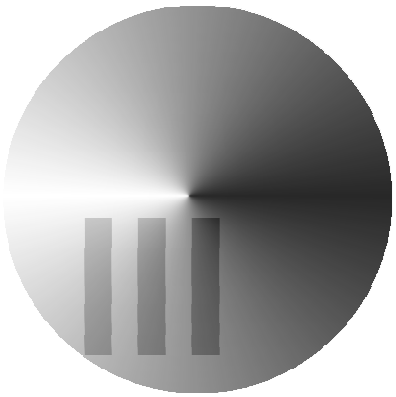
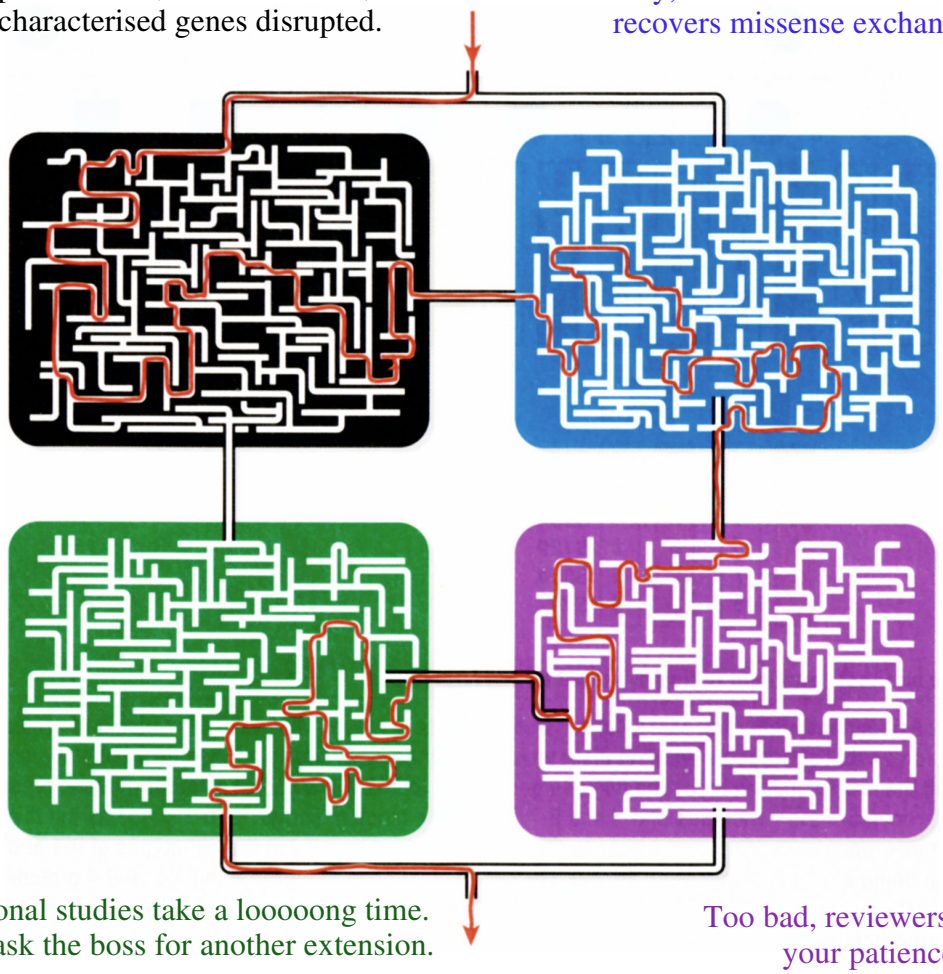


Results

The Lab Rat's Maze

Breakpoint cloned, but ... bad luck;
two uncharacterised genes disrupted.

Nice try, but mutation screening only
recovers missense exchanges.



Functional studies take a looooong time.
Better ask the boss for another extension.

Too bad, reviewers stretch
your patience.

Idea adapted from Dr. R. Smith, University of Iowa, USA.

A. Patient with mental retardation and a *de novo* 46,X,t(X;8)(p11.2;p22.3) balanced translocation

As a starting point in our search for genes involved in MR, we set out to establish a phenotype – genotype correlation in a mildly mentally retarded female patient carrying an apparently balanced X;A translocation. Clinical examination determined the phenotype; karyotyping and molecular analysis of the chromosomal rearrangement established the genotype.

A.1. Clinical characterisation

The subject is mildly mentally retarded and suffers from idiopathic epileptic seizures.

The patient was first examined at the University Hospital Ulm, Germany in 1997. She is the first child of healthy, non-consanguineous, Caucasian parents. A cousin of the mother had idiopathic epilepsy, otherwise the family history was unremarkable and the two younger siblings were healthy. The patient was born spontaneously 10 days after term with a birth weight of 3600 g, a length of 51 cm and a head circumference of 36 cm (all 75th percentile). The pregnancy was complicated by bleeding in the 8th week of gestation. Early motor development was normal. However, with the onset of seizures at the age of 9 months, further motor and mental development were delayed. Unassisted walking was achieved at the age of 15 months, but frequent falling remained a problem for several years. At the age of 8 years, she suffered grand mal seizures, which were initially treated with Primidone. Later, bouts of aggressive behaviour, as well as predominance of focal seizures, appeared. Therefore, treatment was continued with Carbamazepine. At the age of 11 years, growth parameters were normal, with a height of 146 cm (50th percentile), a weight of 32 kg (25th percentile) and a head circumference of 53.5 cm (~50th percentile). Clinical examination revealed only a very mild facial dysmorphology: a long face with mild facial asymmetry, long eyelashes and large ears. Furthermore, non-specific clinical features, such as high palate, mild hyperextensibility of fingers, mild clinodactyly V and cutaneous syndactyly II/III, were noted. The girl presented with mild to moderate MR. She was noted to be quiet and anxious, and could not perform complex requests. Her fine motor skills were poor, but gross neurological examination

was normal. Brain MRI scan was also normal. Until the age of 10, she attended a school for children with learning disabilities. She was then admitted to a school for mentally retarded children, which she attended until she was 20. Since 2006, she has been employed in a sheltered workplace for the mentally handicapped. A lymphoblastoid cell line is available.

A.2. Karyotyping

Courtesy of Dr. G. Barbi, University of Ulm, Ulm, Germany.

The patient carries a *de novo* 46,X,t(X;8)(p11.2;p22.3) balanced translocation.

Standard karyotyping on metaphase spreads from the patient revealed a 46,X,t(X;8)(p11.2;p22.3) apparently balanced translocation (Fig. III-1). Karyotyping of 30 metaphase spreads of each of the patient's parents determined that the translocation was *de novo* and excluded the possibility of a structural mosaic.

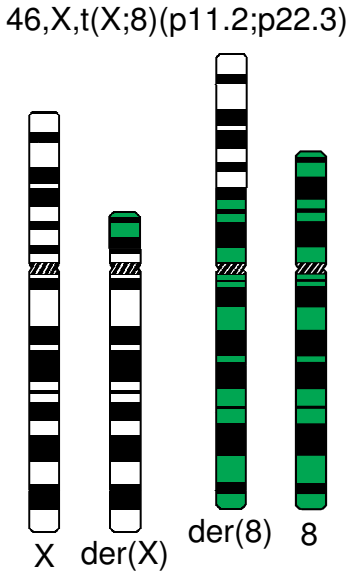


Fig. III-1 | **Ideograms of 46,X,t(X;8)(p11.2;p22.3).**

The ideograms represent normal chromosomes X and 8, and their derivatives der(X) and der(8), which are found in a mildly mentally retarded patient carrying a *de novo* apparently balanced reciprocal 46,X,t(X;8)(p11.2;p22.3) translocation.

A.3. Cytogenetic, molecular and computational analyses

To clarify the patient's genotype, we performed molecular analysis of the apparently balanced t(X;8) translocation, characterising the X-chromosomal and autosomal BPs.

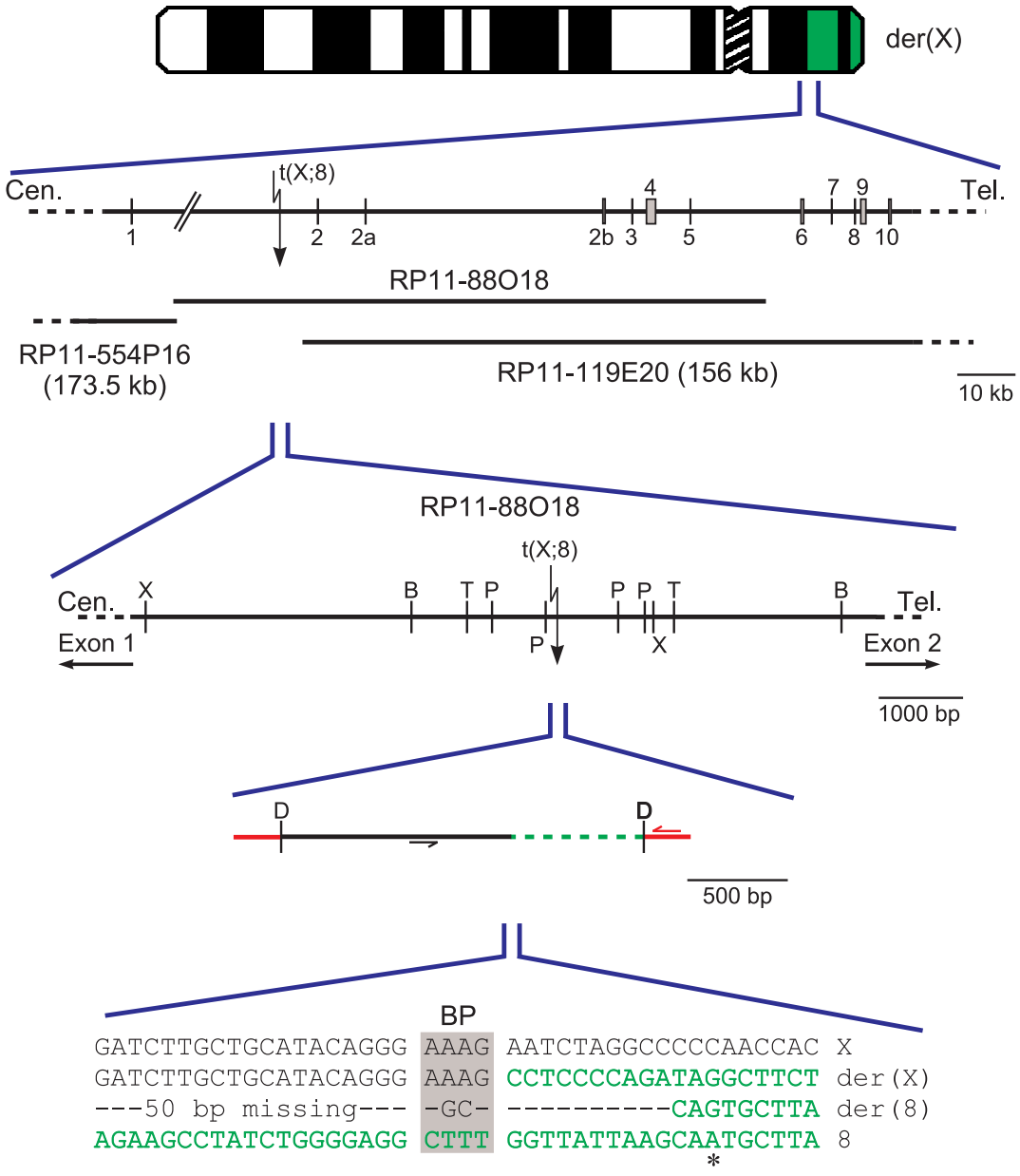


Fig. III-2 | **Positional cloning overview.**

The overview shows the successive steps in the positional cloning approach that was employed to localise the 46,X,t(X;8)(p11.2;p22.3) translocation breakpoint (BP) on der(X). Top to bottom: karyotyping, fluorescent *in situ* hybridisation with bacterial artificial chromosomes resulting in a contig of genomic clones covering the BP region, Southern hybridisation narrowing down the BP, cloning of the chromosome X – chromosome 8 junction fragment by suppression polymerase chain reaction, and sequencing to pinpoint the BP at the single-nucleotide level. Additional figures specify each of these stages (mentioned on the left).

A.3.1. X-chromosomal breakpoint cloning

In order to characterise the translocation's X-chromosomal BP down to the single nucleotide level, we consecutively applied FISH, Southern hybridisation and suppression PCR, and performed *in silico* studies. To determine the sequence flanking the BPs on der(X) and der(8),

we employed BPS PCR. An overview of this positional cloning approach is shown in Fig. III-2.

A.3.1.1. Fluorescent *in situ* hybridisation on chromosome X

Courtesy of C. Menzel and L. Van Zutven, MPI-MG, Berlin, Germany.

BAC RP11-88O18, containing part of the *hKIAA1202* gene, spans the X-chromosomal BP.

Performing FISH on metaphase chromosomes from the 46,X,t(X;8)(p11.2;p22.3) lymphoblastoid cell line with labelled X-chromosomal YAC clones identified ICRF A09026 as a BPS clone. Subsequent FISH with labelled BAC and PAC clones covering the entire ICRF A09026 insert yielded a signal proximal to the BP for BAC clone RP11-554P16, a BPS signal for BAC RP11-88O18 and a signal distal to the BP for BAC RP11-119E20 (Fig. III-3a). From sequence alignments with these three BAC clones, it became apparent that the

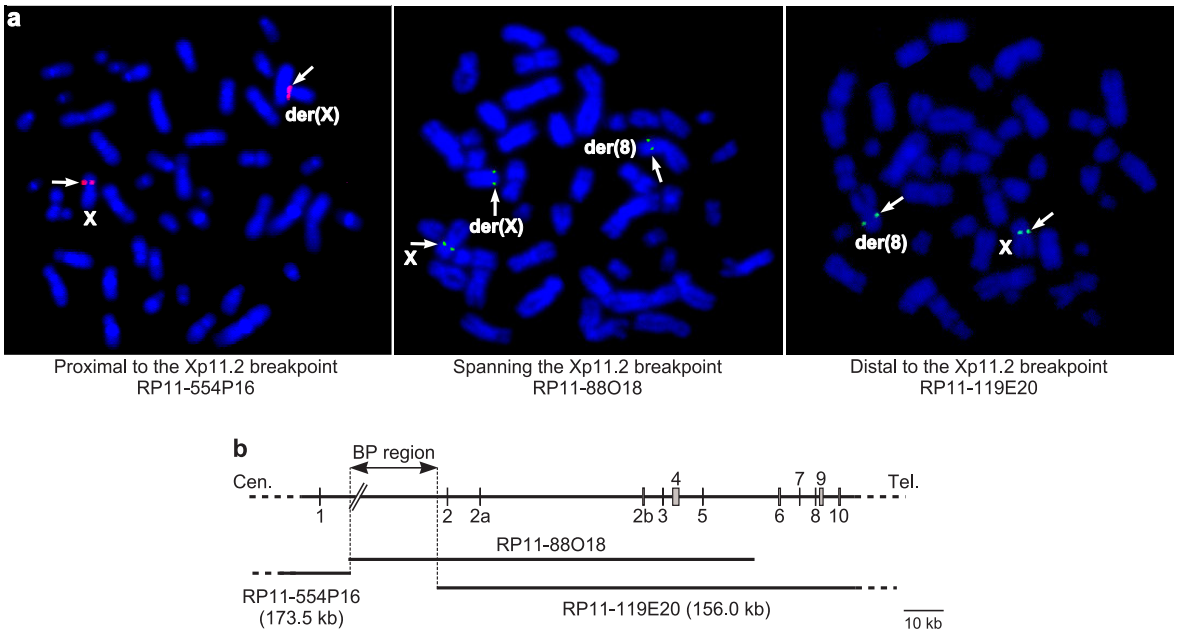


Fig. III-3 | **Fluorescent *in situ* hybridisation on chromosome X narrows down the breakpoint region.**

a. While both the proximal and distal bacterial artificial chromosomes (BACs) show hybridisation signals on two chromosomes (left and right panels, respectively), the breakpoint (BP)-spanning BAC RP11-88O18 yields a signal on the normal X, and a split signal on der(X) and der(8) (middle panel).
b. Sequence analysis of the clones hybridised in panel a point towards a 31 kb region that should contain the X-chromosomal BP ('BP region'). Black lines represent intronic sequences and BACs. Intron 1, 118 kb, is only partially shown (double slash), as are clones RP11-554P16 and RP11-119E20 (dashed lines). Exons are denoted by grey rectangles (resolution permitting). Cen., centromere; Tel., telomere.

Photographs courtesy of C. Menzel, MPI-MG, Berlin, Germany (a).

BP should be located in a 31 kb interval on BAC RP11-88O18, defined by the distal end of RP11-554P16 and the proximal end of RP11-119E20 (Fig. III-3b). NIX analysis of the 31 kb

sequence presumably containing the BP revealed that this sequence lies within the first intron of the uncharacterised *hKIAA1202* gene. While BAC RP11-554P16 encompasses *hKIAA1202* exon 1, the remaining eleven exons are located on clones RP11-88O18 and RP11-119E20 (Fig. III-3b). Results for FISH-mapped clones are listed in Table III-1.

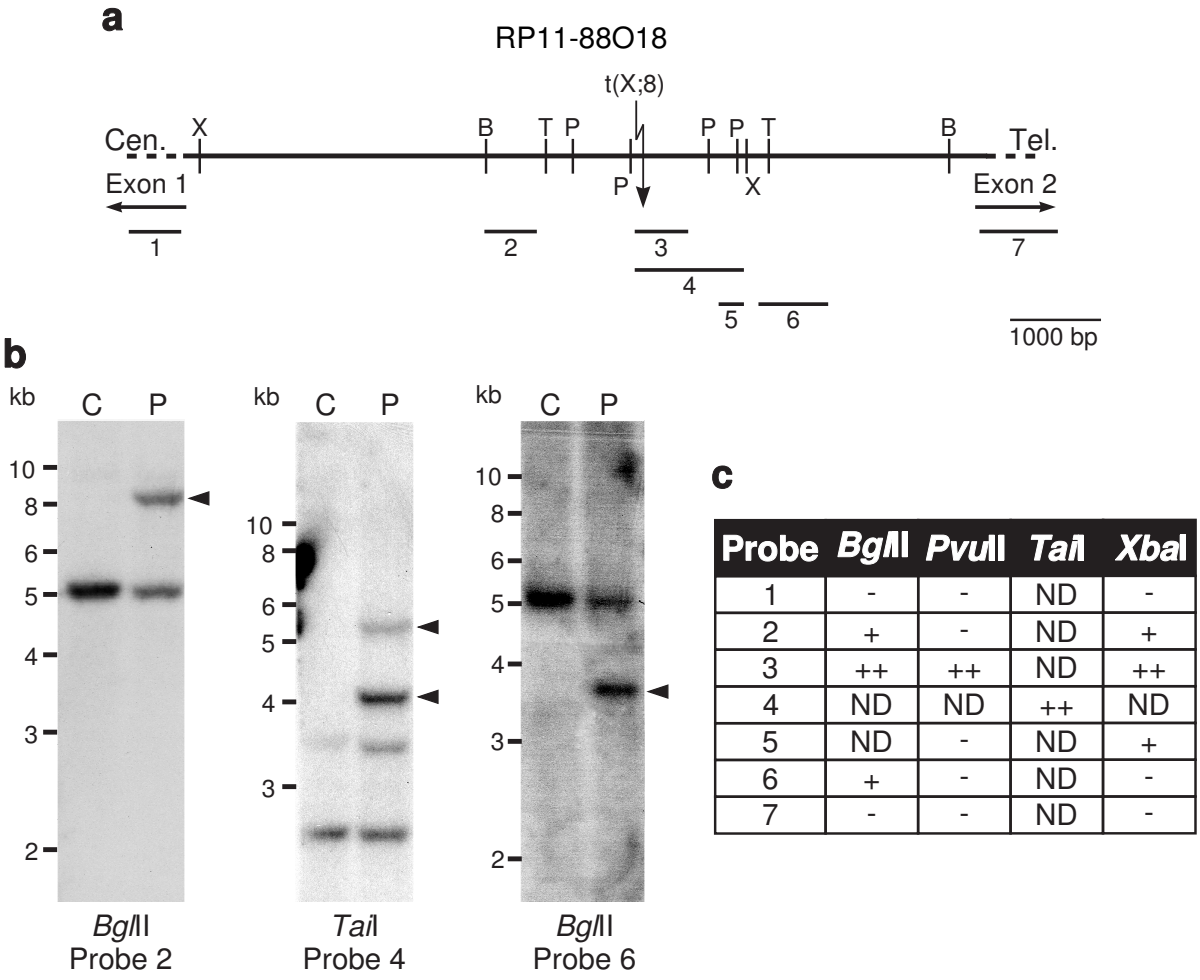
Table III-1 | **FISH mapping results for X-chromosomal YAC, BAC and PAC clones on meta-phases from a 46,X,t(X;8)(p11.2;p22.3) lymphoblastoid cell line**

YAC	**Position on X (cM/cR)**	**Position on X with regard to the BP**
ICRF A02015	~87 cM	Proximal
ICRF D1245	NA	Proximal
ICRF B09128	NA	Proximal
WI 871D6	87 cM – 107 cR	Proximal
WI 810E6	~108 cR	Proximal
ICRF A09026	NA	BPS
BAC/PAC	**Accession #**	**Position on X with regard to the BP**
RP4-589C16	NA	Distal
RP4-620F10	NA	Distal
RP5-1120I12	NA	Distal
RP3-420H8	NA	Distal
RP4-625P15	NA	Distal
RP4-703H20	NA	Distal
RP5-1128E2	NA	Distal
RP4-604E15	NA	Distal
RP11-119E20	AL121865	Distal
RP1-127I1	NA	BPS
RP11-88O18	AL445491	BPS
RP11-554P16	AL359272	Proximal
RP5-1099E4	NA	Proximal
RP4-645A16	NA	Proximal
RP4-589M19	NA	Proximal
RP4-589N19	NA	Proximal
RP4-589O22	NA	Proximal
RP4-807P11	NA	Proximal
RP5-1150M23	NA	Proximal

A.3.1.2. Southern hybridisation on chromosome X

The X-chromosomal BP is located in the first *hKIAA1202* intron.

