERC4* expression is most prominent in fetal brain; expression levels in adult brain and testis are comparable and at an intermediate level. *HERC5* is highly expressed in testis (100 $\times$ ), but also in fetal brain (30 $\times$ ) compared to placenta, heart, and adult brain (0.3 $\times$ ). Expression of *HERC6* is very low and varies only about 5-fold in all tissues examined. The results obtained by quantitative PCR are in agreement with preliminary data for *HERC3*, *HERC4*, and *HERC5* provided by the Kazusa DNA Research Institute and Mitsui et al. [19]. Expression of *HERC* genes in selected human primary cells and in HeLa cells is depicted in Fig. 6B. *HERC3* expression varies about 6-fold in all cells analyzed, with highest expression in HSMC and HSMEC. Expression levels of *HERC4* are constantly moderate; only HSMC express significantly higher amounts (3–4 $\times$ ) than other primary cells tested. *HERC5* mRNA levels are very low in all primary cells, but about 40-fold higher in HeLa cells. *HERC6* expression is comparably low in all cell types examined.

#### Subcellular localization of *HERCs* 3 to 6

According to PSORTII Reinhardt's method for cytoplasmic/nuclear discrimination, the four small *HERC* proteins are likely to be cytosolic. We experimentally determined the subcellular localization of *HERCs* 3–6 by confocal immunofluorescence microscopy of transiently transfected HeLa cells (Fig. 6C). All *HERC* proteins exhibit a punctate staining in the cytoplasm, confirming the bioinformatical