To further narrow down the X-chromosomal 46,X,t(X;8)(p11.2;p22.3) BP, we employed Southern hybridisation. *Bgl*II-, *Pvu*II-, *Tai*I-, and *Xba*I-digested, size-separated genomic DNA preparations from the patient and a control were hybridised with seven RA labelled probes (Fig. III-4a). These probes were PCR-amplified from control genomic DNA



Probe	*Bgl*II	*Pvu*II	*Tai*I	*Xba*I
1	-	-	ND	-
2	+	-	ND	+
3	++	++	ND	++
4	ND	ND	++	ND
5	ND	-	ND	+
6	+	-	ND	-
7	-	-	ND	-

Fig. III-4 | **Southern hybridisation on chromosome X.**

a. A restriction map (B, *Bgl*II; P, *Pvu*II; T, *Tai*I; X, *Xba*I) of clone RP11-88O18 spanning the breakpoint (jagged arrow) on derivative chromosome X is shown from centromere (Cen.) to telomere (Tel.). It shows the positions of seven Southern blot probes (probes 1 – 7).
b. Representative Southern hybridisations on control- and patient-digested (C and P, respectively) genomic DNA, using probes 2, 4 and 6. Arrowheads indicate patient-specific aberrant bands. Size markers on the left.
c. Tabular representation of Southern hybridisation results with all seven probes. -, no aberrant band observed; +, one aberrant band observed; ++, two aberrant bands observed.

using primers as listed in Appendix E. Examples of the resulting hybridisation pattern, characterised by aberrant bands due to the BP, are represented in Fig. III-4b. From this pattern

and the genomic sequence, we deduced that the BP should be situated in an 831 bp *Pvu*II fragment within the first *hKIAA1202* intron (Fig. III-4c). This result is in line with the observations obtained through FISH analysis.

A.3.1.3. Suppression polymerase chain reaction

46,X,t(X;8)(p11.2;p22.3) disrupts *hKIAA1202* on the X chromosome and *hFBXO25* on chromosome 8.

To narrow down the X-chromosomal 46,X,t(X;8)(p11.2;p22.3) BP to the nucleotide level, we employed suppression PCR[756]. This technique also allowed us to clone the unknown chromosome 8 BP sequence, thus circumventing time-consuming characterisation of the autosomal BP as described above, i.e. through extensive FISH mapping and Southern hybridisation. We digested patient and control genomic DNA samples with *Dra*I and ligated

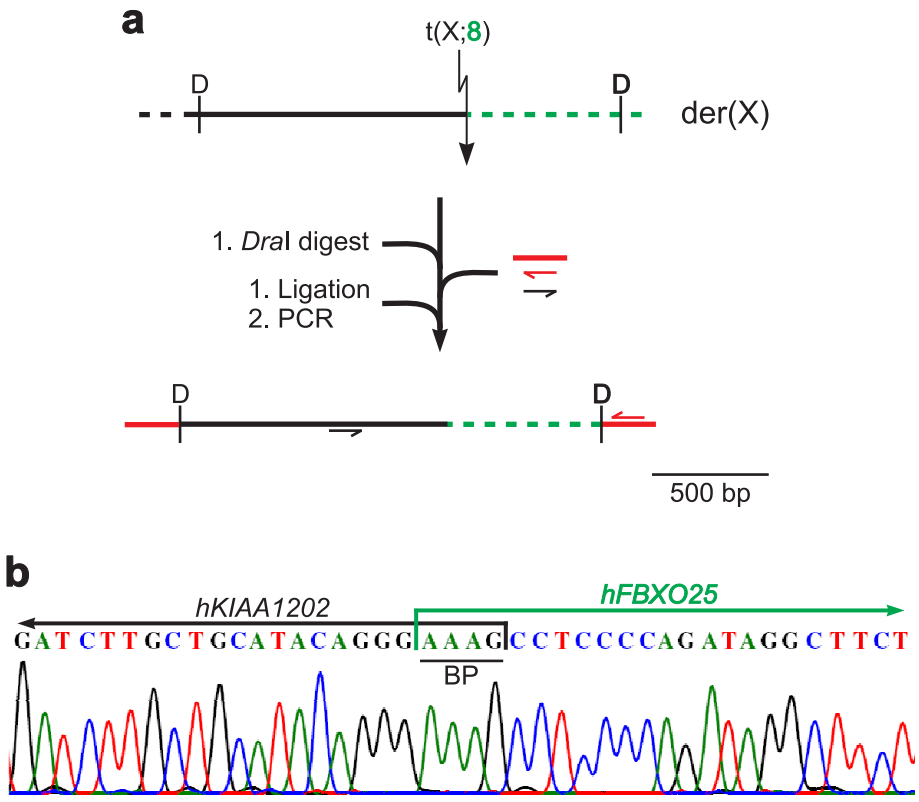


Fig. III-5 | **Suppression polymerase chain reaction.**

a. To clone the breakpoint (BP, jagged arrow) on derivative chromosome X [der(X)], genomic patient DNA was digested (D, *Dra*I), ligated to an adaptor (red bar), and amplified with a gene-specific (black arrow) and an adaptor-specific (red arrow) primer. Note that the chromosome 8 (dashed green line) sequence adjacent to the X-chromosomal sequence (black line) is unknown.
b. Sequencing of the PCR product obtained as described in panel a reveals the exact location of the BP and shows that it disrupts *hKIAA1202* and *hFBXO25*.

an adaptor to the restriction fragments. Subsequent suppression PCR with a primer specific to the adaptor and a primer specific to der(X) sequence flanking the BP allowed the amplification of a patient-specific *Dra*I fragment (Fig. III-5a). Primers are listed in Appendix E. BLAST analyses of the resulting sequence (Fig. III-5b) revealed that the X-chromosomal BP indeed disrupts the uncharacterised *hKIAA1202* gene[757] between exons 1 and 2, and that the chromosome 8 BP disrupts the uncharacterised *hFBXO25* gene[758] between exons 5 and 6[759].

A.3.2. Chromosome 8 breakpoint cloning

The identification of autosomal sequence flanking the der(X) BP allowed a straightforward verification of the results obtained from the X-chromosomal BP cloning and the identification of the sequence flanking the der(8) BP.

A.3.2.1. Fluorescent *in situ* hybridisation on chromosome 8

Courtesy of C. Menzel, MPI-MG, Berlin, Germany.

BAC RP11-91J19 containing the entire *hFBXO25* gene spans the chromosome 8 BP.

BLAST analysis with the autosomal sequence from the BP clone (see previous section) identified BAC clone RP11-91J19 (AC083964), which contains the entire uncharacterised *hFBXO25* gene. Indeed, FISH-mapping with this obligate chromosome 8 BPS clone resulted in a signal on chromosome 8 and a split signal on derivative chromosomes X and 8, which is indicative of its BPS nature (Fig. III-6).

A.3.2.2. Southern hybridisation on chromosome 8

The chromosome 8 BP is located in *hFBXO25* intron 5.

As a last verification of the localisation of the BP, we PCR-amplified a chromosome 8-specific probe (primers are listed in Appendix E) that should span the der(X) BP, and used it in Southern hybridisation on *Bgl*II-, *Dra*I-, and *Xba*I-digested patient and control genomic DNA preparations. As expected, we recovered 6730 bp (*Bgl*II), 2522 bp (*Dra*I) and 3037 bp (*Xba*I) fragments in control and patient DNA, but only in patient DNA did we recover 3707 bp and 8144 bp (*Bgl*II), 2615 bp and 2005 bp (*Dra*I), and 7250 bp and 1793 bp (*Xba*I) BP fragments. It should be noted that the resolution of Southern blot analysis did not suffice to separate the 2522 bp and 2615 bp *Dra*I bands satisfactorily (Fig. III-7). This result proved that the autosomal 46,X,t(X;8)(p11.2;p22.3) BP indeed disrupts the *hFBXO25* gene in intron 5.

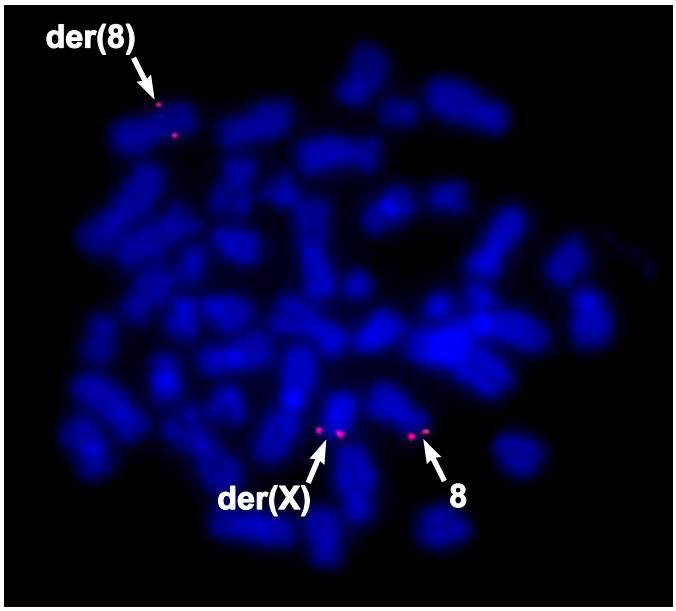


Fig. III-6 | **Fluorescent *in situ* hybridisation on chromosome 8.**

Fluorescent *in situ* hybridisation with the obligatory breakpoint (BP)-spanning bacterial artificial chromosome RP11-91J19 on a metaphase spread of the patient's lymphoblastoid cell line yields signals on chromosome 8, and on derivative chromosomes X and 8 [der(X) and der(8), respectively], which indicates that it indeed spans the BP on der(8).

A.3.2.3. Breakpoint-spanning polymerase chain reaction

Sequences of chromosomes X, 8 and der(X) are as expected. On der(8), ~60 bp are deleted.

In order to clone the der(8) BP and to verify the der(X) BP, we designed primers at either side of the 46,X,t(X;8)(p11.2;p22.3) BPs and set up straightforward PCR reactions across both BPs on patient and control genomic DNA. Primers are listed in Appendix E. As expected, these amplifications resulted in a product in all reactions apart from the BPS PCRs in the control DNA (Fig. III-8a). Sequencing of the PCR products yielded the sequence expected for chromosomes X, 8 and der(X), but revealed a sixty-base pair deletion on der(8). Fifty base pairs of X-chromosomal and ten base pairs of chromosome 8 sequence are missing, and a GC dinucleotide, neither mapping to chromosome X nor to chromosome 8, is inserted at the position of the BP (Fig. III-8b).

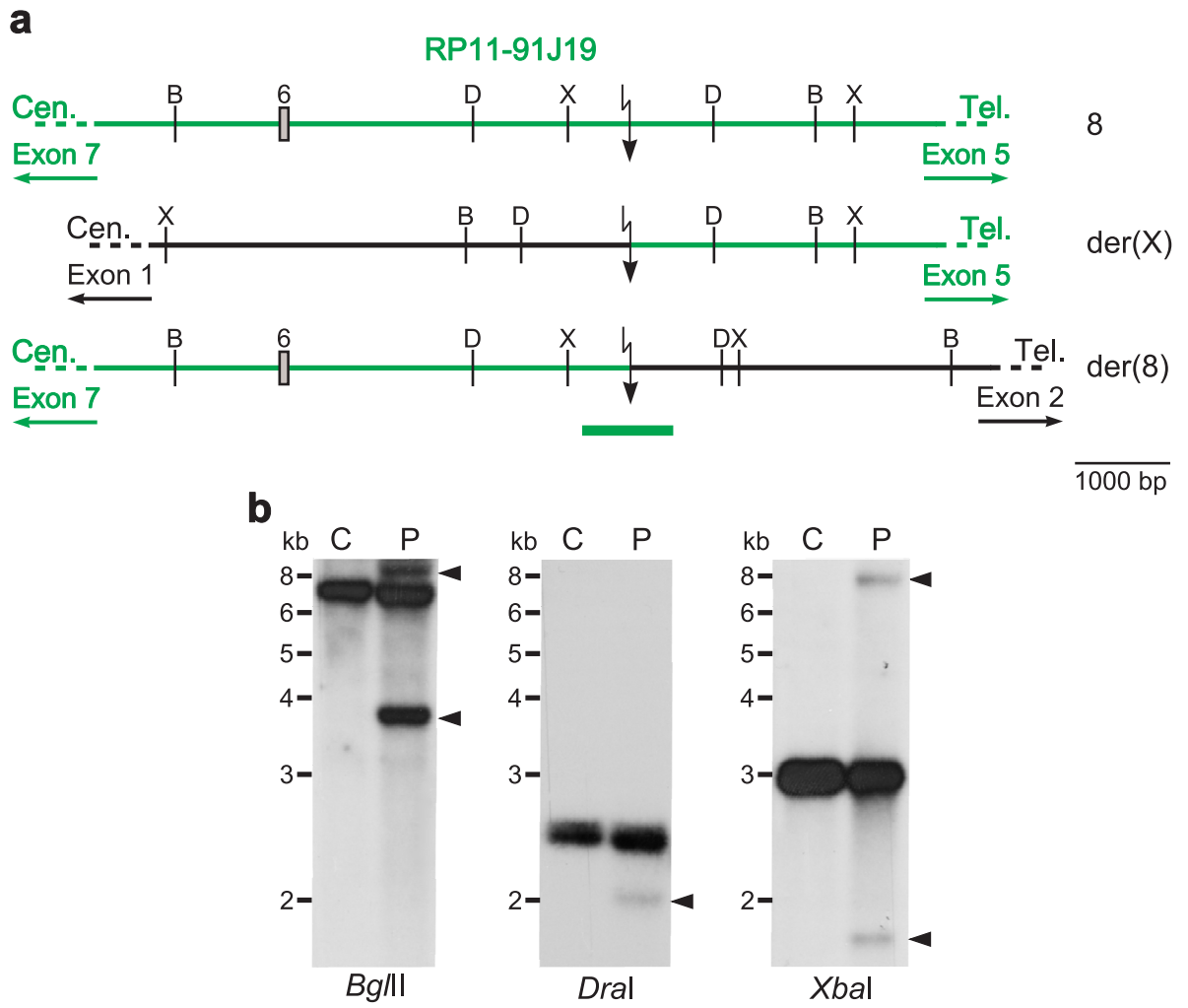


Fig. III-7 | **Southern hybridisation on chromosome 8.**

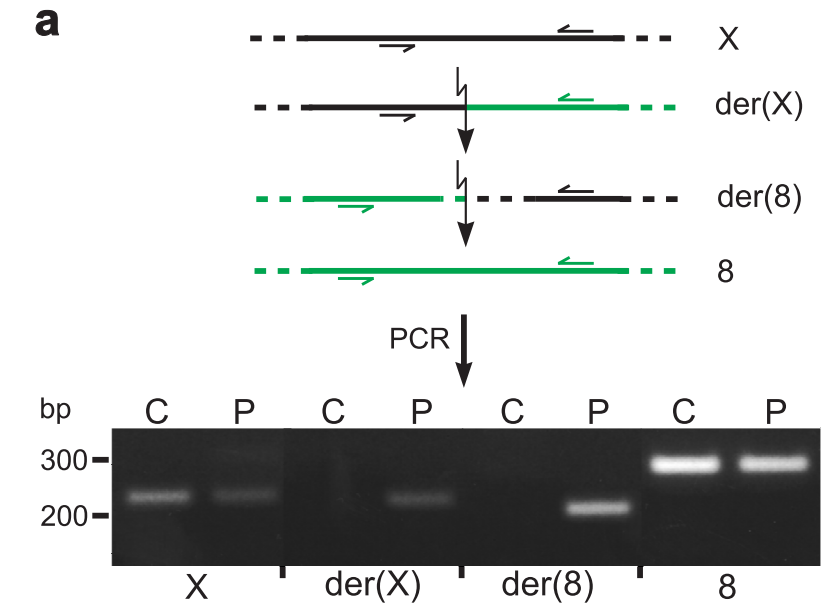
a. The scheme shows a restriction map (B, *Bgl*II; D, *Dra*I; X, *Xba*I) from centromere (Cen.) to telomere (Tel.) of clone RP11-91J19, which is derived from chromosome 8 and spans the breakpoint (BP, jagged arrow) on derivative chromosome 8 [der(8)]. Black and green denote sequences from *hKIAA1202* (X chromosome) and *hFBXO25* (chromosome 8), respectively. A grey rectangle depicts *hFBXO25* exon 6, the green bar represents a chromosome 8-specific BP-spanning probe.
b. Southern hybridisations on control (C) and patient (P) digested genomic DNA, using the probe shown in panel a. Arrowheads indicate patient-specific aberrant bands; see text for an explanation of their pattern. Size markers on the left.

A.4. Computational analysis

Formation of an *hFBXO25 – hKIAA1202* fusion gene is unlikely.

Extensive computational analysis of the genomic sequence around the der(X) and der(8) BPs using NIX analysis of BPS BAC clones and genome browsing on the servers of the UCSC and NCBI indicated that *hFBXO25* is transcribed from telomere to centromere on 8p23.3, whereas *hKIAA1202* is transcribed from centromere to telomere on Xp11.22, leading to a head-to-head fusion of both genes on der(8) and a tail-to-tail fusion on der(X) (Fig. III-9). Presence of fusion transcripts seems unlikely; however, it could not be verified experi-

mentally, as *hKIAA1202* isoforms I – IV are not expressed in lymphoblastoid cells (see III.C.1.2.1).



b

```
                          BP
GATCTTGCTGCATACAGGG AAAG AATCTAGGCCCCCAACCAC X
GATCTTGCTGCATACAGGG AAAG CCTCCCCAGATAGGCTTCT der(X)
---50 bp missing--- -GC- ----------CAGTGCTTA der(8)
AGAAGCCTATCTGGGGAGG CTTT GGTTATTAAGCAATGCTTA 8
                                          *
```

Fig. III-8 | **Breakpoint-spanning polymerase chain reaction.**

a. The top panel depicts a standard polymerase chain reaction (PCR) across the breakpoint (BP, jagged arrow) on genomic DNA with primers specific to chromosomes X (black arrows) and 8 (green arrows). Black and green denote sequences from *hKIAA1202* (X chromosome) and *hFBXO25* (chromosome 8), respectively. Derivative chromosome 8 [der(8)] carries a deletion of sixty bases (not drawn to scale, internal dashed line; see panel b). The bottom panel shows size-separated amplicons of the patient (P) and a control (C). Note the absence of PCR product in the control lanes for the derivative chromosomes. Size marker on the left.
b. Sequencing of all amplicons from the patient shown in panel a pinpoints the BP to a four-base pair region, shadowed in grey. Der(8) misses fifty bases of X-chromosomal and ten bases of chromosome 8 sequence. A GC dinucleotide, neither mapping to chromosome X nor to chromosome 8, is inserted. The patient is heterozygous for a chromosome 8 A/G polymorphism (asterisk).