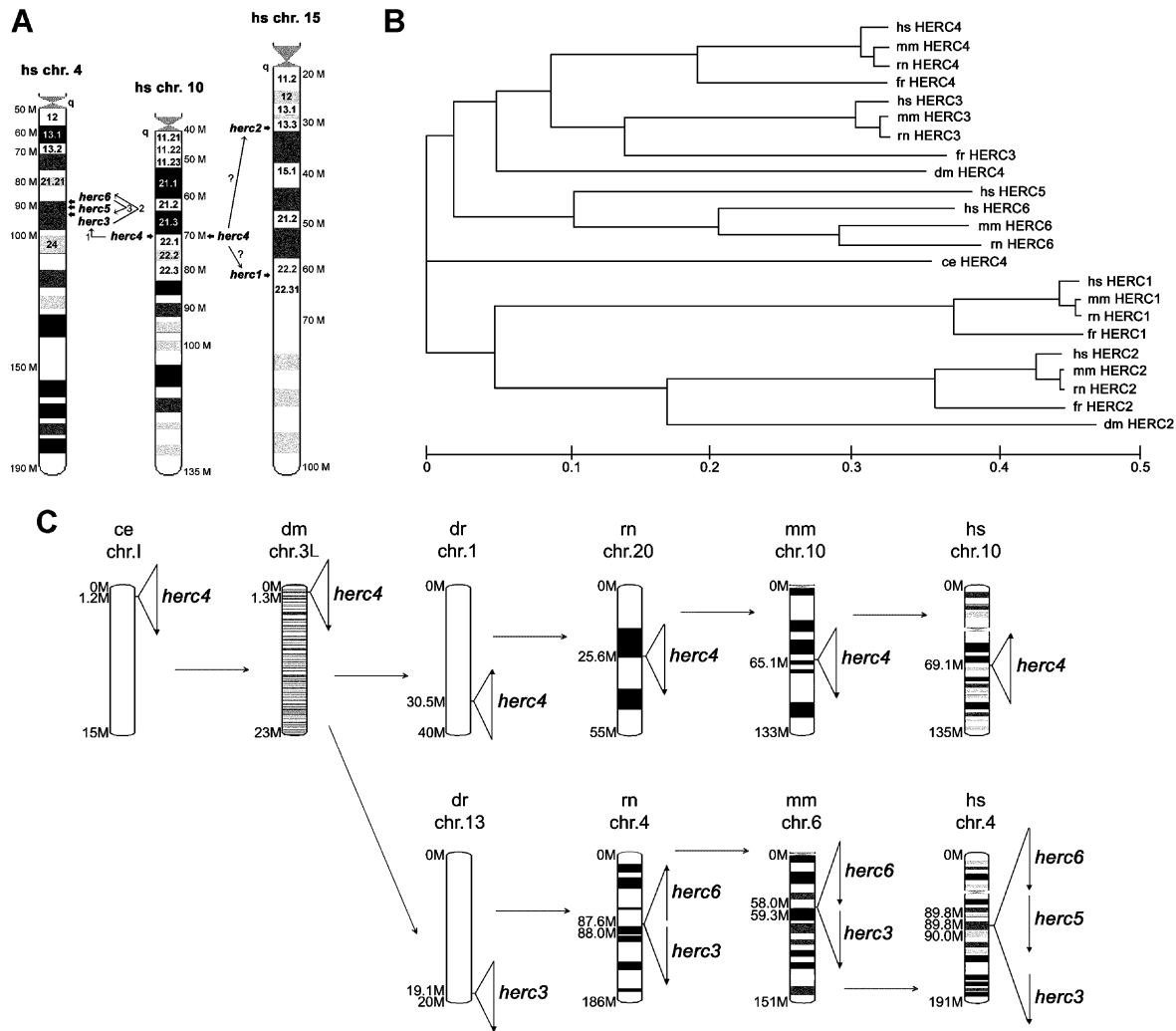


Fig. 5. Evolution of the *HERC* family. (A) Chromosomal localization of the human *HERC* genes. Duplication events are indicated by arrows on the chromosome idiograms. (B) The phylogenetic tree was generated based on a ClustalW alignment of the appropriate sequences. Branch lengths are proportional to the amount of inferred evolutionary changes. (C) Tentative duplication scheme of *HERC3-HERC6*. Positions of loci are displayed in megabases (M), arrowheads indicate orientation of transcription. hs, *Homo sapiens*; mm, *Mus musculus*; rn, *Rattus norvegicus*; fr, *Fugu rubripes*; dr, *Danio rerio*; dm, *Drosophila melanogaster*; ce, *Caenorhabditis elegans*.

prediction. Similar results were obtained when other epitope-tagged *HERC* proteins were expressed (data not shown).

## Discussion

With the identification of the two novel members, the *HERC* family finally consists of two large and four small proteins. Both analysis of the human genome and EST surveys suggest that no further family members are encoded in the human genome. Although there were reports about two other proteins containing a RLD in addition to a HECT domain [27], our detailed analysis revealed no RLD in these proteins. The two novel members are designated *HERC4* and *HERC6*, hence, the naming of the *HERC* genes is not in accordance with their discovery. It should be noticed that Ceb1 [19] was approved as

*HERC5* by HUGO although it was already referred to as *HERC4* [18,26].

Upon currently available sequence data from genome projects of several model organisms we deduce that *HERC* proteins are present neither in *Saccharomyces cerevisiae* nor *Arabidopsis thaliana*. As both RLD and HECT are encoded in these organisms, it is likely that the ancestral *HERC* evolved by gene fusion events. Our analyses identify the nematode *herc4* as the common ancestral gene of the whole family. The existence of a common ancestor was suggested earlier by Yi et al. [13], when they analyzed *HERC2*, comparing it to *HERC1* and *HERC3*. Subsequent gene duplications and chromosomal rearrangements increased the number of family members throughout evolution, resulting in six members in human. The younger members *HERC5* and *HERC6* are more homologous to each other than to *HERC3* or *HERC4*; nevertheless, *HERC3* and *HERC6* share a region of 141 nt with 98% identity, pointing to