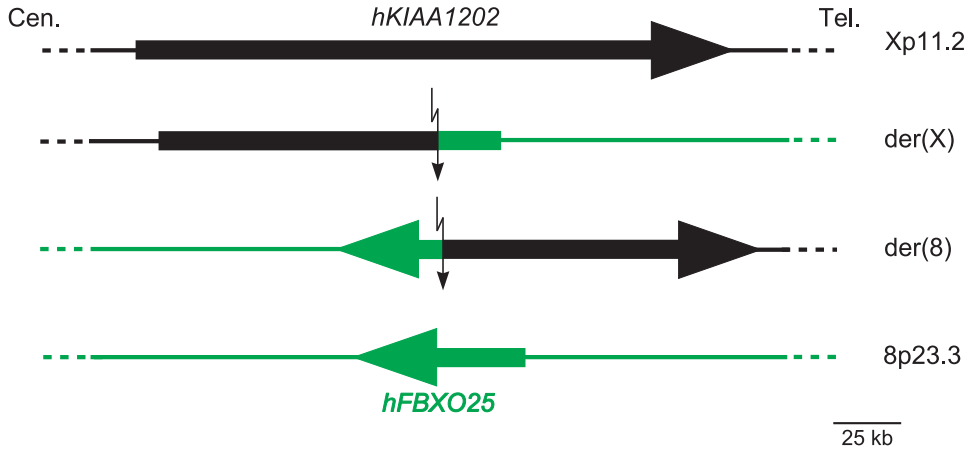


Fig. III-9 | **Genomic organisation at the 46,X,t(X;8)(p11.2;p22.3) breakpoints.**

The diagram shows *hKIAA1202* (black arrow) on Xp11.2 (black lines), *hFBXO25* (green arrow) on 8p23.3 (green lines) and their tail-to-tail and head-to-head fusions on the derivative chromosomes X [der(X)] and 8 [der(8)], respectively. The jagged arrow indicates the breakpoint. Tel., telomere and Cen., centromere.

B. Molecular and computational characterisation of Fbxo25

We established the phenotype and the genotype of the t(X;8) carrier and characterised Fbxo25 on the DNA, RNA and protein levels. We also performed a thorough *in silico* analysis.

B.1. Nucleic acid studies

We determined the organisation of the *Fbxo25* gene and investigated its expression profile.

B.1.1. Genomic organisation of *Fbxo25*

The exonic structure of the *hFBXO25* gene and that of its murine homologue, *mFbxo25*, were studied.

B.1.1.1. Genomic organisation of *hFBXO25*

hFBXO25, which encodes ORFs of 367, 358 and 291 AA, is organised into eleven exons, at least two of which are alternatively spliced.

In silico compilation of a tiling path of *hFBXO25*-related, expressed sequences, using the GAP4[760] and GeneNest[740] sequence alignment tools, resulted in a total of 2655 bp of exonic sequence. Three *hFBXO25* mRNAs (BC050393, AB070270 and BC020249), the Hs.81001.1 Unigene cluster extending the mRNA-based exonic sequence towards the 5' end and an EST (BE834573) extending the exonic sequence towards the 3' end were particularly

informative. 5' and 3' RACE experiments yielded *hFBXO25*-specific products but did not result in additional sequence information. Comparison of the transcribed sequence with the corresponding genomic clone RP11-91J19 established the genomic structure of *hFBXO25*.

hFBXO25 encompasses a 63 kb region on 8p23.3 and codes for 11 exons (Fig. III-10a and Table III-2). In the process of confirming the intron – exon boundaries by inter-exon RT-PCR, we discovered a coding exon, denoted exon 9a (Fig. III-10a), that is not present in the first deposited *hFBXO25* database entry[758], and amplified three alternatively spliced transcripts, translated into ORFs 1 – 3 (Fig. III-10b). ORF-1, which lacks exon 9a, codes for a 358 AA protein; ORF-2, which skips exons 4 and 9a, encodes a 291 AA peptide, and ORF-3 contains the alternatively spliced exon 9a and encodes a 367 AA protein.

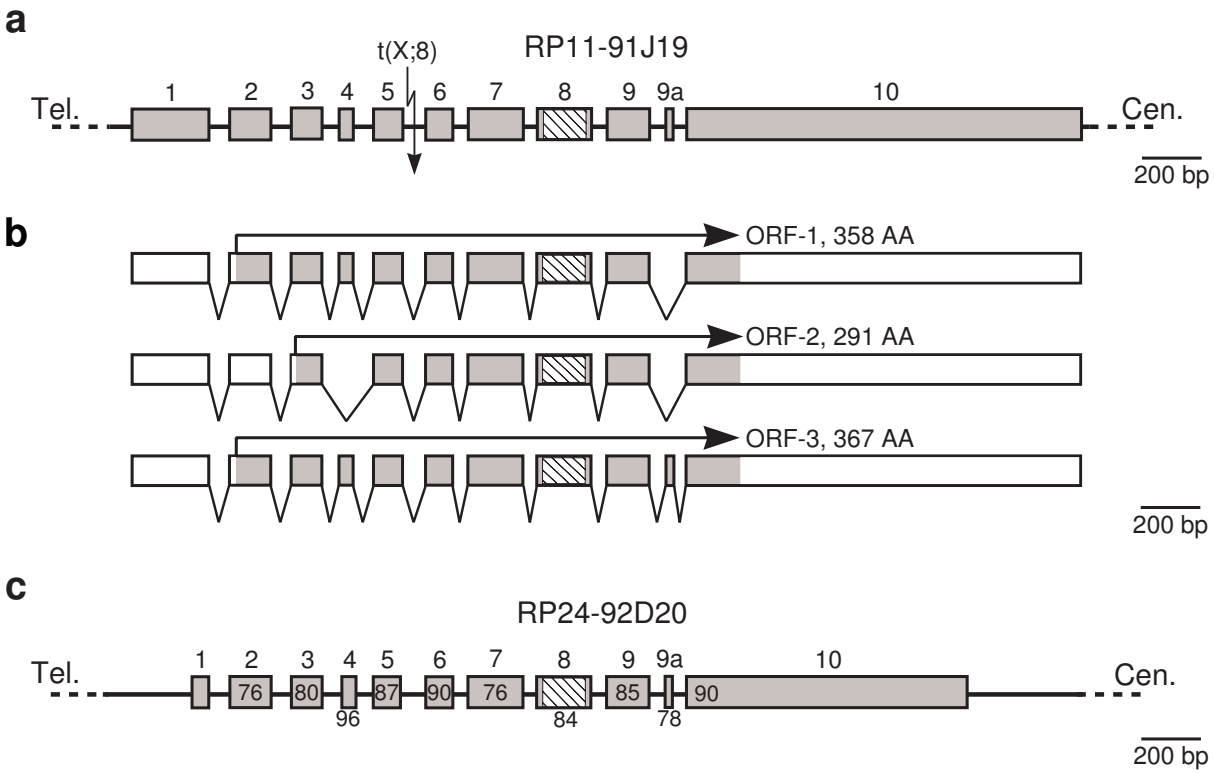


Fig. III-10 | **Genomic organisation of the *hFBXO25* and *mFbxo25* transcripts.**

a. Clone RP11-91J19 harbours the complete *hFBXO25* gene. Grey rectangles denote exons and black lines represent intronic sequences (not drawn to scale). Exon 8 contains a stretch of sequence that encodes an F-box domain (dashed area). The jagged arrow indicates the chromosome 8 breakpoint. Tel., telomere and Cen., centromere.
b. Overview of different splice variants of the *hFBXO25* gene. Hooked arrows represent open reading frames (ORFs) that were amplified using the reverse transcription polymerase chain reaction, open rectangles untranslated regions (UTRs). AA, amino acids.
c. Genomic organisation of the *mFbxo25* gene. Numbers within or underneath the exons indicate the percentage identity between the human and mouse exonic sequences. UTRs are not conserved.

Table III-2 | **Exon numbers, splice site sequences and exon sizes of the human and mouse *Fbxo25* genes**

Exon	*hFBXO25*		*mFbxo25*		Exon size (bp)[†]
	Splice acceptor site[§]	Splice donor site[§]	Splice acceptor site[§]	Splice donor site[§]	
1		GGCGCGCTAG<u>gt</u>aggtgcgg		GGACCCGCAG<u>gt</u>gagaaggg	259/59
2	ttttt<u>ag</u>GAGAACTATG	GTCACAGCAT<u>gt</u>aagttaca	tacctt<u>ag</u>GAGCACGATG	ACCACAGCAT<u>gt</u>gagtccct	141/141
3	tttatttc<u>ag</u>TATCTTAAAT	AATACTCAAA<u>gt</u>aagatcat	tttatttc<u>ag</u>TATCTTAAAT	GCTGCTCACA<u>gt</u>aagagtgt	104/101
4	tttattct<u>ag</u>GTTTTTATCG	CACAAAGGAA<u>gt</u>aagtattc	tttgttct<u>ag</u>GTTTTTATCG	CACAAAGGAA<u>gt</u>aagtgtgc	50/50
5	gcttttcc<u>ag</u>AGGCATGGCT	TGTGGTCAAA<u>gt</u>aagtgttc	tgtcttgt<u>ag</u>AGGCATGGCT	CGTGGTCAAA<u>gt</u>aagtggtt	93/93
6	ttcctta<u>ag</u>CTGTTGCAGC	GTTCAAAAGG<u>gt</u>aagatgat	tttcttca<u>ag</u>CTACTGCAGC	GTCCAAAAGG<u>gt</u>atgattat	94/94
7	tttattcc<u>ag</u>TTCTTGATGA	GATGACTAAG<u>gt</u>ataaatat	ggtttttc<u>ag</u>TTCTTGATGA	GGTGACAAAG<u>gt</u>aggtttct	185/185
8	ttcctttc<u>ag</u>CAAGTGAACA	TGAAAAGCAG<u>gt</u>gagtggga	ttcctttc<u>ag</u>CAGGTAAACA	TGAGCAGCAG<u>gt</u>gagtgcgg	183/183
9	gtttgttt<u>ag</u>TTTTGTAGAC	CTTTTGGAAG<u>gt</u>actgattt	tcttttc<u>ag</u>TTTTGTAGAC	CTTCTGGAAG<u>gt</u>aaggactt	144/144
9a	gatctcga<u>ag</u>GACTACCATC	ACTATTCAAG<u>gt</u>acttccct	ggtgtcca<u>ag</u>GACTACCACC	GTTGTTGAAG<u>gt</u>acttcctg	27/27
10	ctccccac<u>ag</u>GACTCAGGAC		ctctcccc<u>ag</u>GACTCGGGAC		1327/959

[§] Conserved ag and gt dinucleotides at the splice acceptor and splice donor sites, respectively, are underscored.
[†] Exon sizes are given in the format *hFBXO25*/*mFbxo25*.

Multiple entries in the sequence databases support the existence of an 18 bp 5' extension of exon 2. As we never recovered experimental evidence for this putative extension, we did not continue to study it.

The original *hFBXO25* mRNA database entry contains a 30 bp, in-frame, 5' extension of exon 5. RT-PCR experiments on total fibroblast and foetal brain RNAs and on poly-A^+ fibroblast RNA with forward primers (partly) specific to the extension and gene-specific reverse primers confirmed the *in silico* data. Although 5' RACE experiments, using a primer located in the extension, yielded gene-specific amplicons, no additional sequence upstream of the extension was recovered. Taken together, these results suggest an alternative *hFBXO25* transcriptional start 30 bp upstream of exon 5. The putative translational start could be the first downstream ATG codon at the 3' end of exon 7, coding for a 141 AA ORF (ORF-4). However, this ATG codon, and all others downstream of it, are not embedded in a Kozak sequence[761]. Of notice is that the 30 bp extension is only present four times in the EST database (AA463756, N42688, DA32682 and CD515429), whereas a multitude of entries exists for the other isoforms. Due to the apparently low abundance of this putative alternatively spliced transcript and its elusive ORF, we decided not to investigate it any further. Finally, database searches revealed alternative 3' ends of the *hFBXO25* transcript. While mRNA BC020249 is poly-adenylated at position 485 in exon 10, BC050393's poly-A tail starts at position 1309 in exon 10.

B.1.1.2. Genomic organisation of *mFbxo25*

mFbxo25 is organised into eleven exons, encompassing ~33 kb on chromosome 8.

Based on sequence homology with *hFBXO25*, we identified the *mFbxo25* transcript. Inter-exon RT-PCR and 5' and 3' RACE experiments yielded a total of 2036 bp, encoding eleven exons. Comparison of the transcript with mouse genomic clone RP24-92D20 (AC139186) allowed us to establish the genomic organisation of *mFbxo25* (Fig. III-10c and Table III-2). The gene spans 32.7 kb on mouse chromosome 8A1.1. The *mFbxo25* transcript is 71% identical to the *hFBXO25* transcript at the nucleotide level, showing high homology in the coding sequence and no significant homology in the 5' and 3' UTRs (Fig. III-10c). We amplified transcripts containing murine exon 9a by RT-PCR on total brain RNA and found several database entries confirming our results. Using RT-PCR and BLAST analyses, we did not recover sequences homologous to the putative 5' extensions of *hFBXO25* exons 2 and 5.

B.1.2. Expression analysis of *Fbxo25*

Similarly to the delineation of genomic organisation, study of the expression pattern of *Fbxo25* involved the investigation of human and mouse *Fbxo25* expression profiles.

B.1.2.1. Expression analysis of *hFBXO25*

hFBXO25 is expressed in the foetal and adult brain.

Northern hybridisation of MTN poly-A^+ RNA blots using an *hFBXO25* exons 1 – 7-specific probe, which was PCR-amplified from *hFBXO25* cDNA, showed expression of *hFBXO25* in all foetal tissues and all adult brain tissues investigated (Fig. III-11a – b). The *hFBXO25* transcript is present as a single ~2.1 kb band. The large number of *hFBXO25*-related ESTs deposited in the sequence databases also reflects its ubiquitous expression.

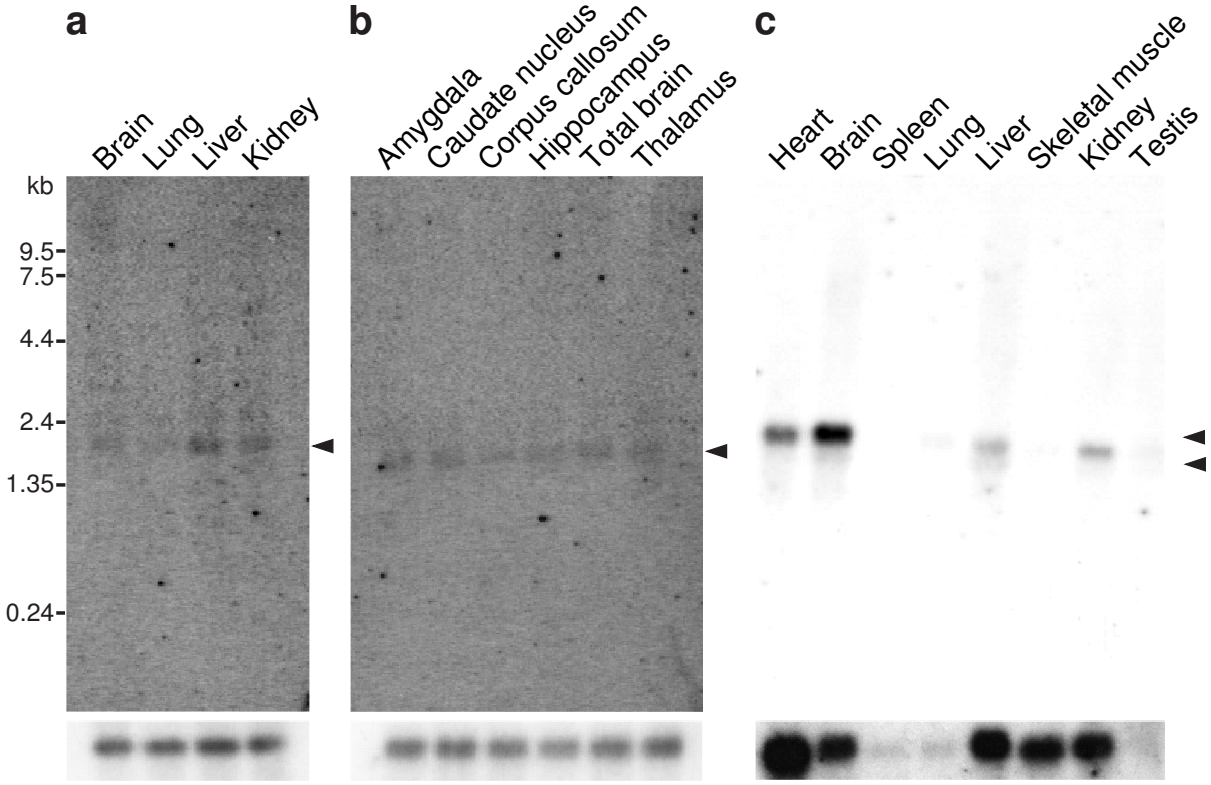


Fig. III-11 | **Expression analyses of *hFBXO25* and *mFbxo25*.**

a. Probing a human foetal multiple tissue northern (MTN) blot with *hFBXO25*, reveals a ~2.1 kb transcript (arrowhead). *β-actin* control, size marker on the left.
b. Hybridising a human brain MTN blot with the same probe as the one used in panel a results in a ~2.1 kb messenger RNA (arrowhead). *β-actin* control, size marker as in panel a.
c. Hybridising a mouse adult MTN blot with an *mFbxo25* probe, divulges transcripts of ~2.1 and ~1.9 kb (arrowheads). *mG3pdh* control, size marker as in panel a.

B.1.2.2. Expression analysis of *mFbxo25*

While *mFbxo25* transcript sizes were estimated using northern hybridisation, their spatial distribution was investigated by RNA ISH.

B.1.2.2.1. Northern hybridisation

mFbxo25 is expressed in adult brain as a ~2.1 kb transcript.

Northern hybridisation of an adult mouse MTN blot using an *mFbxo25* exons 6 – 10-specific probe, which was PCR-amplified from E14.5 mouse brain cDNA, showed a transcript of approximately 2.1 kb that is expressed strongly in brain. Furthermore, we observed expression in heart, liver and kidney. A 1.9 kb transcript is present in brain and liver (Fig. III-11c). Clearly, there is no expression in skeletal muscle, a finding that is confirmed by the absence of expressed sequences from this tissue in the ESTs composing the *mFbxo25* UniGene cluster (Mm.37950). No apparent expression of the *mFbxo25* transcript was detected in spleen, lung or testis. However, the *mG3pdh* control probe revealed only minimal amounts of RNA in these tissues (Fig. III-11c). Still, the expression level may indeed be low, as only 2.5% (3/119), 1.7% (2/119) and 0.84% (1/119) of all *mFbxo25* UniGene cluster-composing ESTs are derived from spleen, lung and testis, respectively.

B.1.2.2.2. RNA *in situ* hybridisation

Collaboration with D. Meunier and Dr. E. Minina, MPI-MG, Berlin, Germany and A. Visel, MPI-EE, Hannover, Germany.

mFbxo25 is expressed in E14.5 neuronal tissues and in the adult hippocampus, cerebral cortex and Purkinje cell layer.

Spatial expression of *mFbxo25* was studied employing RNA ISH. The *mFbxo25* exons 6 – 10 probe used in northern hybridisation was TA-cloned in pGEM-T Easy. The vector's T7 (anti-sense) and SP6 (sense) promoters were used in *in vitro* transcription reactions to label the probes with $^{33}\alpha$[P]UTP, $^{35}\alpha$[S]UTP or digoxigenin. Each of these differently labelled probes was applied to paraffin-embedded mouse tissue sections using protocols established in three different laboratories.

To decide which embryonic stage to employ in the ISH, semi-quantitative RT-PCR analyses were performed. Linear *mFbxo25* amplification was obtained with up to 27 cycles (Fig. III-12a). Using 25 cycles, the *18S rRNA* primer:competimer ratio at which *mFbxo25* and *18S rRNA* amplification were approximately equally strong was established on cDNA prepared from mouse adult brain total RNA (Fig. III-12b). Finally, *mFbxo25* expression levels in E12 (head and body), E14.5 and E18 (both stages, brain and body) were compared employing

a 1:40 *18S rRNA* primer:competimer ratio. Results from these analyses showed the highest level of expression in E14.5 brain and body (Fig. III-12c). Therefore ISH was performed on E14.5 sections.