Table 1  
Intron–exon organization of the human small HERC genes

hs <i>HERC3</i>				hs <i>HERC4</i>				hs <i>HERC5</i>				hs <i>HERC6</i>			
Exon	Intron			Exon	Intron			Exon	Intron			Exon	Intron		
1	79	1	2982	1	140	1	1457								
2	58	2	10,180	2	30	2	470								
3	255	3	43,790	3	304	3	28,319	1	378	1	382	1	382	1	4100
<b>4</b>	<b>160</b>	4	2012	<b>5</b>	<b>160</b>	5	6234	<b>2</b>	<b>124</b>	2	160	<b>2</b>	<b>160</b>	2	2118
<b>5</b>	<b>77</b>	5	780	<b>7</b>	<b>77</b>	7	3906	<b>3</b>	<b>77</b>	3	77	<b>3</b>	<b>77</b>	3	5076
<b>6</b>	<b>222</b>	6	951	<b>8</b>	<b>222</b>	8	1095	<b>4</b>	<b>222</b>	4	228	<b>4</b>	<b>228</b>	4	2608
<b>7</b>	<b>92</b>	7	1040	<b>9</b>	<b>92</b>	9	7101	<b>5</b>	<b>92</b>	5	95	<b>5</b>	<b>95</b>	5	2432
8	131	8	570	10	131	10	11,359	6	31	6	128	6	128	6	708
9	161	9	2379	11	161	11	21,343	7	146	7	137	7	137	7	1154
<b>10</b>	<b>77</b>	10	3939	<b>13</b>	<b>77</b>	13	282	<b>8</b>	<b>77</b>	8	68	<b>8</b>	<b>68</b>	8	6666
11	125	11	1597	14	125	14	999	9	104	10	122	10	122	10	556
<b>12</b>	<b>60</b>	12	3164	<b>15</b>	<b>60</b>	15	124	<b>10</b>	<b>60</b>	11	60	<b>11</b>	<b>60</b>	11	2910
13	112	13	403	16	112	16	503	11	94	12	94	12	94	12	4459
<b>14</b>	<b>190</b>	14	1778	<b>17</b>	<b>190</b>	17	1375	<b>12</b>	<b>190</b>	13	190	<b>13</b>	<b>190</b>	13	4558
15	164	15	115	18	173	18	21,860	13	155	14	155	14	155	14	6280
<b>16</b>	<b>114</b>	16	5965	<b>19</b>	<b>120</b>	19	7546	<b>14</b>	<b>114</b>	15	114	<b>15</b>	<b>114</b>	15	621
17	24	17	92	20	24	20	2136	15	111	16	108	16	108	16	3877
18	90	18	1540	21	99	21	1747								
<b>19</b>	<b>171</b>	19	1958	<b>22</b>	<b>168</b>	22	224	<b>16</b>	<b>171</b>	18	171	<b>18</b>	<b>171</b>	18	2412
<b>20</b>	<b>144</b>	20	922	<b>23</b>	<b>144</b>	23	13,489	<b>17</b>	<b>144</b>	19	144	<b>19</b>	<b>144</b>	19	4419
<b>21</b>	<b>167</b>	21	5410	<b>24</b>	<b>167</b>	24	1284	<b>18</b>	<b>167</b>	20	167	<b>20</b>	<b>167</b>	20	1013
<b>22</b>	<b>67</b>	22	414	<b>25</b>	<b>67</b>	25	3352	<b>19</b>	<b>67</b>	21	67	<b>21</b>	<b>67</b>	21	730
23	83	23	16,798	26	83	26	3372	20	71	22	71	22	71	22	2093
<b>24</b>	<b>184</b>	24	226	<b>27</b>	<b>184</b>	27	7409	<b>21</b>	<b>184</b>	23	184	<b>23</b>	<b>184</b>	23	90
<b>25</b>	<b>103</b>	25	2141	<b>28</b>	<b>103</b>	28	1945	<b>22</b>	<b>103</b>	24	103	<b>24</b>	<b>103</b>	24	1983
26	1738			29	1232			23	491	25	866	25	866		

Data are based on the analysis of human genome sequences in the UCSC and GenBank databases. Exons with identical size in all four members are indicated in bold, exons with sizes identical in three members are shown in bold italic. hs, *Homo sapiens*.

*HERC3* as the origin of *HERC6*. Gene duplication is still ongoing within the *HERC* gene family, as partial duplicated paralogs of the human *HERC2* are present on chromosomes 15 and 16 [28]. In rat, *Herc6* localizes to chromosome 4, but a 492-nt fragment of the 3' UTR also matches with 98% identity to chromosome 1q11.

The chromosomal localization of the *HERC* gene family reflects their common evolutionary history. In human, *HERC3*, *HERC5*, and *HERC6* localize to chromosome 4q22 and the most ancient member *HERC4* to chromosome 10. In all mammalian model organisms examined *HERC4* and the *HERC3–6* cluster are located in the same synteny regions. Noticeable, *HERC5* is the youngest member existing only in primates. A chimp ortholog is detectable by analysis of sequence data available at the chimpanzee genome browser at Ensembl. Similar to the human system, *HERC5* is present in the same synteny region as *HERC3* and *HERC6*.