On sagittal sections of E14.5 mouse embryos, expression was observed almost exclusively in neuronal tissues. Specific signals were detected in the medulla oblongata, cerebellar primordium, lateral wall of the midbrain, roof of the neopallial cortex (the future cerebral cortex), medial wall of the cerebral hemisphere, left trigeminal (V) ganglion and spinal cord (Fig. III-13a). We also found expression in the intestine. Hybridisations on coronal sections of E14.5 heads revealed specific signals in the medulla oblongata, trigeminal (V) and facial (VII) ganglia, Rathke's pouch (the future pituitary gland), neopallial cortex, posterior semicircular canal and crus commune of the labyrinth (Fig. III-13b), and in the olfactory epithelium and neural layer of the retina (Fig. III-13c). Expression in adult mouse brain was confined to the cerebral cortex, hippocampus and Purkinje cell layer (Fig. III-13d).

Identical results were obtained with all three RNA ISH protocols (Figs. III-13 and III-14). The *mFbxo25* sense probe never yielded a specific pattern.

Fig. III-12 | *Next page.* **Semi-quantitative reverse transcription polymerase chain reaction establishing *mFbxo25* expression levels.**

a. Plotting the normalised intensity of the *mFbxo25* amplicons, shown in the inset (size marker on the right), against the number of polymerase chain reaction (PCR) cycles, shows linear amplification up to 27 cycles. The *mFbxo25* expression level is established (panels b and c) using 25 cycles (dashed line).
b. Three different *18S ribosomal RNA* (*rRNA*) primer:competimer ratios (1:50, 1:40 and 1:35) are used to establish conditions resulting in approximately equal amplification of *mFbxo25* and *18S rRNA* (dashed line). The *mFbxo25* expression level is determined using a primer:competimer ratio of 1:40 (panel c). Intensity measurements were based on the multiplex PCR reactions shown in the inset. While the *mFbxo25* amplicon migrates as a 630 bp band, the *18S rRNA* amplicon migrates at 477 bp. Size marker on the left.
c. Using the conditions as established in panels a and b, complementary DNA (cDNA) preparations from embryonic day (E) 12 – 18 mice were employed to test relative *mFbxo25* expression levels. Expression is highest in E14.5 embryos. This is not due to degradation of the other cDNA samples; in fact, the rate of *18S rRNA* amplification shows that preparations from E14.5 embryos contain significantly less cDNA than those from the other stages (inset). The bottom panel illustrates the size-separated multiplex PCR reactions used in the expression analysis. Size marker on the left.

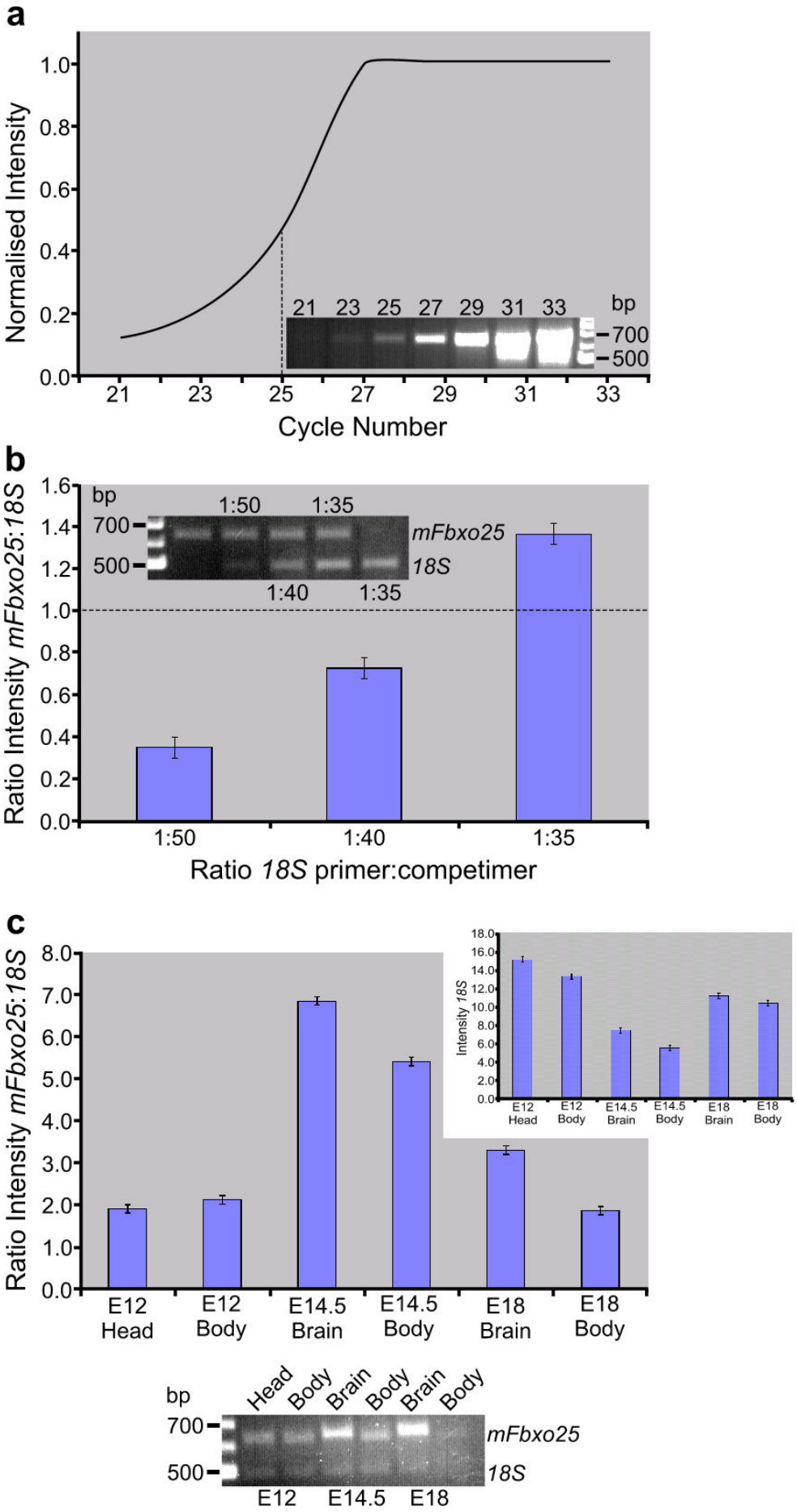

a
Normalised Intensity
1.0
0.8
0.6
0.4
0.2
0.0
21 23 25 27 29 31 33
bp
700
500
21 23 25 27 29 31 33
Cycle Number
b
Ratio Intensity mFbxo25:18S
bp
700
500
1:50 1:35
1:40 1:35
mFbxo25
18S
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1:50 1:40 1:35
Ratio 18S primer:competimer
c
Ratio Intensity mFbxo25:18S
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
E12 Head
E12 Body
E14.5 Brain
E14.5 Body
E18 Brain
E18 Body
Intensity 18S
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
E12 Head
E12 Body
E14.5 Brain
E14.5 Body
E18 Brain
E18 Body
bp
Head
Body
Brain
Body
Brain
Body
700
500
mFbxo25
18S
E12 E14.5 E18

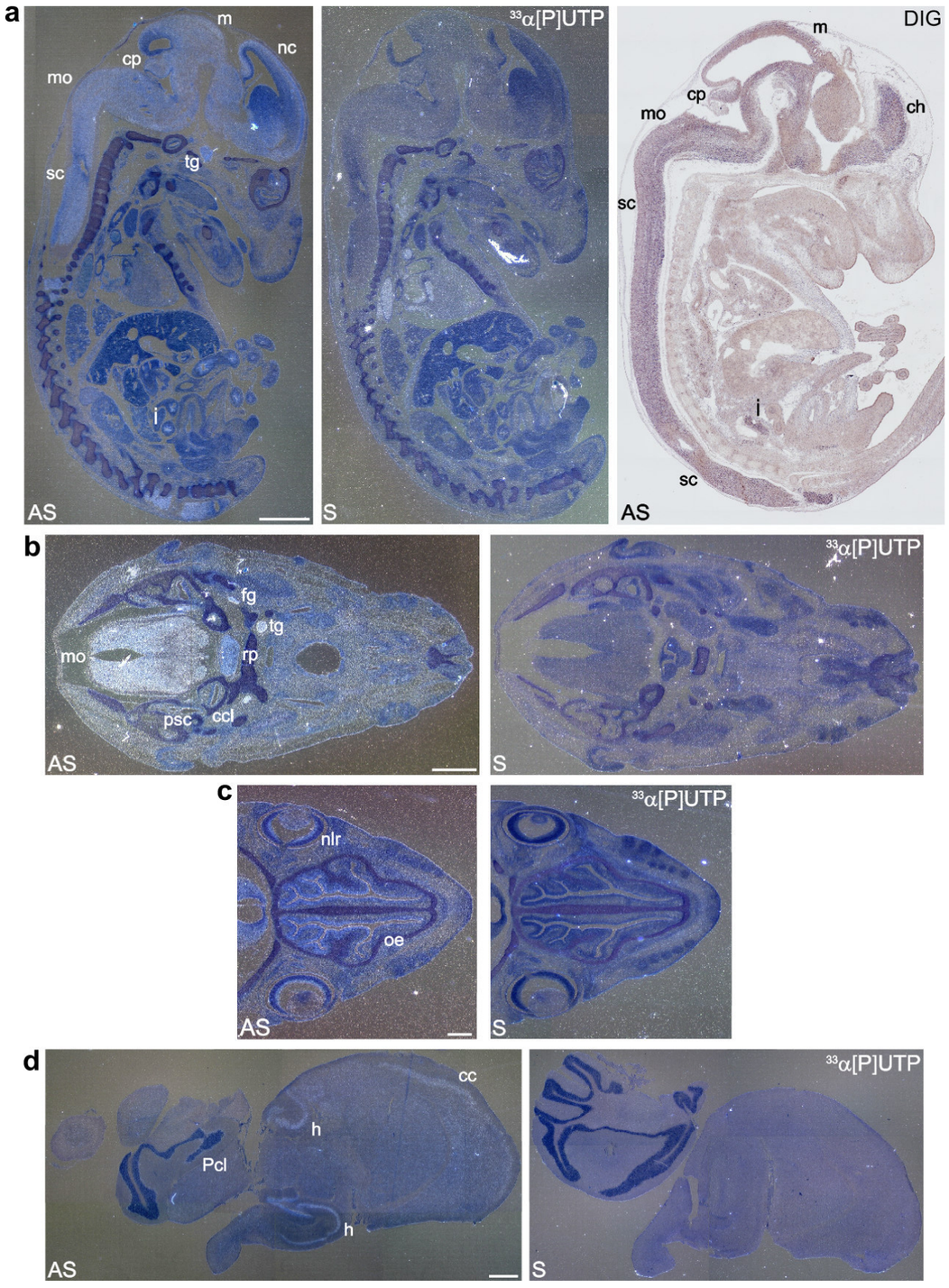
Fig. III-13 | *Next page, left.* **Expression of *mFbxo25* in tissues from embryonic day 14.5 and adult brain ($^{33}\alpha$[P]UTP- and digoxigenin-labelled probes).**

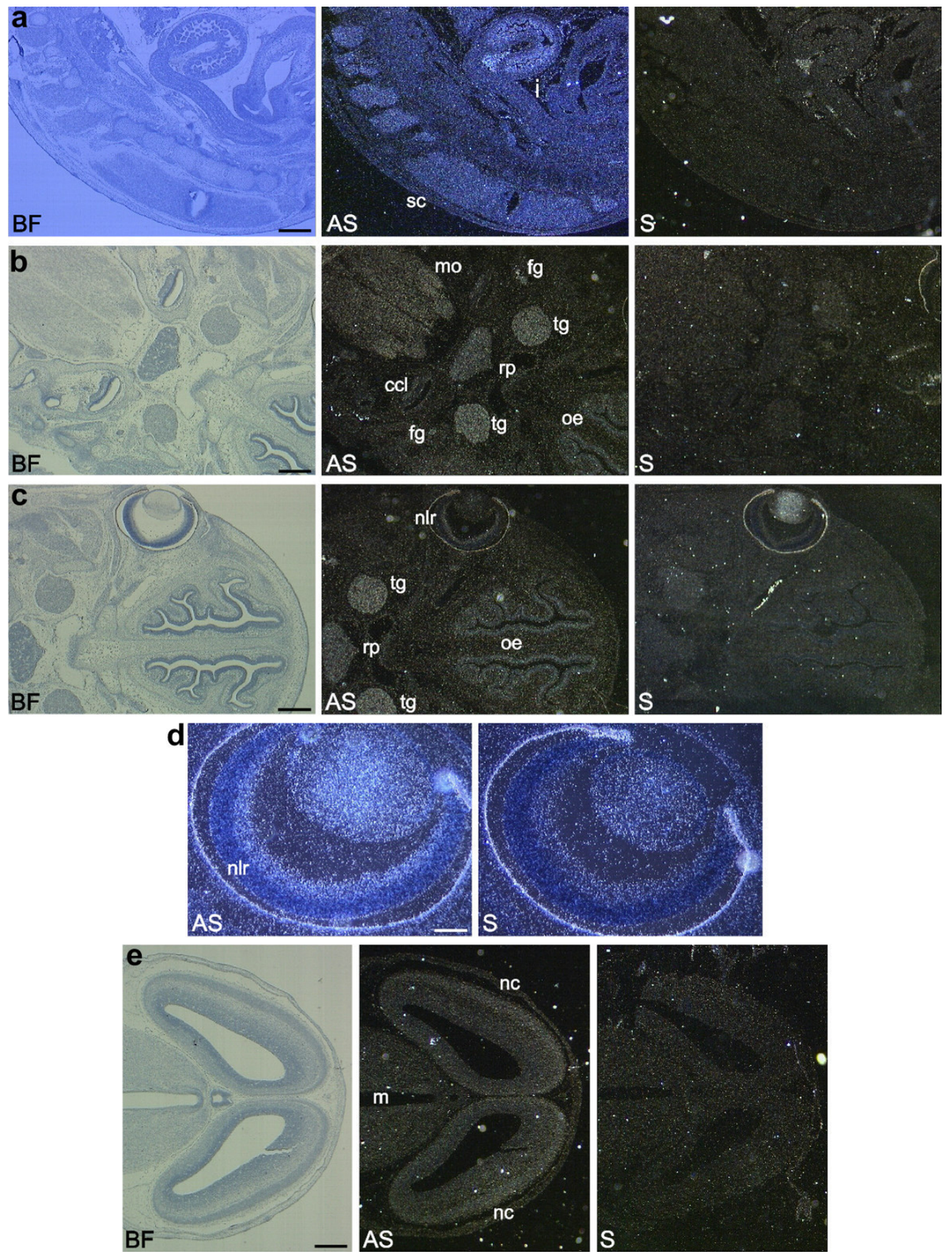
a. Hybridisation with radioactively ($^{33}\alpha$[P]UTP)- or digoxigenin (DIG)-labelled *mFbxo25* probes on sagittal sections of embryonic day (E) 14.5 mice reveals expression in the cerebellar primordium (cp), lateral wall of the midbrain (m), medulla oblongata (mo), roof of the neopallial cortex (nc), medial wall of the cerebral hemisphere (ch), spinal cord (sc), left trigeminal ganglion (tg) and intestine (i). Anti-sense (AS) and sense (S) probes are hybridised on serial sections. Images are composed of overlapping microscopic photographs. Anterior is up, dorsal to the left. Scale bar, ~1 mm.
b. On coronal sections of E14.5 heads, the $^{33}\alpha$[P]UTP anti-sense probe yields a signal in the medulla oblongata (mo), trigeminal (tg) and facial (fg) ganglia, Rathke's pouch (rp), posterior semicircular canal (psc) and crus commune of the labyrinth (ccl). Lateral is up, anterior to the right. Scale bar, ~0.25 mm.
c. Sections more dorsal from those depicted in panel b show expression in the olfactory epithelium (oe) and neural layer of the retina (nlr). Scale bar, ~0.5 mm.
d. As seen from sagittal sections of the adult mouse brain, expression is confined to the cerebral cortex (cc), hippocampus (h) and Purkinje cell layer (Pcl). Dorsal is up, anterior to the right. Scale bar, ~1 mm.

Image courtesy of A. Visel, MPI-EE, Hannover, Germany (a – right).

Fig. III-14 | *Next page, right.* **Expression of *mFbxo25* in tissues from embryonic day 14.5 ($^{35}\alpha$[S]UTP-labelled probe).**

Hybridisation with a $^{35}\alpha$[S]UTP-labelled *mFbxo25* probe on sections from embryonic day 14.5 mice essentially yields the same expression pattern as that shown in Fig. III-13. BF, bright field; other acronyms as in Fig. III-13. Anti-sense and sense probes were hybridised on serial sections. Unless otherwise noticed, scale bars ~0.25 mm.
a. Sagittal section showing the dorsal posterior third of the body. Anterior is up, dorsal to the left.
b. Coronal section showing the central structures of the head. Lateral is up, anterior to the right. Image rotated ~45° clockwise.
c. Coronal section showing the lateral anterior third of the head. Lateral is up, anterior to the right.
d. High magnification recording of the eye. Orientation as in panel c. Scale bar, ~100 μm.
e. Coronal section showing the dorsal half of the head. Orientation as in panel c.


a
m
nc
mo
cp
sc
tg
i
AS
$^{33}\alpha$[P]UTP
S
DIG
m
cp
mo
ch
sc
i
sc
AS
b
fg
tg
mo
rp
psc
ccl
AS
$^{33}\alpha$[P]UTP
S
c
nlr
oe
AS
$^{33}\alpha$[P]UTP
S
d
cc
h
Pcl
h
AS
$^{33}\alpha$[P]UTP
S


a
BF
i
sc
AS
S
b
BF
mo
fg
tg
rp
ccl
fg
tg
oe
AS
S
c
BF
nlr
tg
rp
oe
AS
tg
S
d
nlr
AS
S
e
BF
nc
m
nc
AS
S

B.2. Computational studies

In silico investigations focused on the search for hFBXO25 homologues and analysis of F-box sequences.

B.2.1. Identification of hFBXO25 homologues

hFBXO25 has several putative homologues among vertebrates.

Subtraction from the sequence databases combined with Ensembl predictions resulted in putative hFBXO25 homologues in several different organisms. An alignment of these homologues, including hFBXO25 ORF-1, is shown in Fig. III-15a. Putative *P. troglodytes* Fbxo25 was omitted from the alignment, as the available prediction was obviously erroneous. As hFBXO25 and hFBXO32 share 61.4% identity and 77.2% similarity, the protostomia and their subranks were evolutionarily too distant from *H. sapiens* to distinguish unambiguously between hFBXO25 and hFBXO32 homologues. For this reason, these sequences were also excluded from the analysis. However, Jin *et al.* claim that CG11658 and DY3.6 are the fly and worm orthologues, respectively, from hFBXO25/32[645]. Sequence identity among Fbxo25 homologues is remarkably high, ranging from 60.7% (*G. gallus* vs. *T. nigroviridis*) to 99.7% (*H. sapiens* and *M. fascicularis* vs. *M. mulatta*; Fig. III-15b). As *P. pygmaeus*' Fbxo25 evidently lacked 24 AA of N-terminal sequence (Fig. III-15a), it was not included in the sequence identity analysis.

B.2.2. Analysis of F-box sequences

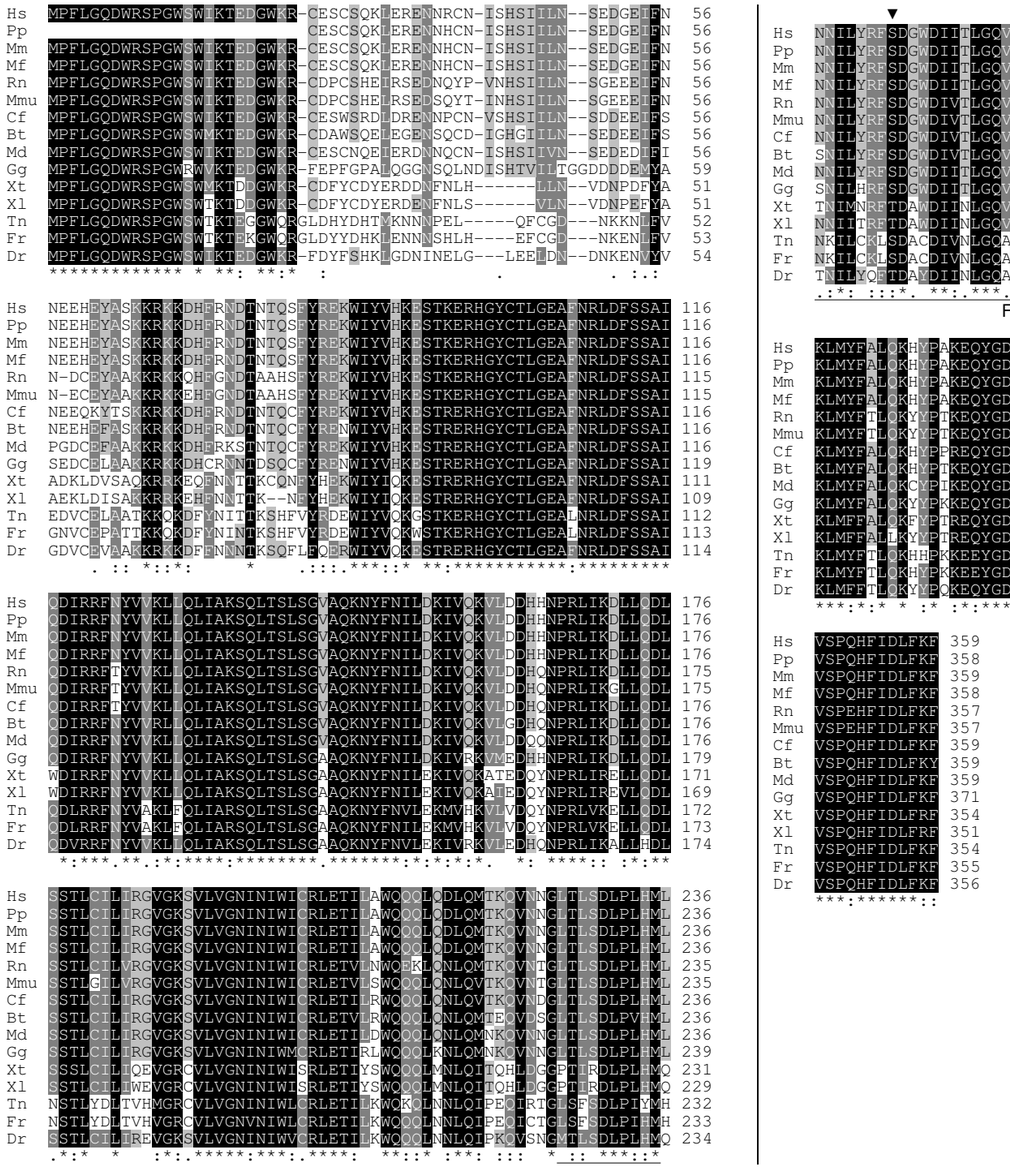
The hFBXO25 S244 residue, located in the F-box, is uncommon, but is not indicative of a new subgroup of F-boxes.

Bai *et al.* showed that even moderate mutations in conserved residues of the scCdc4 F-box abolish its interaction with Skp1. Indeed, P279A, I286A, L290A, W316A and LP278-279AA mutants failed to bind Skp1[610]. Similarly, Cenciarelli *et al.* published a

Fig. III-15 | *Next page.* **hFBXO25 vertebrate homologues.**

a. Cross-species global multiple sequence alignment of hFBXO25 homologues among vertebrates. The F-box is underlined and an arrowhead indicates residue S244 (see Fig. III-16). Shades of grey highlight similarities within subsets of sequences. An asterisk marks identities, a colon conserved substitutions and a dot semi-conserved substitutions. Hs, *H. sapiens*; Pp, *P. pygmaeus;* Mm, *M. mulatta*; Mf, *M. fascicularis*; Rn, *R. norvegicus*; Mmu, *M. musculus*; Cf, *C. familiaris*; Bt, *B. taurus*; Md, *M. domestica*; Gg, *G. gallus*; Xt, *X. tropicalis*; Xl, *X. laevis*; Tn, *T. nigroviridis*; Fr, *F. rubripes*; Dr, *D. rerio*.
b. Sequence identity matrix based on the alignment shown in panel a.

a



b

Sequence Identity Matrix for Fbxo25 Vertebrate Homologues

	Hs	Mm	Mf	Rn	Mmu	Cf	Bt	Md	Gg	Xt	Xl	Tn	Fr	Dr
Hs	100	99.7	99.4	88.3	88.3	91.9	90.2	91.9	81.4	70.1	70.7	62.7	65.5	76.3
Mm		100	99.7	88.3	88.3	91.9	90.2	91.9	81.4	70.1	70.7	62.7	65.8	76.3
Mf			100	88.5	88.5	91.6	89.9	91.6	81.4	69.9	70.9	62.9	66.0	76.5
Rn				100	96.3	86.6	85.2	86.3	77.0	68.2	69.5	63.2	64.9	75.4
Mmu					100	87.1	85.7	85.2	77.0	68.8	70.1	63.2	65.7	75.4
Cf						100	89.6	88.8	78.9	70.1	70.7	61.9	64.4	74.6
Bt							100	87.4	79.2	69.6	69.6	62.7	65.8	74.6
Md								100	81.6	71.0	70.1	63.8	65.5	76.3
Gg									100	67.9	68.4	60.7	62.9	74.6
Xt										100	93.7	61.6	62.7	72.8
Xl											100	61.7	62.9	71.6
Tn												100	92.6	69.7
Fr													100	72.2
Dr														100

reduced binding affinity between Skp1 and the L51A and W76A hFBXL3a F-box mutants[758]. scCdc4 L290 and W316 correspond to hFBXL3a L51 and W76, respectively (Fig. III-16a). All conserved residues investigated in both studies are present in the human and mouse Fbxo25 F-boxes, apart from the L290/L51 residue; a serine residue occupies this position in the Fbxo25 F-boxes (S244 in hFBXO25 and S243 in mFbxo25, Fig. III-16a).

Apart from muscle-specific rFbxo32, which was shown to be up-regulated during muscle atrophy[762], all other F-box domains from 20 FBPs ranging from human to yeast (hCYCLIN-F, hSEL-10, hFBXL1, hFBXL2, hFBXL3a, hFBXL4, hFBXL5, hFBXW1a, hFBXW2, hFBXO2, hFBXO4, hFBXO5, hFBXO7, mFbxl8, mFbxo8, mFbxo12, mFbxo15, mFbxo16, scCdc4 and scGrr1), which were shown experimentally to bind Skp1[609,610,758,762-764], contain a leucine residue at the position corresponding to hFBXO25 S244 (25-244) (Fig. III-16a). In addition, crystal structure studies of the F-box – Skp1 interaction in the Skp1 – hFBXL1[643] and hFBXW1a – Skp1 – β-catenin complexes[641] showed that this leucine residue directly contacts Skp1.

Within the F-box domain of hFBXO25 homologues, sequence identity is as high as 100% (Fig. III-16b). Although *P. troglodytes* and *P. pygmaeus* proteins were excluded from Fbxo25 sequence identity calculations, their F-boxes were included in the cross-species F-box comparison. Unexpectedly, *X. tropicalis*' and *T. nigroviridis*' F-boxes showed the lowest homology (56.0%; Fig. III-16b).

Alignment of 1740 non-redundant F-box sequences (Pfam PF00646) showed that non-polar residues, such as leucine, valine and methionine, occupy position 25-244 in 62.4%, 15.8% and 8.3% of F-boxes, whereas threonine, serine and tyrosine – characterised by hydroxyl groups in their side chains – were rare at the same position (1.5%, 1.0% and 0.2%, respectively). Out of 61 non-redundant human F-box sequences, only hFBXO25 and hFBXO32, which cluster together in a phylogenetic analysis of this set of sequences, contain a serine at 25-244 (Fig. III-16c). The fact that the human and murine Fbxo25/32 F-box domains are classified as false negatives when applying them to the Prosite F-box domain matrix (PS50181) further points towards their uniqueness.

a

```
                                                                                
scCdc4 (P07834)     272 RDLITSLPFEISL-KIFNYLQFE-DIINSL-GVSQNWNKIIRKSTS--------LWKKL-- 319
hFBXL3a (Q9UKT7)     34 -CDWGNLLQDIIL-QVFKYLPLL-DRAHAS-QVCRNWNQVFHMPD---------LWRCFEF 81
hFBXO4 (Q9UKT5)      56 ASTLTRLPIDVQL-YILSFLSPH-DLCQLG-STNHYWNETVRNPI---------LWRYF-- 102
hFBXO7 (Q9Y3I1)     329 VFGLVVLPLELKL-RIFRLLDVR-SVLSLS-AVCRDLFTASNDPL---------LWRFL-- 375
hFBXO25 (Q8TCJ0)    226 -LTLSDLPLHMLNNILYRFSDGW-DIITLG-QVTPTLYMLSEDRQ---------LWKKLCQ 274
mFbxo25 (Q9D2Y6)    225 -LTLSDLPLHMLNNILYRFSDGW-DIVTLG-QVTPTLYMLSEDRR---------LWKRLCQ 273
hFBXO32 (Q969P5)    223 -LTFTDLPLCLQLNIMQRLSDGR-DLVSLG-QAAPDLHVLSEDRL---------LWKKLCQ 271
mFbxo32 (Q9CPU7)    223 -LTITDLPVCLQLNIMQRLSDGR-DLVSLG-QAAPDLHVLSEDRL---------LWKRLCQ 271
                              *   :    :           .          .          **: :    

scGrr1 (P24814)     314 VFALNMLPSEILH-LILDKLNQKYDIVKFL-TVSKLWAEIIVKIL---------YYRPH-- 361
mFbxo8 (Q9QZN3)      68 FINLEMLPPELSF-TILSYLNAI-DLCLAS-CV---WQDLANDEL---------LWQGL-- 111
mFbxo12 (Q9QZN2)      6 ---EIHLPSVPMM-EILSYLDAY-SLLQAA-QVNKNWNELASSDV---------LWRKLCQ 51
mFbxo15 (Q9QZN0)     82 ---LDRMPSEILV-KILSYLDAV-TLVCIG-CVSRRFYHLADDNL---------IWVRKYA 127
mFbxo16 (Q9QZM9)     86 LDFTTKLPRVLSV-YIFSFLDPR-SLCRCA-QVSWYWKSLAELDQ---------LWMLK-- 132
mFbx18 (Q8CIG9)       2 GELVDNLPEEVLA-LIFRDLPLR-DLAVAT-RVCRAWAAAAANST---------VWSDK-- 48
hFBXO2 (Q9UK22)      43 AAYLDELPEPLLL-RVLAALPAA-ELVQACRLVCLRWKELVDGAP---------LWLLK-- 90
hFBXO5 (Q9UKT4)     250 ---LFRRGLRHVLATILAQLSDM-DLINVS-KVSTTWKKILEDDKG--------AFQLYS- 296
hFBXL1 (Q13309)      94 ------LPDELLL-GIFSCLCLP-ELLKVS-GVCKRWYRLASDES---------LW----- 140
hFBXL2 (Q9UKC9)       9 ------LPKELLL-RIFSFLDIV-TLCRCA-QISKAWNILALDGS---------NW----- 55
hFBXL4 (Q9UKA2)     277 NGYFDKLPYELIQ-LILNHLTLP-DLCRLA-QTCKLLSQHCCDPLQYIHLNLQPYWAKL-- 332
hFBXL5 (Q9UKA1)     202 STGITHLPPEVML-SIFSYLNPQ-ELCRCS-QVSMKWSQLTKTGS---------LWKHL-- 248
hFBXW1a (Q9Y297)    190 ----DHIAEN-----ILSYLDAK-SLCAAE-LVCKEWYRVTSDGM---------LWKKL-- 228
hFBXW2 (Q9UKT8)      54 RDFLKLLPLELSF-YLLKWLDPQ-TLLTCCL-VSKQWNKVISACT--------EVWQTA-- 101
hCYCLIN-F (P41002)   29 NLTILSLPEDVLF-HILKWLSVE-DILAVR-AVHSQLKDLVDNHAS--------VWACA-- 76
hFBXW7 (Q96LE0)     163 -DFISLLPKELAL-YVLSFLEPK-DLLQAA-QTCRYWRILAEDNL---------LWREKCK 208
```

b

Sequence Identity Matrix for Fbxo25 F-boxes

	Hs	Pt	Pp	Mm	Mf	Rn	Mmu	Cf	Bt	Md	Gg	Xt	Xl	Tn	Fr	Dr	S/T
Hs	100	100	100	100	100	94.0	94.0	92.0	90.0	100	96.0	70.0	72.0	60.0	64.0	76.0	S
Pt		100	100	100	100	94.0	94.0	92.0	90.0	100	96.0	70.0	72.0	60.0	64.0	76.0	S
Pp			100	100	100	94.0	94.0	92.0	90.0	100	96.0	70.0	72.0	60.0	64.0	76.0	S
Mm				100	100	94.0	94.0	92.0	90.0	100	96.0	70.0	72.0	60.0	64.0	76.0	S
Mf					100	94.0	94.0	92.0	90.0	100	96.0	70.0	72.0	60.0	64.0	76.0	S
Rn						100	100	94.0	88.0	94.0	90.0	66.0	68.0	60.0	64.0	72.0	S
Mmu							100	94.0	88.0	94.0	90.0	66.0	68.0	60.0	64.0	72.0	S
Cf								100	92.0	92.0	88.0	68.0	70.0	62.0	66.0	70.0	S
Bt									100	90.0	90.0	66.0	66.0	60.0	64.0	70.0	S
Md										100	96.0	70.0	72.0	60.0	64.0	76.0	S
Gg											100	70.0	70.0	58.0	62.0	74.0	S
Xt												100	94.0	56.0	60.0	72.0	T
Xl													100	58.0	62.0	70.0	T
Tn														100	96.0	62.0	S
Fr															100	66.0	S
Dr																100	T

c

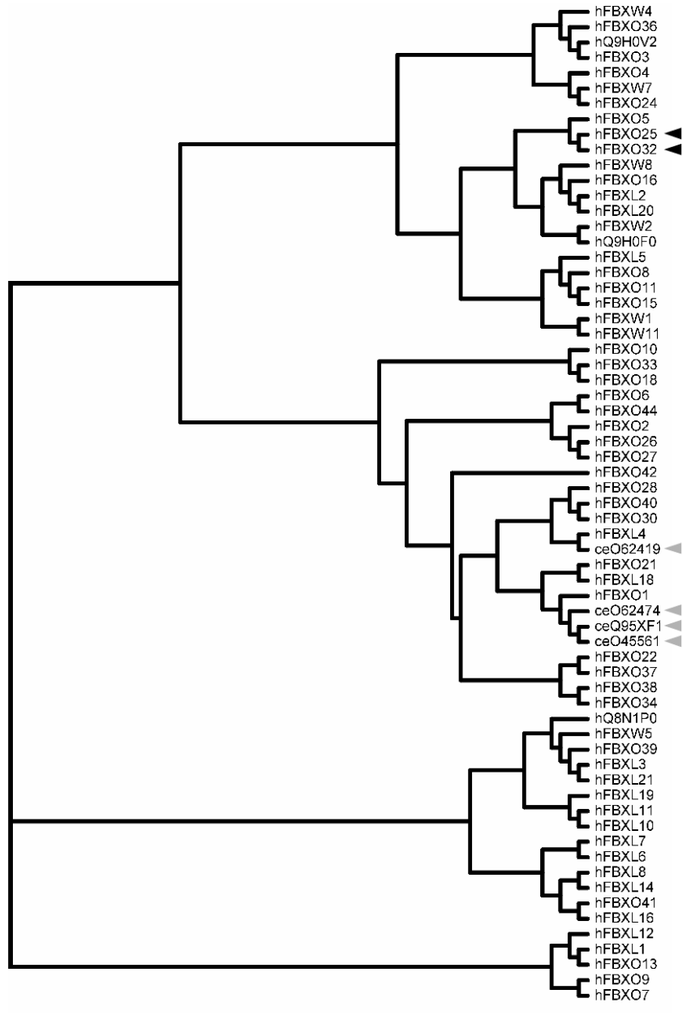


Fig. III-16 | **hFBXO25 S244 is unusual, but does not define a new subclass of F-boxes.**

a. Multiple alignment of F-box domains that interact with Skp1. Arrowheads indicate residues that abolish binding of the F-box with Skp1 when they are mutated, as was shown for scCdc4 (black) and hFBXL3a (grey). Bold print highlights the peculiar occurrence of a serine residue at 25-244 (see text for explanation) in human and mouse F-boxes 25 and 32. Swiss-Prot accession numbers are given in brackets. An asterisk marks identities, a colon conserved substitutions and a dot semi-conserved substitutions.

b. Sequence identity matrix for vertebrate Fbxo25 F-boxes. As can be appreciated from the last column, hFBXO25 residue S244 is well conserved among vertebrate homologues. Pt, *P. troglodytes*; other abbreviations as in Fig. III-15a.

c. Phylogenetic analysis of a non-redundant set of human F-boxes shows that hFBXO25 and hFBXO32, both containing a serine residue at position 25-244 (see text), cluster together (black arrowheads). However, four *C. elegans* F-boxes that also feature serines at the positions corresponding to 25-244 (grey arrowheads) do not cluster together with hFBXO25/32. Nomenclature is according to Jin *et al.*[645]. In case of uncharacterised proteins, the Swiss-Prot identifier is used. h, *H. sapiens*; ce, *C. elegans*.

Taken together, these data made us hypothesize that there might exist a subgroup of F-box domains that could be identified by a serine residue at 25-244. However, this hypothesis does not seem to hold true: the only four *C. elegans* F-box sequences carrying a serine residue neither cluster together in a phylogenetic analysis of 420 *C. elegans* F-boxes, nor do they cluster together with hFBXO25 and hFBXO32 in the human phylogenetic analysis (Fig. III-16c).

B.3. Protein studies

Protein analysis was initiated by cloning the hFBXO25 ORF in several mammalian expression vectors. Next, we studied the subcellular localisation of hFBXO25 and determined some of its interaction partners.

B.3.1. Cloning of the hFBXO25 open reading frame

hFBXO25 ORFs 1 and 2 and an F-box deletion construct were cloned into mammalian expression vectors.

ORFs 1 and 2 omitting the stop codon were amplified by RT-PCR on human fibroblast cDNA, introducing *Xho*I and *Sac*II restriction sites. Resulting PCR products were directionally cloned in the pEGFP-N3 and pEGFP-C1 mammalian expression vectors in order to generate C-terminal and N-terminal EGFP fusion proteins, respectively. Subsequently, *Nhe*I/*Bam*HI fragments from the EGFP-N3 fusion constructs were directionally subcloned into the pcDNA4/V5-HisB mammalian expression vector in order to construct V5-tagged ORFs 1 and 2. Apart from tagged versions of ORFs 1 and 2, we also designed a V5-tagged F-box deletion construct, hFBXO25-ORF1ΔF-V5. Essentially, this construct contains ORF 1, but lacks AA 226-274, which encode the F-box domain. It was PCR-amplified in two amplicons flanking the sequence encoding AA 226-274. The amplicon 5' of the F-box-encoding-sequence was amplified from the hFBXO25-ORF1-V5 clone, introducing *Xho*I and *Eco*RI restriction sites. The amplicon 3' of the F-box encoding sequence was amplified from the same clone including *Eco*RI and *Xba*I sites. Both amplicons were digested with the appropriate endonucleases and ligated together before being directionally cloned between the *Xho*I/*Xba*I sites of the pcDNA4/V5-HisB vector. For reasons of consistency, hFBXO4 and hFBXO7 ORFs[758], which served as positive controls in co-IP experiments (see III.B.3.3.1), were subcloned from pcDNA3 into pcDNA4/V5-HisB. Employing PCR technology, a *Not*I site was introduced at the 3' end of the PCR product, and hFBXO4 and hFBXO7 amplicons were directionally cloned between the *Bam*HI/*Not*I and *EcoR*I/*Not*I sites of the pcDNA4/V5-HisB vector, respectively.

B.3.2. Subcellular localisation of over-expressed hFBXO25

hFBXO25 localises predominantly, but not exclusively, to the nucleus, but not to the nucleoli.