Several splice products were observed for both *HERC4* and *HERC6*, most of them remarkably affecting translation. Truncated proteins lacking the HECT domain might serve as negative regulators of active ubiquitin ligases controlling substrate turnover, whereas alteration of the RLD would influence their putative GEF activity. Interestingly, occurrence of exon 16 in *HERC6* as well as the corresponding exon 20 in *HERC4* does not alter the ORF. Absence or presence of the encoded amino acids might cause a conformational change of the ubiquitin ligase and

affect substrate recognition. We observed alternative splicing of exon 20 in human *HERC4* to be cell-type dependent, possibly reflecting different functional requirements. The corresponding splice forms also exist in the mouse ortholog (BC060033 versus BC043082), for other species no information is available. Additionally, alternative splicing of exons 24 and 25 in *HERC4* may result in altered binding to the ubiquitin conjugase, a further possibility for interfering with ubiquitin ligase activity. Alternative splicing was also recently reported for the HECT ubiquitin ligase UBE3B [27], but no functional consequences are known so far. It will be interesting to elucidate further the possible regulation of ubiquitin ligases by alternative splicing and find out its functional consequences.

Two well-defined protein domains characterize the *HERC* family; however, associated functions are not confirmed for all members. So far, the ability to transfer ubiquitin via the HECT domain is demonstrated for *HERC3* and *HERC5* [20,21]; a GEF activity is known only for the RLD1 of *HERC1*, stimulating guanine nucleotide dissociation on several ARF and Rab family GTPases [15]. However, for the RLD2 of *HERC1* and the RLD of *HERC3* no GEF activity was detectable [15,20], therefore it remains open whether all RLDs encoded by *HERC* function as GEF.

Localization of *HERC1* to inner membranes together with its binding to clathrin suggests a role in vesicular

Table 2  
Human HERC genes and their orthologs

Gene name	GenBank Accession No.	TPA Accession No.	ENSEMBL Gene ID	Locus ID
hs <i>HERC1</i>	U50078	—	00000103657	8925
mm <i>Herc1</i>	AK083823 Partial	<b>BK005507</b>	00000042835 Partial	235439
rn <i>Herc1</i>	XM_236362 Partial	<b>BK005532</b>	00000017592 Partial	LOC 315771
fr <i>herc1</i>	—	<b>BK005525</b>	00000134686 Partial	—
dr <i>herc1</i>	—	—	00000002331 Partial	—
hs <i>HERC2</i>	AF071172	—	00000128731	8924
mm <i>Herc2</i>	AF061529	—	00000030451	15204
rn <i>Herc2</i>	XM_218720 Partial	<b>BK005526</b>	00000013718 Partial	LOC 308669
fr <i>herc2</i>	—	<b>BK005529 BK005530</b>	00000152536 Partial	—
dr <i>herc2</i>	—	—	00000020787 Partial	—
dm <i>herc2</i>	AY113293 Partial	—	CG11734	33035
hs <i>HERC3</i>	D25215	—	00000138641	8916
mm <i>Herc3</i>	BC042574	—	00000029804	73998
rn <i>Herc3</i>	XM_342701	—	00000007304	LOC 362377
fr <i>herc3</i>	—	<b>BK005527 BK005528</b>	00000135502 Partial	—
dr <i>herc3</i>	—	—	00000011556 Partial	—
hs <i>HERC4</i>	<b>AY221963</b>	—	00000148634	26091
hs <i>Herc4</i>	BC039600 BC060033	—	00000050498 Partial	67345
rn <i>Herc4</i>	XM_228147	—	00000000381	LOC 309758
fr <i>herc4</i>	—	<b>BK005524</b>	00000128236 Partial	—
dr <i>herc4</i>	—	—	00000013498 Partial	—
dm <i>herc4</i>	AY128479	—	CG9153	38151
ce <i>herc4</i>	—	—	Y48G8AL	171700
hs <i>HERC5</i>	<b>AY337518</b>	—	00000138646	51191
hs <i>HERC5</i>	AB027289 <b>AF336798</b> BC042047	—	00000138642	55008
mm <i>Herc6</i>	AK011047 Partial	<b>BK005506</b>	00000029798 Partial	67138
rn <i>Herc6</i>	XM_342700 Partial	<b>BK005531</b>	00000024024 Partial	LOC 362376

hs, *Homo sapiens*; mm, *Mus musculus*; rn, *Rattus norvegicus*; fr, *Fugu rubripes*; dr, *Danio rerio*; dm, *Drosophila melanogaster*; ce, *Caenorhabditis elegans*. Bold accession numbers are based on our submission.