In an attempt to gain insight into the function of the hFBXO25 protein, we established its subcellular localisation. We transfected N- and C-terminal EGFP-tagged versions of ORF-1 to make sure that the observed subcellular localisation was not an effect of the position of the tag. Similarly, we generated C-terminal V5-tagged ORF-1 to investigate a putative (steric) interference of the EGFP-tag. The same set of constructs was generated for hFBXO25 ORF-2. In order to monitor possible cell specific localisation, we transfected all constructs in five different cell lines. HeLa and COS-7 cells were chosen as straightforward experimental systems. Because of the neuronal expression of *mFbxo25* in E14.5 mice (see III.B.1.2.2.2), we decided to transfect all constructs into Neuro-2A, U373 MG and SH-SY5Y cell lines as well. These experiments revealed predominant, but not exclusive, localisation to the nucleus, a distribution which has also been reported for other FBPs[764,765]. The nucleoli, primarily involved in the biogenesis of ribosomes[766], seem devoid of hFBXO25. We did not detect any difference in subcellular localisation of either of the ORFs, nor was there a noticeable interference of the EGFP-tag or a cell-specific effect (Fig. III-17).

B.3.3. Investigation of hFBXO25 interactions

hFBXO25 functional studies involved co-IP and ligase activity assays, and *in vitro* mutagenesis experiments.

B.3.3.1. Co-immunoprecipitation experiments

hFBXO25 is part of the $SCF^{hFBXO25}$ complex.

It has been shown repeatedly that FBPs interact through their F-box domains with Skp1 and – indirectly – with Cul1. To verify the biological relevance of the predicted hFBXO25 F-box, we investigated its putative interactions with Skp1 and Cul1. To this end, we used an α-V5 antibody to immunoprecipitate over-expressed V5-tagged hFBXO25 ORF1 (hFBXO25-ORF1-V5; predicted size, 48 kDa) and over-expressed V5-tagged hFBXO25 ORF1 lacking the F-box domain (hFBXO25-ORF1ΔF-V5; predicted size, 40 kDa) from HeLa total cell lysates. As positive controls, we over-expressed hFBXO4 and hFBXO7, which have been shown to bind Cul1 and Skp1[758]. Immunoprecipitates were size-separated on gradient SDS-PAGE gels and immunoblotted with antibodies against Skp1 and Cul1,

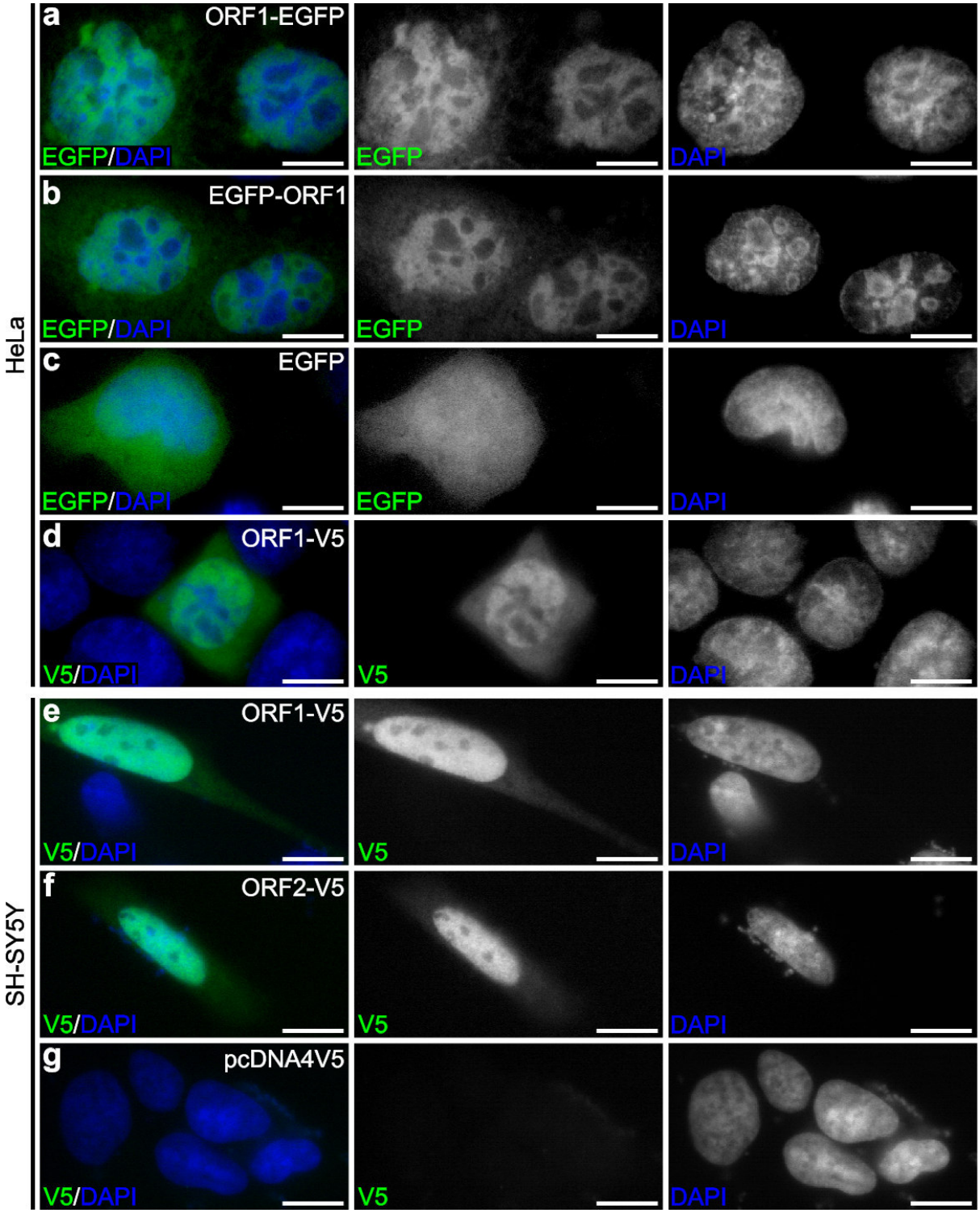


Fig. III-17 | **Subcellular localisation of hFBXO25 open reading frames 1 and 2.**

Indicated hFBXO25 open reading frames (ORFs) and expression vectors are transiently expressed in HeLa (panels a – d) and SH-SY5Y (panels e – g) cells. Cells are recorded using direct fluorescence (enhanced green fluorescent protein, EGFP) or indirect immunofluorescence (V5). DNA is counter-stained with 4'6-diamidino-2-phenylindole·2 HCl (DAPI). No difference in subcellular localisation is observed between N- and C-terminally tagged constructs, EGFP- and V5-tagged clones, transfection in HeLa and SH-SY5Y cells, and transfection of ORFs 1 and 2. Scale bars, 5 μm.

showing binding between the hFBXO25 F-box domain and endogenous Skp1 and Cul1 (Fig. III-18a), thereby establishing hFBXO25 as a *de facto* FBP. Using the same procedure, we also showed that Roc1 is a member of the $SCF^{hFBXO25}$ E3 ligase complex (Fig. III-18a). The hFBXO25 – Skp1/Cul1/Roc1 interactions are reduced compared to those of the hFBXO4 and hFBXO7 controls with Skp1/Cul1/Roc1. Over-expressed hFBXO25-ORF1-V5 and hFBXO25-ORF1ΔF-V5 consistently ran as multiple-band patterns, and both hFBXO4 and hFBXO7 also showed additional bands (Fig. III-18a). Immunoblotting of over-expressed constructs immediately after cell lysis resulted in essentially the same pattern, strongly reducing the possibility of non-specific protein degradation due to prolonged incubations inherent to the immunoprecipitation protocol. We therefore assume that these bands represent differently modified FBP species.

B.3.3.2. E3 ligase activity assay

The $SCF^{hFBXO25}$ complex is a functional E3 ligase.

To investigate whether the $SCF^{hFBXO25}$ complex is a functional E3 ligase or whether the reduced binding between hFBXO25 and Skp1/Cul1/Roc1 abolishes $SCF^{hFBXO25}$ E3 ubiquitin ligase activity, we performed immunoprecipitation experiments in HeLa cells with over-expressed hFBXO25-ORF1-V5 and subsequent α-Ubiquitin immunoblotting. As a positive control, we over-expressed V5-tagged hFBXO4. We observed a high MW smear, indicative of ubiquitinylated proteins, which was not present in the vector control. The increase in intensity of the high MW smear upon LLnL-induced proteasome blocking is comparable to that observed in the positive control (Fig. III-18b). Taken together, these results suggest a functional $SCF^{hFBXO25}$ E3 ubiquitin ligase complex.

B.3.3.3. *In vitro* mutagenesis

hFBXO25's unusual serine at position 244 is crucial to the hFBXO25 – Skp1 binding.

To estimate the importance of the atypical S244 residue in the hFBXO25 F-box, we mutated S244 into a much more prevalent leucine residue. The hFBXO25-S244L-V5 construct, containing the S244L exchange in its F-box, was created from the hFBXO25-ORF1-V5 plasmid using *in vitro* mutagenesis methodology. We performed co-IP experiments in HeLa cells with over-expressed hFBXO25-S244L-V5. Interestingly, mutation of S244 abolished the hFBXO25 – Skp1 interaction, showing the significance of S244 (Fig. III-19a). The weak residual Cul1 – hFBXO25-S244L-V5 interaction (Fig. III-19a) is likely due to their common binding partner Roc1, a central member of the E3 complex[628].

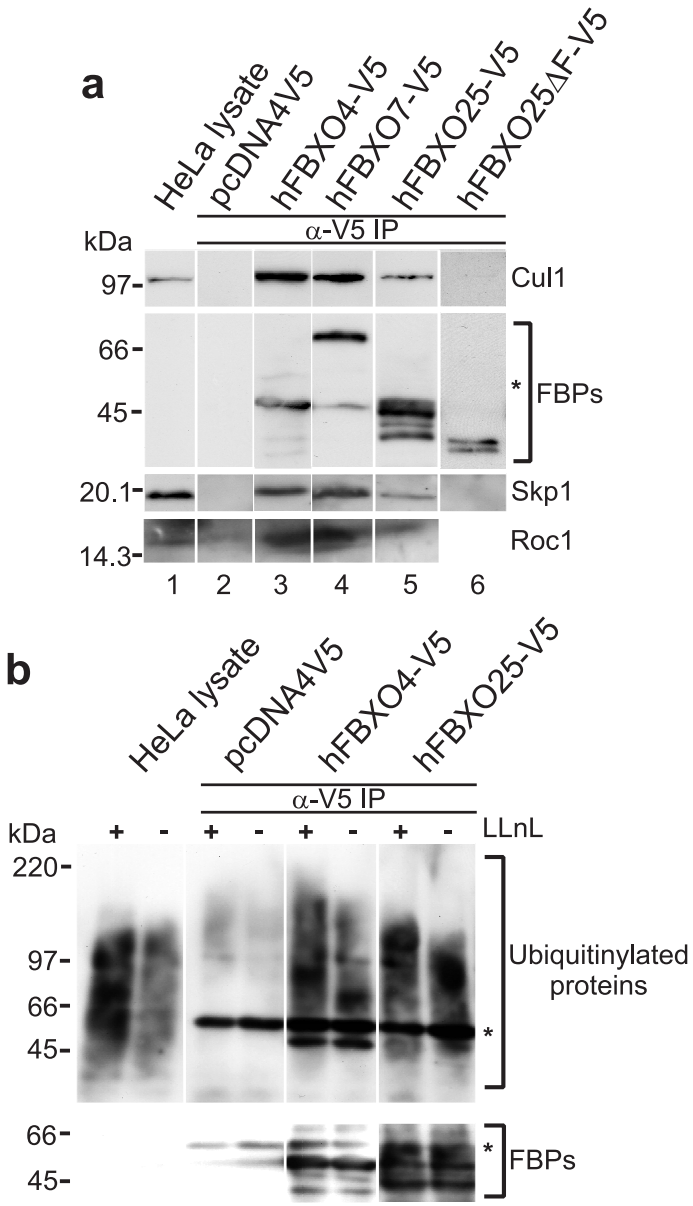


Fig. III-18 | **hFBXO25 is a *de facto* F-box protein.**

a. HeLa cells are transfected with vector (lane 2) or plasmids encoding V5-tagged proteins (lanes 3 – 6). Over-expressed proteins are immunoprecipitated (IP) with α-V5 antibody, and immunoblotted with α-Cul1, α-Skp1, α-Roc1 and α-V5 antibodies. These experiments show that hFBXO25 interacts with Cul1, Skp1 and Roc1. The hFBXO25 F-box mediates the interaction with Skp1. Binding between hFBXO25ΔF-V5 and Roc1 was not assessed. The different bands in lanes 3 – 6 possibly represent differently modified protein species (see text). The asterisk indicates the position of the antibody heavy chain, which becomes visible upon extended exposure. FBPs, F-box proteins; size marker on the left.

b. Same procedure as in panel a, but immunoblotting with α-Ubiquitin antibody and incubation of cells with (+) or without (-) the proteasome inhibitor N-Acetyl-Leu-Leu-Norleucinal (LLnL) two hours prior to lysis. The result suggests a functional $SCF^{hFBXO25}$ Ubiquitin ligase complex. Asterisks indicate the antibody heavy chain. Size marker on the left.

To test the effect of a lack of binding between hFBXO25-S244L-V5 and Skp1 on the $SCF^{hFBXO25}$ Ubiquitin ligase activity, we performed a ligase assay as described for WT hFBXO25-V5 using the F-box mutant. The difference in intensity of Ubiquitin-positive

smear between cells over-expressing WT hFBXO25-V5 and mutant hFBXO25-S244L-V5 suggests that the absence of binding to Skp1 negatively affects the ligase activity, highlighting the importance of S244 (Fig. III-19b). The high MW smear observed in the hFBXO25-S244L-V5 samples presumably consists of substrate molecules ubiquitinylated by endogenous hFBXO25 and/or other SCF complexes. Such substrate molecules occur in soluble form or are bound to the FBPs. As hFBXO25-S244L-V5, which is not impaired in its ability to bind substrate, is overly abundant, the equilibrium between the two substrate states will tend towards the bound form, resulting in their pull-down and subsequent detection. As can be appreciated from Fig. III-19b, blocking of the proteasome in cells over-expressing hFBXO25-S244L-V5 does not have the obvious effect observed with over-expressed WT FBPs. This is likely due to the fact that the effect of excess functional FBP molecules in the cell is no longer present. Instead, only the actions of endogenous cell components are assayed. In comparison to over-expression, these endogenous components are too few to make a noticeable difference between LLnL-treated and untreated cells using our assay.

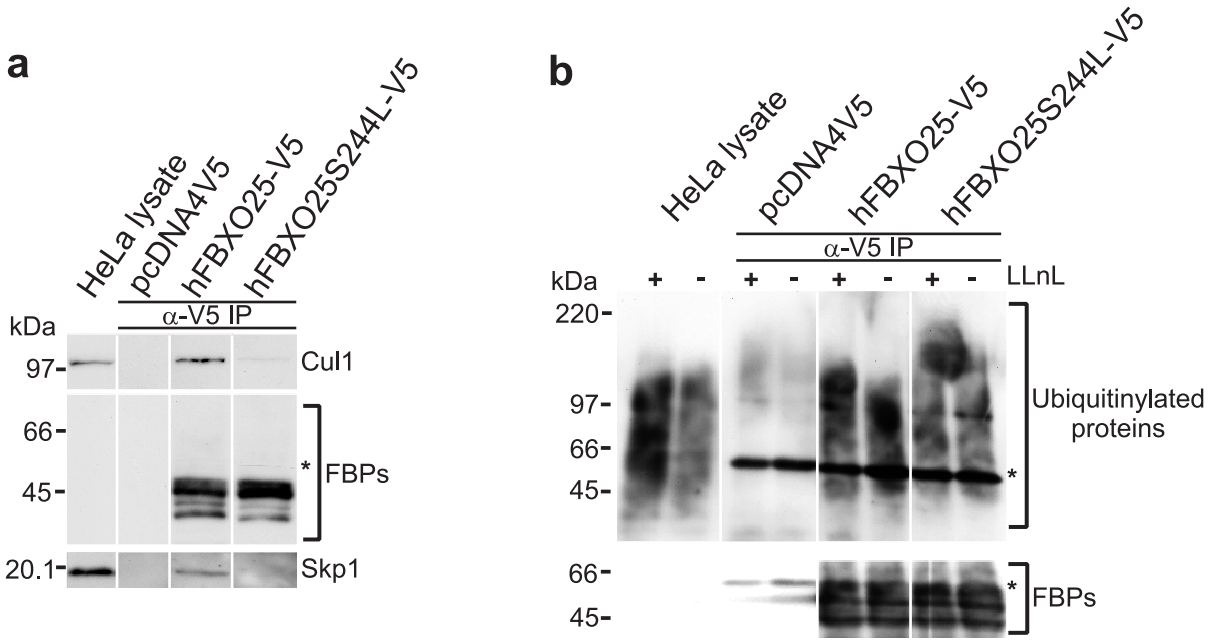


Fig. III-19 | **The serine at position 244 in the hFBXO25 F-box is crucial for its interaction with Skp1.**

a. hFBXO25 binds Skp1 and Cul1, but the S244L mutant F-box does not bind Skp1 anymore, thereby also abolishing the hFBXO25 – Cul1 interaction. The residual Cul1 – hFBXO25-S244L-V5 interaction is probably due to their common binding partner Roc1. Experimental procedure as in FigIII-18a. The asterisk indicates the position of the antibody heavy chain, which becomes visible upon extended exposure. FBPs, F-box proteins; size marker on the left.
b. Comparison of the hFBXO25-V5 and hFBXO25S244L-V5 lanes shows the importance of S244 in the $SCF^{hFBXO25}$ ligase activity. See text for a further interpretation of the ubiquitinylation pattern. Experimental procedure as in FigIII-18b. Asterisks indicate the antibody heavy chain; size marker on the left.

C. Molecular and computational characterisation of Kiaa1202

Apart from the gene disrupted by the autosomal BP, we also characterised the X-chromosomal gene affected by the translocation. We investigated the *hKIAA1202* gene, its transcript and its protein product. In addition, we performed computational studies on the protein sequence.

C.1. Nucleic acid studies

Characterisation of *hKIAA1202* involved establishing the intron – exon structure of the gene, the organisation of its transcripts and its pattern of expression. The obtained information was useful in screening *hKIAA1202* for mutations.

C.1.1. Genomic organisation of *Kiaa1202*

Establishing the structure of the *hKIAA1202* gene involved bioinformatic approaches, study and alignment of known ESTs and mRNAs, and RT-PCR and RACE experiments. We also established the genomic organisation of *mKiaa1202*, hKIAA1202's mouse homologue.

C.1.1.1. Genomic organisation of *hKIAA1202*

hKIAA1202, which encodes ORFs of 1498, 1453 and 1382 AA, is organised into 12 exons, at least 3 of which are alternatively spliced. Exon scrambling may occur in *hKIAA1202* transcripts.

A *hKIAA1202* cDNA was first identified in a high-throughput screen for uncharacterised human genes coding for large (> 50 kDa) proteins *in vitro*. A 6029 bp *hKIAA1202* cDNA clone, including a 15 bp poly-A tail, was recovered from a large-insert adult brain cDNA library[757]. A second mRNA, termed *SHAP-A* and derived from human heart, represents an alternatively spliced *hKIAAA1202* transcript. The first SHAP-A exon (Sh.1) is located between the first and second *hKIAA1202* exon. Alignment of these cDNAs with a genomic contig compiled from the BAC clones RP11-554P16, RP11-88O18 and RP11-119E20 resulted in a first crude genomic map of the *hKIAA1202* locus, and revealed eleven exons spanning 222 kb of genomic sequence on Xp11.22.

Next, we performed an extensive computational analysis to find additional expressed sequences and to refine the *hKIAA1202* genomic map. The compiled genomic *hKIAA1202* contig was used to extract corresponding mouse sequences from the genome databases at the NCBI and EMBL. These mouse sequences were in turn compiled into an *mKiaa1202* genomic contig and both contigs were subsequently submitted to the NIX server, resulting in a multitude of exon predictions. Next, the contigs were compared using PIP-maker, which identified regions of conservation between human and mouse *Kiaa1202*, and hence suggested

possible expressed sequences. In addition, the *hKIAA1202* contig was screened for ORFs ≥ 150 AA after conceptual translation of the strand corresponding to the hKIAA1202 protein. Finally, available databases of expressed sequences were screened for human non-spliced ESTs mapping to the *hKIAA1202* genomic contig.

From this combined computational information, putative expressed regions were considered according to the following criteria:

- an EST, or
- a ≥ 150 AA ORF, or
- human – mouse conservation over 150 bp with at least 75% identity, or
- exon prediction by at least two algorithms and human-mouse conservation over 100 bp with at least 75% identity, or
- an ORF ≥ 100 AA if within 1 kb of an exon or any of the above features.

Twenty-four regions complying with at least one of these selection criteria were found (IDs A – X), and 91% (20/22) of those examined by RT-PCR on total fibroblast and foetal brain RNAs turned out to be expressed (Table III-3). The same strategy also led to the identification of three putative extensions of known exons (IDs Ext.1 – Ext.3). Expression of two such extensions was experimentally tested and one of them was expressed (Table III-4). Finally, analysis of six spliced ESTs and mRNAs (IDs Sp.1 – Sp.6) also mapping to the *hKIAA1202* locus indicated the existence of another two uncharacterised exons and one putative extension of a known exon. Expression of these spliced ESTs and mRNAs was not assessed experimentally (Table III-5).

Table III-3 | **Putatively expressed sequences within the *hKIAA1202* gene**

ID	Alignment[§]	Criteria	Expressed?
		RP11-554P16	
A	83630 – 83148	227 AA ORF	Yes
B	82468 – 81896	BF922581; 81%, 135 bp; 250 AA ORF; human pred. GM	Yes
C	74463 – 74139	157 AA ORF	Yes
D	71521 – 71393	83%, 133 bp; human pred. FG; mouse pred. FG	Yes
E	69720 – 69570	77%, 194 bp	Yes
F	69117 – 68896	86%, 254 bp; mouse pred. MZEF, FGs	Yes
G	50088 – 49767	153 AA ORF	Yes
H	42484 – 42149	AW945473	Yes

Table III-3 | **Putatively expressed sequences within the *hKIAA1202* gene**

ID	Alignment[§]	Criteria	Expressed?
I	36894 – 36603	83% – 93%, 424 bp; 103 AA ORF	Yes
J	25868 – 25600	157 AA ORF	Yes
		RP11-88O18	
K	10408 – 10277	85%, 148 bp; mouse pred. MZEF	No
L	9714 – 9464	87%, 257 bp	Yes
M	8650 – 8409	88%, 244 bp	Yes
		RP11-119E20	
N	13544 – 13673	BF362446; 77%, 98 bp; mouse pred. GM	Yes
O	24513 – 24744	167 AA ORF	Yes
P	28159 – 28676	230 AA ORF	Yes (only fibroblast)
Q	34337 – 34497	83%, 174 bp	Yes
R	45496 – 45914	250 AA ORF	Yes (only fibroblast)
S	46942 – 47238	147 AA ORF	Yes
T	59515 – 60003	200 AA and 153 AA ORFs	No
U	79451 – 79564; 79651 – 80045; 80056 – 80861	BF724336; BF914967; BF724335; 78%, 115 bp ; human pred. FGs; mouse pred. MZEF	Yes (79451 – 79564); Yes (79651 – 80045, only fibroblast); No (80056 – 80861)
V	80862 – 80958	110 AA ORF	Yes
W	105134 – 105780	AA515902; AI005420; AI381269; AW972020; BQ016983	ND
X	105876 – 106528	BF107195	ND

[§] When expression was evaluated, coordinates of putative exons are based on the primers used in the RT-PCR reaction; when expression was not assayed, coordinates are based on alignments with the entities mentioned under 'criteria'.

Table III-4 | **Evidence for 5' and 3' extensions of *hKIAA1202* exons**

ID	Alignment[§]	Criteria	Expressed?	Putative 5' extension[†]
Ext.1	RP11-554P16, 84482 – 84229 (primers) and 84502 – 84079 (entire region)	87%, 159 bp; human pred. GF, GS, FG; mouse pred. GF, FG	No	Exon 1
Ext.2	RP11-119E20, 96145 – 96389 (primers) and 96094 – 96494 (entire region)	133 AA ORF	Yes	Exon 7
Ext.3	RP11-119E20, 106820 – 107451 (entire region)	AA147750; AA147516	ND	Exon10

[§] Coordinates of putative extensions of known exons are based on the primers used in the RT-PCR reaction ('primers') and on alignments with the entities mentioned under 'Criteria' ('entire region'). The latter are an indication for the degree of overlap with the exon.

[†] Putative extensions that are shaded in grey are considered to be indeed expressed in the context of a *hKIAA1202* transcript. See text for details.

Table III-5 | **Overview of spliced ESTs and mRNAs mapping to the *hKIAA1202* locus**

ID	EST/mRNA	Alignment	Transcript organisation[§]
Sp.1	BF932432	RP11-554P16, 83979 – 83954; RP11-119E20, 3159 – 3313, 60788 – 60889	Exons 1-2-3-...
Sp.2	6848591H1	RP11-554P16, 84054 – 83951, 74748 – 74682; RP11-119E20, 3159 – 3313, 60788 – 60924, 63428 – 63541	Exons 1-Sp.2-2-3-4-...
Sp.3	AW001038	RP11-554P16, 84008 – 83951; RP11-119E20, 3159 – 3312, 11915 – 12059	Exons 1-2-Sp.3-...
Sp.4	5968254H1	RP11-554P16, 84166 – 83951; RP11-119E20, 3159 – 3312, 63428 – 63615	Exons 1-2-4-...
Sp.5	BF881299	RP11-119E20, 65921 – 65606, 91072 – 90912, 71485 – 71420	Exons 4-5-6-...
Sp.6	BU194023	RP11-119E20, 103263 – 103400, 106952 – 107672	Exons 9-10-Sp.6

[§] Sp.2, putative additional exon between exons 1 and 2; Sp.3, putative additional exon between exons 2 and 3; Sp.6, putative 3' extension of exon 10. Shaded putative exons are considered to be indeed expressed in the context of a *hKIAA1202* transcript. See text for details.

We performed inter-exon RT-PCR reactions to:

- verify the presence of *hKIAA1202* exons in tissue,
- experimentally define the intron – exon boundaries,
- try to identify different *hKIAA1202* splice variants, and
- try to link putative uncharacterised exons to *hKIAA1202*.

We designed primers specific to each of the exons present in the *hKIAA1202/SHAP-A* cDNAs and to putative exons with IDs W, Sp.2 and Sp.3, as they were either covered by several unspliced ESTs (W, Table III-3) or present in spliced ESTs (Sp.2 and Sp.3, Table III-5). RT-PCR reactions were carried out on total fibroblast and foetal brain RNAs between exons 1 and Sp.2, 1 and 2, 1 and Sp.3, 1 and 3, 1 and 4, Sp.2 and 2, Sp.2 and Sp.3, Sp.2 and 3, 2 and Sp.3, 2 and 3, 2 and 4, Sp.3 and 3, Sp.3 and 4, Sh.1 and 3, 3 and 4, 3 and 5, 3 and 6, 4 and 5, 4 and 6, 4 and 7, 5 and 6, 5 and 7, 5 and 8, 6 and 7, 6 and 8, 6 and 9, 7 and 8, 7 and 9, 7 and W, 8 and 9, 8 and W, 8 and 10, 9 and W, 9 and 10, and W and 10.

In addition, we performed 5' and 3' RACE on (putative) exons 1, Sp.2, L, M, 2 and X; 5' RACE on (putative) exons R, 3 and 6; and 3' RACE on (putative) exons Sp.3, S and 5. All reactions were carried out on total fibroblast and foetal brain RNA.

Our combined computational and experimental efforts resulted in the identification of five transcripts.

Isoform I, corresponding to the *hKIAA1202* cDNA, was recovered from foetal brain and fibroblast RNA, and codes for a 1498 AA ORF (Figs. III-20b – c).

Isoform II, coding for a 1453 AA ORF and lacking exon 3, was amplified from foetal brain and fibroblast RNA (Fig. III-20b). This transcript was also present in a private EST collection derived from human brain (Sp.4, Table III-5; courtesy of Dr. B. Hinzmann, metaGen, Berlin, Germany).

Isoform III, containing an additional exon 2a localised between exons 2 and Sh.1, was only recovered from fibroblast RNA. Exon 2a is also present on a kidney-derived EST (Sp.3, Table III-5). Introduction of exon 2a shifts the supposed translational start of the ORF (1382 AA) to exon 3 (Fig. III-20b).

Isoform IV corresponds to *SHAP-A* and is transcribed from the alternatively spliced exon Sh.1, localised between exons 2a and 3 (Fig. III-20b). As *SHAP-A* is transcribed from the *hKIAA1202* locus, we will from now on refer to exon Sh.1 as exon 2b. We did not experimentally verify the 5' end of exon 2b, but it may well be complete at its 5' end since the deposited *SHAP-A* mRNA sequence was derived from clones of a cDNA-library and cloned 5'-RACE products.

Finally, isoform V, only recovered from fibroblast RNA, is characterised through a 73 bp 5' extension of exon 1. We assume that the rest of the transcript is as in the *hKIAA1202* cDNA.

We did not find evidence for expression of the putative exon Sp.2, present in a spliced EST from a human kidney library (Table III-5; courtesy of Dr. B. Hinzmann, metaGen, Berlin, Germany). Although we amplified sequence from region X on several occasions, we were unable to link it to any of the known *hKIAA1202* exons, and therefore did not consider it any further. An overview of the recovered intron – exon boundaries is given in Table III-6.

Recovery of full-length clones of these proposed new isoforms by screening a Human Foetal Brain cDNA library (RZPD Lib. No. 564) and both the Human Large cDNA Collection I (derived from skeletal muscle and spinal cord) and II (derived from bone marrow, adipose tissues and foetal skin) (RZPD Lib. Nos. 451 and 313, respectively) was unsuccessful, presumably due to the low expression of *hKIAA1202* and to the considerable size of its transcripts. A graphical overview of *hKIAA1202*'s genomic organisation and its transcripts found in fibroblasts and foetal brain is shown in Figs. III-20a – b.

Table III-6 | ***hKIAA1202*** **exons' splice site sequences, genomic coordinates and sizes**

Exon	Splice acceptor site[§]	Splice donor site[§]	Exon coordinates[†]	Exon size (bp)	Intron size (kb)
1[‡]		AGTGTCTAAGgtaagaactg	~84166 – 83951	at least 216	118.0
1[¿]		AGTGTCTAAGgtaagaactg	~84093 – 83951	at least 143	118.0
2	tgcaacacagATTGAAGATG	TTGTCAGGAGgtaaggtaca	3161 – 3312	152	8.6
2a	ggcccttcacAGACTAAGTT	TTAAGCCAACaaacatggta	11915 – 12059	145	43.0
2b (Sh.1)	ND	AACTTCCATGgtaggtacag	~55415 – 55790	at least 377	5.0
3	cttgctgtagGAGGAACGCC	GCAACACAAGgtgagtctag	60790 – 60924	135	2.5
4	tgtcttacagTGACGTGTGT	GACAGTGCAGgtgaggactt	63430 – 65920	2491	5.5
5	ttctttacagGAATTTCCTG	AGTCAGGGAGgtaagtgagc	71423 – 71484	62	19.4
6	tgttttgcagGGAAATGGCT	AATCCGCCAAgtaagtacat	90914 – 91717	804/816[¶]	4.6
7	gtttccacagACAAGAGTTT	TCAAAAAAAGgtaaagtttt	96285 – 96465	181	4.1
8	attcccctagATACAGCTTA	CCAGGAGAAGgtagagttgg	100563 – 100832	270	1.3
9	gcttccccagTTGGTACTGA	CCGTGGCTAGgtaagcaagt	102134 – 103400	1267	3.4
10	cttatcccagGCATATTCTT		106954 – 107451	490/494/498/518[¶]	

[§] Conserved ag and gt dinucleotides at the splice acceptor and splice donor sites, respectively, are underscored.
[†] Coordinates for exon 1 on clone RP11-554P16, for exons 2 – 10 on clone RP11-119E20.
[‡] Exon 1 from isoform V.
[¿] Exon 1 from isoforms I – IV.
[¶] Depending on the haplotype (see III.C.1.3.3).

In the process of determining the genomic organisation of the *hKIAA1202* gene, we amplified several 'scrambled' *hKIAA1202* transcripts, both in cultured fibroblast cells and foetal brain. In such mRNAs, exons lack internal sequence, are split or incomplete, or the order of the exons is shuffled[767,768]. The ORF in these messages is often abolished. We recovered aberrant mRNAs performing RT-PCR and RACE. The amplification products were assayed by direct sequencing and are listed in Table III-7. Although positions of BPs within exons proved to be very variable, there are also indications that breakage may not be entirely random, as we found identical BP positions in different transcripts. Moreover, on at least one occasion, we amplified an identical scrambled transcript from fibroblasts using RT-PCR and 3' RACE (Table III-7; note that the RT-PCR-amplified product has 89 bp of additional exon 8 sequence due to primer design). Using 5' and 3' RACE methodology, we also amplified several transcripts with 5' and 3' diversions from the exon boundaries. As these diversions occurred at the end of the amplicon, we could not distinguish between true novel exon boundaries and artifactual cDNA synthesis. Therefore, we will not specify these results any further.

Fig. III-20 | *Next page.* **Organisation of the *Kiaa1202* gene, its transcripts and its protein product.**

a. The *hKIAA1202* gene spans 222 kb of genomic DNA on the short arm of the X chromosome. The gene is depicted from centromere (Cen.) to telomere (Tel.). A jagged arrow indicates the breakpoint. Black lines represent intronic sequences. Intron 1, 118 kb, is only partially shown (double slash). Grey rectangles denote exons (resolution permitting).
b. The diagram depicts four *hKIAA1202* transcripts (I – IV). A fifth isoform (V, see text) is not shown. Rectangles symbolise exons, which are numbered above isoform I, hooked arrows open reading frames (ORFs). ORF IV, as presented in the sequence databases, is likely to be incomplete at the 3' end (dashed arrow). White rectangles represent untranslated regions (UTRs), grey ones depict coding sequence, and diagonal stripes depict repetitive sequence. Introns are not drawn to scale. While filled arrowheads indicate local patient-specific sequence exchanges, open arrowheads denote sequence alterations found in both patients and controls. Numbering of sequence variants corresponds to Table IV-2.
c. hKIAA1202 corresponding to ORF I. Dashed lines represent exon boundaries. Grey rectangles symbolise domains [PSD-95/Dlg/ZO-1, PDZ (PS50106) and APX Shroom domain 2, ASD2]. Open arrowheads represent binding sites (BS) (Enabled/Vasodilator-stimulated phosphoprotein homology domain 1 BS, EVH1-BS and PDZ-BS), black arrowheads sequence motifs [bipartite nuclear localisation signal (NLS, PS00015) and leucine zipper (LZ, PS00029)] and the grey arrowhead the α-hKIAA1202 antigenic site. 'C' and 'QE' highlight cysteine (PS 50311)- and glutamine/glutamic acid (PS50322/PS50313)-rich regions, respectively. Grey lines indicate the positions of clones 1 – 6, which contain parts of the ORF that were used in the yeast two-hybrid screen (Y2H, see III.C.3.4.3).
d. Genomic structure of the *mKiaa1202* gene, drawn from centromere to telomere. Introns are not drawn to scale. Numbers within or underneath the exons indicate the percentage identity between the human and mouse exonic sequences. UTRs are not conserved. See text for additional information on exons 1, 2b and 9.

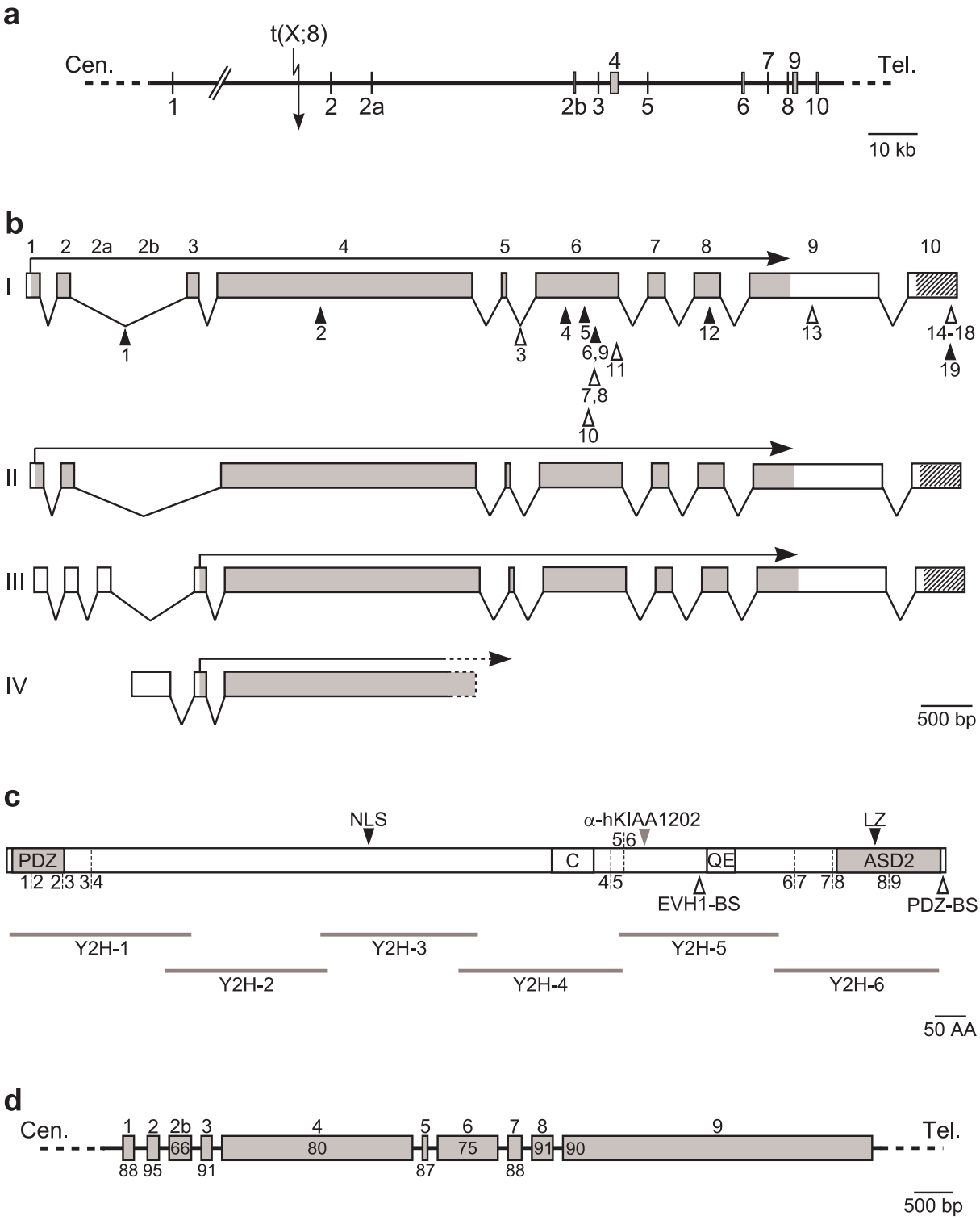

a
t(X;8)
Cen.
Tel.
1
2
2a
2b
3
4
5
6
7
8
9
10
10 kb
b
I
II
III
IV
500 bp
c
NLS
α-hKIAA1202
LZ
PDZ
C
QE
ASD2
EVH1-BS
PDZ-BS
Y2H-1
Y2H-2
Y2H-3
Y2H-4
Y2H-5
Y2H-6
50 AA
d
Cen.
Tel.
500 bp

Table III-7 | **Scrambled *hKIAA1202* transcripts recovered from cultured primary fibroblasts and foetal brain by RT-PCR and RACE**

Scrambled transcript[§]	Experimental methodology
Primary fibroblasts	
Exons 3 – 5' 4 – 3' 4 – 5	RT-PCR
Exons 3' 4 – 5 – 179 bp of 4 – 3' 6 – 7	RT-PCR
Exons 4 – 3' 6 – 7	RT-PCR
Exons 4 – 5' 5 – 3' 7	RT-PCR
Exons 156 bp of 4 – 30 bp of 1	5' RACE from exon 1
Exons 5 – 5' 6 – 31 bp of 9 – 3' 7 – 8	RT-PCR
Exons 5 – 5' 6 – 3' 7 – 8	RT-PCR
Exons 5' 5 – 3' 7 – 8	RT-PCR
Exons 5' 5 – 272 bp of put. X – 3' 8	RT-PCR
Exons 5 – 5' 6 – 118 bp of 9 – 3' 7 – 8	RT-PCR
Exons 5 – 5' 6 – 118 bp of 9 – 3' 7	3' RACE from exon 5
Exons 5 – 5' 6 – 28 bp of 5	3' RACE from exon 5
Exons 68 bp of 6 – 3' 6 – 7 – 8	RT-PCR
Foetal brain	
Exons 3 – 5' 4 – 3' 4 – 5 – 6	RT-PCR
Exons 3 – 5' 4 – 3' 4 – 5	RT-PCR
Exons 56 bp of 4 – 3' 4 – 5	RT-PCR
Exons 5' 6 – 3' 6 – 7 – 8 – 9	RT-PCR

[§] Identical BP positions within an exon are highlighted with the same shade of grey. Intact exon '*x*' boundaries are noted as 5' *x* or 3' *x*. For exons lacking either boundary, the size of the internal sequence is specified. Putative exons are indicated as 'put.' (Table III-3).