trafficking [29]. The granular distribution in the cytosol of all small HERC proteins might indicate a similar function. The greatest difference among all HERC family members is the domain architecture enabling additional functionalities of the large subfamily members. The small members consisting of only one RLD and a HECT domain are highly homologous to each other. Despite their similarities we observed greatly varying mRNA expression levels in all tissues and cells examined. These differences might reflect the specific substrate recognition of the putative ubiquitin ligases. In general, expression of *HERC3* and *HERC4* is more easily detectable than the expression of *HERC5* and *HERC6*, with the exception of *HERC5* in testis and HeLa cells. This is in agreement with the suggested role in the cell cycle based on its ability to bind to cyclins and a correlation with the tumor suppressors p53 and pRB in several cell lines [19]. Moreover we have detected a notable expression of *HERC3*, *HERC4*, and *HERC5* in fetal brain, whereas expression is decreased in adult brain. This might indicate a putative role in brain development. Although nothing is known about human *HERC2* expression it might also play a role in neuronal tissue because the mouse *Herc2* is involved in a genetic disorder including neuromuscular defects [13].

Our study provides an extensive comparison of HERC family members revealing both similarities and differences. Identification of substrates of these ubiquitin ligases as well

as deciphering the function of the encoded RLDs will be necessary to elucidate the involvement of the HERC proteins in cellular processes.

## Materials and methods

### Bioinformatical analyses

Databases used for sequence analyses were the NCBI Server (<http://www.ncbi.nlm.nih.gov/>), the Ensembl database (<http://www.ensembl.org/>), and the UCSC Genome Browser (<http://genome.ucsc.edu/>). Sequence comparisons were performed using BLAST at NCBI and BLAT at UCSC. Pairwise protein alignment was carried out at NCBI using the BLAST2 SEQUENCES interface and BLOSUM62 scoring matrix. Multiple protein alignment was performed using ClustalW (<http://www.ebi.ac.uk/clustalw/>) exerting a Gonnet250 scoring matrix for calculating evolutionary distances followed by Boxshade 3.21 server ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)). Amino acid sequences were analyzed with PSORTII (<http://psort.nibb.ac.jp/form2.html>), protein motif searches were executed using Pfam [25]. Sizes of both exons and introns and their boundaries were determined by alignment of cDNA to genomic data. Chromosome models and chromosomal distances were derived from Ensembl.

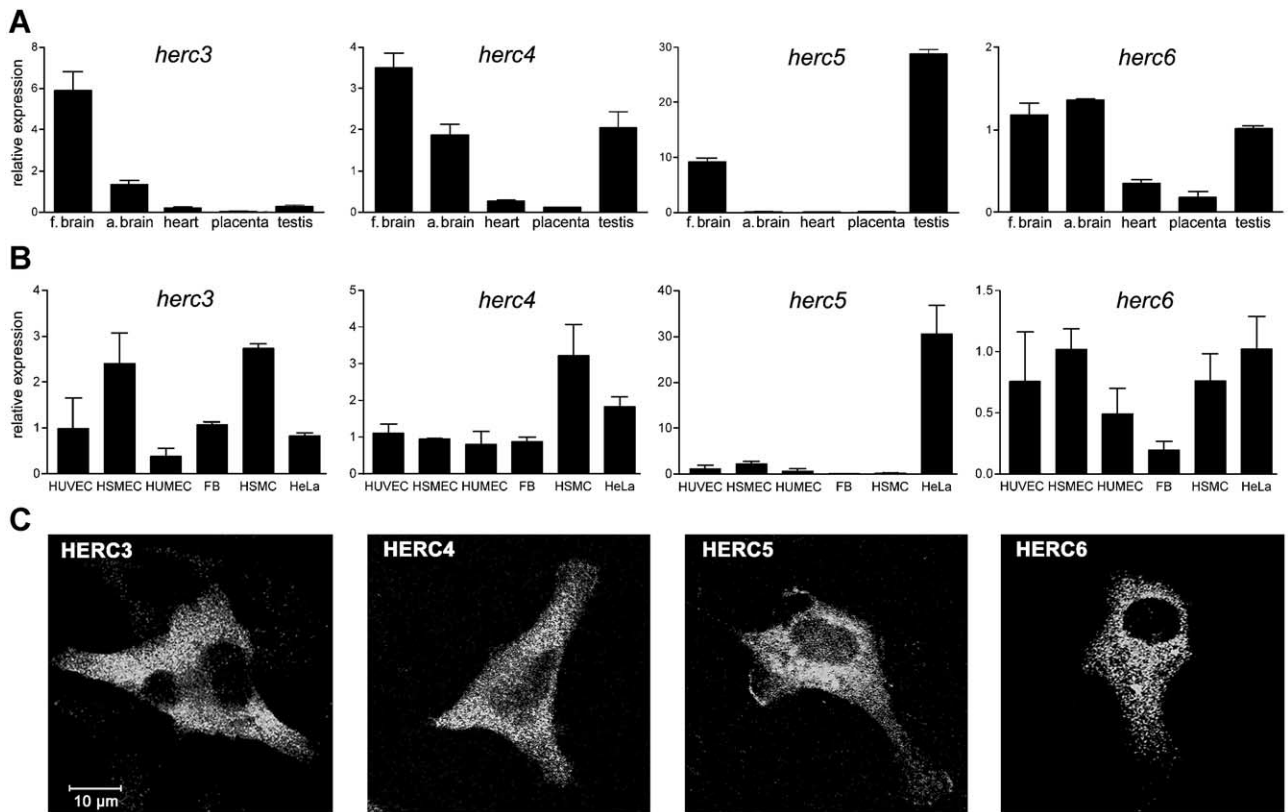


Fig. 6. mRNA expression profiling and protein expression of human HERCs 3–6. (A) Expression of *HERC3–HERC6* in selected tissues analyzed by quantitative real-time RT-PCR. Results are shown as mean values  $\pm$  SEM ( $n = 2–4$ ). f. brain, fetal brain; a. brain, adult brain. (B) Analysis of expression as in (A) in selected primary cells and HeLa cells. HUVEC, human umbilical vein endothelial cells; HSMEC, human skin microvascular endothelial cells; HUMEc, human uterus microvascular endothelial cells; FB, human skin fibroblasts; HSMC, human uterus smooth muscle cells. (C) Confocal images of HeLa cells expressing Myc-tagged *HERC3*, *HERC4*, or *HERC6* or FLAG-tagged *HERC5*.

### Cloning of *HERC* genes

KIAA0032/D25215/*HERC3* cDNA [30] and KIAA1593/AB046813/*HERC4* [23] were provided by the Kazusa DNA Research Institute (<http://www.kazusa.or.jp/huge/index.html>). The *HERC3* cDNA was cloned into pcDNA3.1-HisA (Invitrogen) through an *EcoRI–BglII* PCR fragment corresponding to nt 167–1263 of *HERC3* cDNA fused to *BglII–SmaI*-digested nt 1264–3370. PCR was carried out with High Fidelity DNA Polymerase (Roche). The *EcoRI–NotI* fragment of pcDNA3.1-HisA-*herc3* was subcloned into pCMV-Myc (Clontech). *HERC4* major transcript was generated by amplifying a cDNA fragment out of HeLa mRNA using primers with restriction sites *EcoRI–HindIII* corresponding to nt 250–1367 of *HERC4* cDNA. The fragment was ligated to *HindIII–EcoRI*-digested nt 1368–3490 and subsequently cloned into pCMV-Myc. pFLAG-*herc5* was generated by subcloning the *XhoI* fragment of pCS2 + MT [19] into pcDNA3.1-FLAG-C. To obtain full-length *HERC6* a cDNA fragment isolated from a HSMEC derived cDNA library was elongated by 3' and 5' methods as described previously [21]. A fragment including bp 184–1613, amplified by PCR to add a *BglII* site at its 5' end,

was generated and fused to a fragment consisting of bp 1613–3890 followed by *NotI*. The two fragments were cloned into *BamHI–NotI*-digested pCMV-Myc to obtain pCMV-Myc-*herc6*. All cDNA clones were sequenced using the ABI Prism Dye-Terminator Kit (Perkin-Elmer).

### Reverse transcription

Primary cells were grown as described previously [21]. HeLa and Jurkat cells were cultured in RPMI medium, 10% fetal bovine serum, and antibiotics. Total RNA was isolated using either High Pure RNA Isolation (Roche) or TRIzol (Invitrogen). mRNA was isolated using Dynabeads (Dyna). Total RNA from placenta, heart, testis, and adult brain was obtained from Clontech, fetal brain RNA from Stratagene. Either 1  $\mu$ g of total RNA or 50 ng of mRNA was reverse transcribed using TaqMan and random hexamer primers (Roche) and the resulting cDNA was used for further analysis of alternative splicing and expression profiling.

### Analysis of alternative splicing

Occurrence of exons was controlled by PCR using *Taq* polymerase (Roche) under the following cycling condi-

tions: initial denaturation at 94°C for 2 min and then 30 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min. PCR fragments were analyzed by ethidium bromide staining of 1.5–2% agarose gels. Primers are listed in Supplementary Table S6. All primers were obtained from Invitrogen and tested for their functionality.

#### Quantitative real-time RT-PCR

Quantitative PCR was performed by LightCycler technology using SYBR Green I detection (Roche). In all assays, cDNA was amplified using a standardized program (10 min denaturing step; 55 cycles of 5 s at 95°C, 5 s at 65°C, and 15 s at 72°C; melting point analysis in 0.1°C steps; final cooling step). The final amount of cDNA per reaction corresponded to 5 ng of total RNA used for reverse transcription. Primer design, relative quantification of target gene expression, and PCR efficiency were performed as described previously [31]. As reference values, mean values of either  $\beta 2m$  or *HERC* expression were calculated. Results are shown as mean ( $n = 2-4$ ) values  $\pm$  SEM. Real-time PCR efficiencies were 1.85 for *HERC3*, 2.22 for *HERC4*, 2.17 for *HERC5*, 2.08 for *HERC6*, and 1.82 for  $\beta 2m$ . Primers are listed in Supplementary Table S6.

#### Immunofluorescence

HeLa cells grown on gelatin-coated glass coverslips were transfected with the respective *HERC* plasmids using Lipofectamine Plus (Invitrogen). Twenty hours posttransfection cells were fixed in 3% paraformaldehyde for 15 min at room temperature and permeabilized with 0.5% Triton X-100 for 10 min at room temperature. Samples were incubated with antibodies recognizing either Myc (Oncogene, 9E10) or FLAG tag (Stratagene, M2) followed by an Alexa 568-conjugate (Molecular Probes) and analyzed by a Zeiss LSM 510 confocal laser scanning microscope.

#### Acknowledgments

We thank Johannes Schmid and Rupert Ecker for their technical help with the confocal microscope and Peter Hufnagl for his helpful comments on LightCycler quantification. We also thank M. Ohtsubo, N. Nomura, T. Nagase, and S. Sugano for the *ceb1*, *KIAA0032*, *KIAA1593*, and *AK026808* cDNAs, respectively. This work was supported by BioMolecular Therapeutics and by grants to BRB (SFB F509) and JL (SFB F505) from the Austrian Science Foundation and was performed partially within the EU 6th framework program EVGN (contract number LSHM-CT-2003-503254) and the Center for Molecular Medicine (CeMM) of the Austrian Academy of Sciences.

#### Appendix A. Supplementary data

Supplementary data for this article may be found on ScienceDirect.

#### References

- [1] C.M. Pickart, Back to the future with ubiquitin, *Cell* 116 (2004) 181–190.
- [2] M.H. Glickman, A. Ciechanover, The ubiquitin–proteasome proteolytic pathway: destruction for the sake of construction, *Physiol. Rev.* 82 (2002) 373–428.
- [3] M. Muratani, W.P. Tansey, How the ubiquitin–proteasome system controls transcription, *Nat. Rev. Mol. Cell. Biol.* 4 (2003) 192–201.
- [4] P.P. Di Fiore, S. Polo, K. Hofmann, When ubiquitin meets ubiquitin receptors: a signalling connection, *Nat. Rev. Mol. Cell. Biol.* 4 (2003) 491–497.
- [5] V. Jesenberger, S. Jentsch, Deadly encounter: ubiquitin meets apoptosis, *Nat. Rev. Mol. Cell. Biol.* 3 (2002) 112–121.
- [6] M. Pagano, Cell cycle regulation by the ubiquitin pathway, *FASEB J.* 11 (1997) 1067–1075.
- [7] A.M. Weissman, Themes and variations on ubiquitylation, *Nat. Rev. Mol. Cell. Biol.* 2 (2001) 169–178.
- [8] L. Hicke, R. Dunn, Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins, *Annu. Rev. Cell Dev. Biol.* 19 (2003) 141–172.
- [9] C.A. Joazeiro, A.M. Weissman, RING finger proteins: mediators of ubiquitin ligase activity, *Cell* 102 (2000) 549–552.
- [10] S. Hatakeyama, K.I. Nakayama, U-box proteins as a new family of ubiquitin ligases, *Biochem. Biophys. Res. Commun.* 302 (2003) 635–645.
- [11] J.M. Huibregtse, M. Scheffner, S. Beaudenon, P.M. Howley, A family of proteins structurally and functionally related to the E6-AP ubiquitin–protein ligase, *Proc. Natl. Acad. Sci. USA* 92 (1995) 2563–2567.
- [12] M. Scheffner, J.M. Huibregtse, R.D. Vierstra, P.M. Howley, The HPV-16 E6 and E6-AP complex functions as a ubiquitin–protein ligase in the ubiquitination of p53, *Cell* 75 (1993) 495–505.
- [13] Y. Ji, et al., 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, *Hum. Mol. Genet.* 8 (1999) 533–542.
- [14] T. Nishimoto, A new role of ran GTPase, *Biochem. Biophys. Res. Commun.* 262 (1999) 571–574.
- [15] J.L. Rosa, R.P. Casaroli-Marano, A.J. Buckler, S. Vilaro, M. Barbacid, p619, a giant protein related to the chromosome condensation regulator RCC1, stimulates guanine nucleotide exchange on ARF1 and Rab proteins, *EMBO J.* 15 (1996) 4262–4273.
- [16] A.L. Lehman, et al., A very large protein with diverse functional motifs is deficient in *rjs* (runty, jerky, sterile) mice, *Proc. Natl. Acad. Sci. USA* 95 (1998) 9436–9441.
- [17] C. Cruz, et al., Assignment of the human P532 gene (*HERC1*) to chromosome 15q22 by fluorescence in situ hybridization, *Cytogenet. Cell Genet.* 86 (1999) 68–69.
- [18] C. Cruz, et al., The human *HERC3* gene maps to chromosome 4q21 by fluorescence in situ hybridization, *Cytogenet. Cell Genet.* 87 (1999) 263–264.
- [19] K. Mitsui, et al., A novel human gene encoding HECT domain and RCC1-like repeats interacts with cyclins and is potentially regulated by the tumor suppressor proteins, *Biochem. Biophys. Res. Commun.* 266 (1999) 115–122.
- [20] C. Cruz, F. Ventura, R. Bartrons, J.L. Rosa, *HERC3* binding to and regulation by ubiquitin, *FEBS Lett.* 488 (2001) 74–80.

- [21] R. Kroismayr, et al., HERC5, a HECT E3 ubiquitin ligase tightly regulated in LPS activated endothelial cells, *J. Cell Sci.* 117 (2004) 4749–4756.
- [22] S.E. Schwarz, J.L. Rosa, M. Scheffner, Characterization of human hect domain family members and their interaction with UbcH5 and UbcH7, *J. Biol. Chem.* 273 (1998) 12148–12154.
- [23] T. Nagase, R. Kikuno, M. Nakayama, M. Hirotsawa, O. Ohara, Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro, *DNA Res.* 7 (2000) 273–281.
- [24] R.L. Strausberg, et al., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences, *Proc. Natl. Acad. Sci. USA* 99 (2002) 16899–16903.
- [25] A. Bateman, et al., The Pfam protein families database, *Nucleic Acids Res.* 30 (2002) 276–280.
- [26] F.R. Garcia-Gonzalo, et al., Interaction between HERC1 and M2-type pyruvate kinase, *FEBS Lett.* 539 (2003) 78–84.
- [27] T.W. Gong, L. Huang, S.J. Warner, M.I. Lomax, Characterization of the human UBE3B gene: structure, expression, evolution, and alternative splicing, *Genomics* 82 (2003) 143–152.
- [28] Y. Ji, et al., Structure of the highly conserved HERC2 gene and of multiple partially duplicated paralogs in human, *Genome Res.* 10 (2000) 319–329.
- [29] J.L. Rosa, M. Barbacid, A giant protein that stimulates guanine nucleotide exchange on ARF1 and Rab proteins forms a cytosolic ternary complex with clathrin and Hsp70, *Oncogene* 15 (1997) 1–6.
- [30] N. Nomura, et al., Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001–KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1, *DNA Res.* 1 (1994) 27–35.
- [31] A. Kadl, et al., Analysis of inflammatory gene induction by oxidized phospholipids in vivo by quantitative real-time RT-PCR in comparison with effects of LPS, *Vasc. Pharmacol.* 38 (2002) 219–227.