C.1.1.2. Genomic organisation of *mKiaa1202*

mKIAA1202 is organised into 10 exons, at least 2 of which are alternatively spliced.

Since initial investigations on *mKiaa1202*, the mouse homologue of *hKIAA1202*, were limited, we determined the gene's intron – exon structure to a degree accurate enough to serve as a template for primer design; for this, a purely computational analysis was sufficient. First, we established a genomic contig containing the *mKiaa1202* locus by tiling BAC clones RP24-391O14 (BX511313) and RP23-313P3 (AL671501). Next, we aligned ESTs and mRNAs showing sequence similarity to *hKIAA1202* to this contig and found that *mKiaa1202* encompasses 237 kb on XA1.1. Two expressed sequences were particularly informative. First, an EST recovered from dendritic cells (BY187865) contained exons 1 and 2, whereby exon 1 corresponded to its human counterpart from isoform V. Second, a recently published *mKiaa1202* cDNA clone (AK173129) derived from adult mouse pancreatic islets represents the mouse variant of *SHAP-A*, starting with exon 2b[769]. Together with the Ensembl prediction

for an *mKiaa1202* transcript (ENSMUST00000073888) and the alignment of *hKIAA1202* to the mouse genomic contig, we established the intron – exon boundaries, listed in Table III-8. We were not able to accurately set the 5' boundaries of exons 1 and 2b, as we could not be sure whether the ends of the expressed sequences under study represented true 5' boundaries or were artefacts of the cDNA synthesis. It is also unclear whether the *mKiaa1202* cDNA clone still contains intronic sequence in the 3' UTR or whether it is an accurate representation of a transcript. Splicing of the 3' UTR in the *hKIAA1202* cDNA clone may be an indication that the former is the case, but a lack of sequence similarity between human and mouse in the UTR makes it difficult to draw a conclusion. A pictorial overview of *mKiaa1202*'s genomic organisation is given in Fig. III-20d, including the percentage identity between human and mouse exonic sequence.

C.1.2. Expression analysis of *Kiaa1202*

To get a first insight into the physiological role of *Kiaa1202*, we determined its expression in several human and mouse foetal and adult tissues.

C.1.2.1. Expression analysis of *hKIAA1202*

hKIAA1202 is expressed in foetal and adult brain.

To gain insight into the physiological role of *hKIAA1202*, its expression pattern was investigated in human foetal and adult tissues. Conventional northern hybridisation with a *hKIAA1202* exon 4-specific probe, PCR-amplified from the *hKIAA1202* cDNA, yielded a 9.0 kb mRNA in all human adult tissues that were investigated, and an additional 5.0 kb transcript expressed in heart, lung and skeletal muscle (Fig. III-21a). Upon extended exposure, the 5.0 kb transcript also appeared in all other tissues apart from brain and pancreas. In line with these observations, the probe only detected the 9.0 kb mRNA in all human adult brain regions that were investigated; expression was strongest in the medulla and weakest in the cerebral cortex and spinal cord (Fig. III-21b). Similarly, only the 9.0 kb transcript was found in human foetal brain. Although not apparent from Fig. III-21c, the 5.0 kb mRNA was observed in the other investigated foetal tissues upon extended exposure.

Expression in fibroblast (Fig. III-21d) and lymphoblastoid cells (Fig. III-21e) was tested by RT-PCR. PCR conditions were optimised so that all primer pairs reliably amplified PCR products of the expected size when using human foetal brain RNA as a template (Fig. III-21e, bottom panel). To avoid spurious amplification from genomic DNA, only intron-spanning

primers were employed. Moreover, amplification was obtained when reverse transcriptase had been added to the RT reaction, but not without the addition of the enzyme. This expression study showed that *hKIAA1202* is expressed in cultured fibroblasts (Fig. III-21d). In lymphoblastoid cell lines, no amplification occurred with primers located in exons 1, 2, 2a and 2b, showing the lack of expression of *hKIAA1202* isoforms I – IV. Interestingly, we were able to amplify specific products with primers located in exons 3, 4, 5, 6, 7 and 9, indicating the presence of as yet unidentified isoforms. No difference was observed between four controls and the patient cell line (Fig. III-21e, top and middle panels).

C.1.2.2. Expression analysis of mouse and zebrafish *Kiaa1202*

mKiaa1202 and *drKiaa1202* are expressed in adult tissues. ISHs were unsuccessful.

Expression of *mKiaa1202* in adult tissues was assayed by northern hybridisation with a *mKiaa1202*-specific probe on a MTN blot. The probe was PCR-amplified from mouse adult brain cDNA. As shown in Fig. III-21f, *mKiaa1202* is expressed in all tissues under study. In contrast to *hKIAA1202* expression, we found transcripts of 9.5, 4.5, 2.0 and 1.4 kb in the mouse brain. Not all of these transcripts are present in the other investigated tissues (Fig. III-21f).

To get spatial information on *Kiaa1202*'s expression, considerable effort was put into *mKiaa1202* ISH on paraffin-embedded sections. To decide which embryonic stage to employ, semi-quantitative RT-PCR analyses were performed as described for *mFbxo25* (see III.B.1.2.2.2). Thirty cycles resulted in linear *mKiaa1202* amplification and an *18S* primer:competimer ratio of 1:45 was used. Although ~10-fold reduced in comparison to *mFbxo25* amplification, *mKiaa1202* amplification was best on E14.5 brain cDNA. Therefore, ISH was performed on E14.5 sections.

Attempts were made with two probes, PCR-amplified from mouse adult brain cDNA. At least one of them yielded a clear pattern of transcripts in northern hybridisations (Figs. III-21a – c); the other one was not tested. Three established protocols from three different laboratories (Dr. R. Fundele, MPI-MG, Berlin, Germany; Dr. A Vortkamp, MPI-MG, Berlin, Germany and Prof. Dr. G. Eichele, MPI-EE, Hannover, Germany), each employing a different labelling technique ($^{35}\alpha$[S]UTP, $^{33}\alpha$[P]UTP and digoxygenin, respectively), were tested under guidance of or in collaboration with an experienced lab member (courtesy of D. Meunier, Dr. E. Minina and A. Visel, respectively). In each case *mFbxo25* ISH was performed in parallel.

Table III-8 | ***mKiaa1202*** **exons' splice site sequences, genomic coordinates and sizes**

Exon	Splice acceptor site[§]	Splice donor site[§]	Alignment[†]	Exon size (bp)	Intron size (kb)
1[‡]		GGTGTCGAAGgtaaggagcg	~19791 – 20016	at least 226	120.5
1[¿]		GGTGTCGAAGgtaaggagcg	~19873 – 20016	at least 144	120.5
2	tgcaacacagATTGAAGATG	TTGTCAGGAGgtaaggcaca	140548 – 140699	152	56.4
2b	ND	AACTTCCATGgtaggtacag	~20089 – 20379	at least 291	3.2
3	cttgccatagGAGAAACACC	GCAACACTAGgtgagtctgg	23612 – 23746	135	2.2
4	tgtcttacagTGACGTGAGT	AGCCATACAGgtgaggactt	25975 – 28438	2464	5.3
5	ccctttgcagGAATTTCCTG	ATTCTGGGAGgtaagtgact	33739 – 33800	62	32.6
6	tgttttgcagGGAAATGGCT	TGACCCCAAAgtaagtgtat	66443 – 67219	777	4.6
7	gtttccacagACAAGAGTTT	TCAAAAAAAGgtaaggtttt	71884 – 72064	181	2.9
8	actcttttagATTCAGCTCA	CCAGGAGAAGgtacagatga	75006 – 75275	270	1.0
9	gcttccccagTTGGTGCTGA		76231 – 80231	4001	
10	See text				

[§] Conserved ag and gt dinucleotides at the splice acceptor and splice donor sites, respectively, are underscored.
[†] Coordinates for exons 1 and 2 on clone RP24-391O14, for exons 2b – 9 on clone RP23-313P3.
[‡] Exon 1 corresponding to *hKIAA1202* isoform V.
[¿] Exon 1 corresponding to *hKIAA1202* isoforms I – IV.

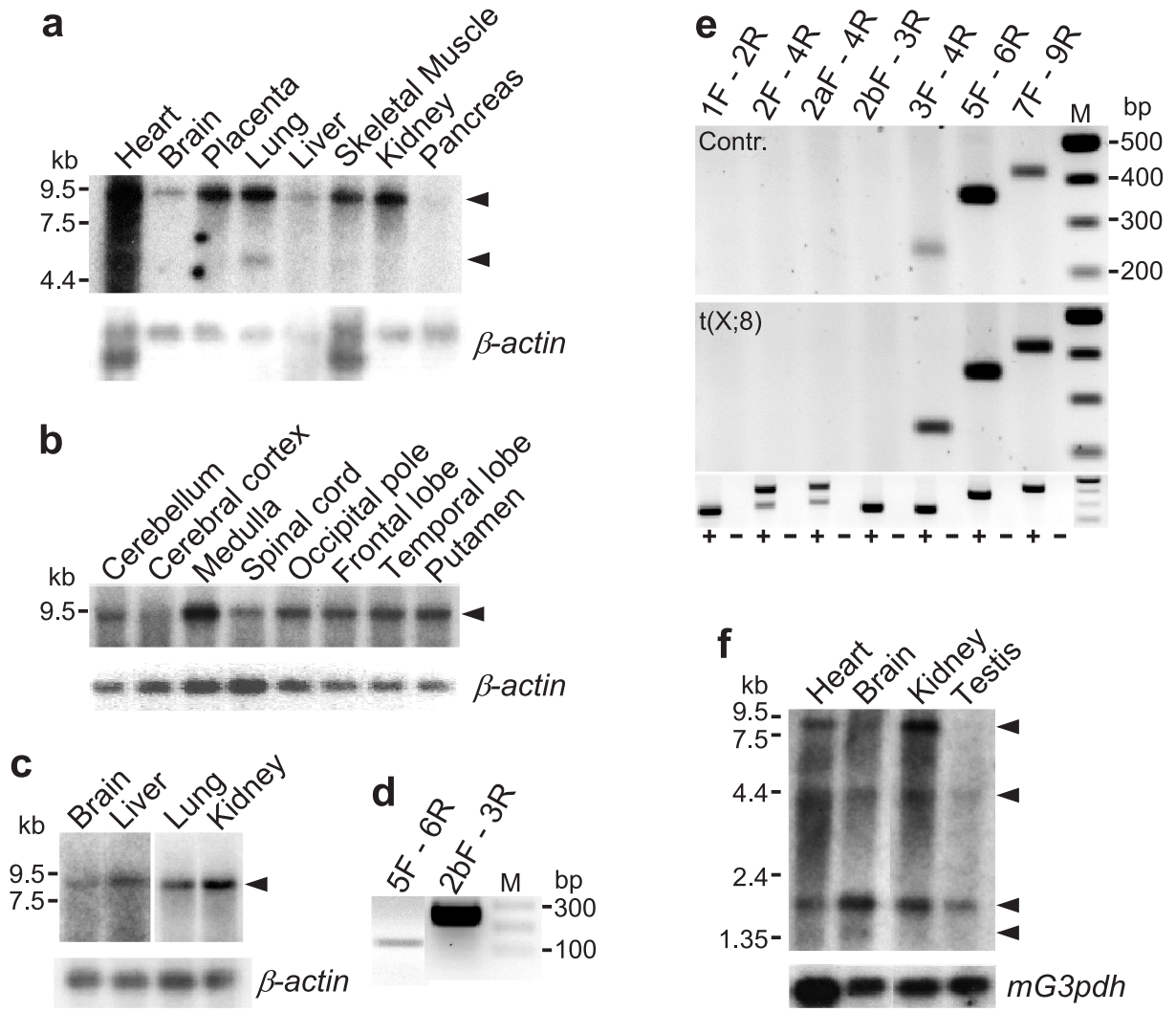


Fig. III-21 | **Expression analyses of human and mouse *Kiaa1202*.**

a. Northern blot hybridisation with a *hKIAA1202* exon 4-specific probe shows that the *hKIAA1202* gene is ubiquitously expressed with a 9.0 kb transcript present in all tissues examined, and a 5.0 kb mRNA expressed in heart, lung and skeletal muscle (arrowheads). *β-actin* control, size marker on the left.

b. Within the brain, the *hKIAA1202* probe recognises a single 9.0 kb transcript (arrowhead) in all tissues studied by northern blot hybridisation. *β-actin* control, size marker on the left.

c. Upon hybridisation of a human foetal brain multiple tissue northern blot with the *hKIAA1202* probe, the 9.0 kb transcript becomes apparent (arrowhead). The brain and liver lanes are exposed 25% longer than the lung and kidney lanes. *β-actin* control, size marker on the left.

d. Reverse transcription polymerase chain reaction (RT-PCR) on total RNA assays expression in primary human fibroblasts. The picture shows size-separated amplicons from two different experiments obtained with intron-spanning primers specific to *hKIAA1202* exons as indicated at the top. Size marker (M) on the right.

e. RT-PCR experiments are conducted on total RNA from control 46,XX (Contr., top panel) and 46,X,t(X;8)(p11.2;p22.3) [t(X;8), middle panel] lymphoblastoid cell lines, and from foetal brain (bottom panel). PCR is performed with intron-spanning primers specific to *hKIAA1202* exons as indicated at the top. Primers specific to exons 1, 2, 2a and 2b do not obtain amplification in lymphoblastoid cells. Specific amplicons are recovered for exons 3, 4, 5, 6, 7 and 9. These data are representative for four different 46,XX control cell lines. Using foetal brain RNA, amplification works with all primer pairs when reverse transcriptase was added to the transcription reaction (+), but not in the absence of the enzyme (-). The size difference between both amplicons in the 2F – 4R and 2aF – 4R reactions corresponds to the alternatively spliced exon 3 (Fig. III-20b). Size markers (M) on the right.

f. Northern hybridisation with a *mKiaa1202* 3'-specific probe reveals expression in all tissues. Transcripts of 9.5, 4.5, 2.0 and 1.4 kb (arrowheads) are identified. *mG3pdh* control, size marker on the left.

Even though the *mFbxo25* ISH worked well in all cases, a specific *mKiaa1202* signal was never obtained. As stage-specific RT-PCR experiments revealed that *mKiaa1202* expression from E8.5 – E10.5 embryos is likely to be below the threshold of detection by ISH and does not change significantly between stages E11.5 and E16.5[199], no ISH was undertaken on embryos from stages other than E14.5.

We also attempted to study *drKiaa1202* expression by stage-specific whole mount ISH (collaboration with S. Rohr, MDC, Berlin, Germany). A set of intron-spanning primers was designed using the Ensembl prediction for *drKiaa1202,* which itself was confirmed by extensive multiple sequence alignment analyses. These primers are listed in Appendix E. The primers were tested on WIK genomic DNA and cDNA derived from total RNA. The DNA and RNA were isolated from adult body. A single PCR product of the expected size was amplified from the cDNA but not from the genomic DNA. Surprisingly, conventional RT-PCR on total RNA from WIK embryonic and larval stages 1 – 16 cells (1.5 hpf), 1000 cells (3 hpf), Shield (6 hpf), 18 – 19 somites (16 hpf), Prim-5 (24 hpf), Long-pec (48 hpf), and Day 4 (96 hpf) indicated that *drKiaa1202* is not expressed at detectable levels during zebrafish embryogenesis up to 96 hpf. Therefore, whole mount ISH was not pursued.

C.1.3. Mutation screening of *hKIAA1202*

Using *hKIAA1202*'s established genomic organisation as a guideline, we screened the *hKIAA1202* locus for pathogenic changes. Our extensive mutation analysis included searching for small localised mutations, large chromosomal rearrangements and expansion of simple repeats.

C.1.3.1. Single-nucleotide exchanges

SSCP and DHPLC analyses recovered three novel *hKIAA1202* sequence variants, c.4116G>T which results in p.L1372F, and c.2957+21T>C and c.*333G>A.

Guided by *hKIAA1202*'s genomic structure, we screened DNA samples from 196 patients diagnosed with XLMR for small localised mutations, such as missense and nonsense exchanges, or minor deletions or insertions. To this end, we performed SSCP analysis on PCR-amplified 170 – 350 bp products covering the 5' UTR, the complete coding sequence (including exon 2a), the 3' UTR up to the start of the L1M4 and *Alu*Y repeats at position 5565, and approximately 40 nucleotides of flanking intronic sequence around each exon. Exon 2b was not included in the analysis. As a control, we used identical conditions to screen 168 control X chromosomes (Controls A). We detected a missense exchange, c.4116G>T, leading to

p.L1372F and occurring in a Dutch family, but absent in the control panel (Table IV-2 and Fig III-20b). The p.L1372F alteration is located in the ASD2 domain and the affected leucine residue is semi-conserved from human to pufferfish. It lies just N-terminal of an amino acid stretch, of which 78% (18/23) of the residues are identical among vertebrates and which is part of a predicted leucine zipper motif (PS00029) (Fig. III-24a). However, this transversion did not segregate with the disease within the proband's family; the putative mutation was found in two affected brothers, but also in a brother considered to be within the normal range of intelligence. Through several international collaborations, four additional sequence exchanges in *hKIAA1202* were recovered using SSCP (Table IV-2 and Fig III-20b)[199].

DNA samples from 25 Euro-MRX families were screened for mutations by DHPLC, using the same PCR conditions and primers as in the SSCP screen. All 25 families were linked to Xp11, a gene-dense region on the X chromosome also harbouring *hKIAA1202*. An overview of the mutation screening performed on each of the Xp11-linked families is given in Table III-9. We detected two novel single-nucleotide exchanges (Table IV-2 and Fig III-20b)[a]. c.2957+21T>C and c.*333G>A are exchanges in the non-coding sequence of the *hKIAA1202* gene, the first one 3' of exon 5, the second one in the 3' UTR. For c.2957+21T>C, we PCR-amplified the corresponding DHPLC amplicon from male control genomic DNAs and digested the products with *Alu*I. *Alu*I digestion resulted in three restriction fragments (138, 55 and 47 bp) for the T-variant and four fragments (113, 55, 47 and 25 bp) for the C-variant. Both variants were found in 46 controls, indicating that c.2957+21T>C is not a disease-causing mutation. Although we did not formally test if c.*333G>A alters the function of the hKIAA1202 protein, we believe it to be unlikely.

Table III-9 | **Mutation analysis of families linked to Xp11**

DNA	SSCP[§]	DHPLC[§]	Genomic rearrangement[§]	Variable repeats[§]	Exon6/2 direct sequencing[§]
D2					
L17					
L38					
L45					
MRX1					

[a] We screened sixteen DNA samples from individuals with suspected XLMR by employing both SSCP and DHPLC (Table III-9). SSCP did not uncover any sequence exchange among these patients. However, DHPLC analysis revealed two different sequence exchanges (c.2957+21T>C and c.*333G>A, see Table IV-2) in eight DNA samples. Although our sample number is small, this observation seems to be in accordance with the literature[770,771] and implies superior sensitivity of DHPLC compared to SSCP.

Table III-9 | **Mutation analysis of families linked to Xp11**

DNA	SSCP[§]	DHPLC[§]	Genomic rearrangement[§]	Variable repeats[§]	Exon6/2 direct sequencing[§]
MRX12					
MRX15					
MRX17					
MRX26					
MRX31					
MRX44					
MRX45					
MRX65					
N9					
N17					
N39					
N40					
N42					
N45					
N61					
T3					
T25					
T40					
T50					
T102					
4291					
HA					
MO					
SHS					
WTS					
WWS					

[§] Grey shading indicates screened samples.

C.1.3.2. Genomic rearrangements

No obvious genomic rearrangements involving *hKIAA1202* are apparent.

As disease has been shown to arise from global chromosomal rearrangements[772], we considered the possibility of large-scale inversions and deletions.

We searched a genomic region of ~1900 kb centred around, and including, *hKIAA1202* for direct and inverted repeats using REPuter software. The genomic region under investigation harboured one inverted repeat (2.3 kb, 98 % ID) and one direct repeat (1.5 kb, 97 % ID).

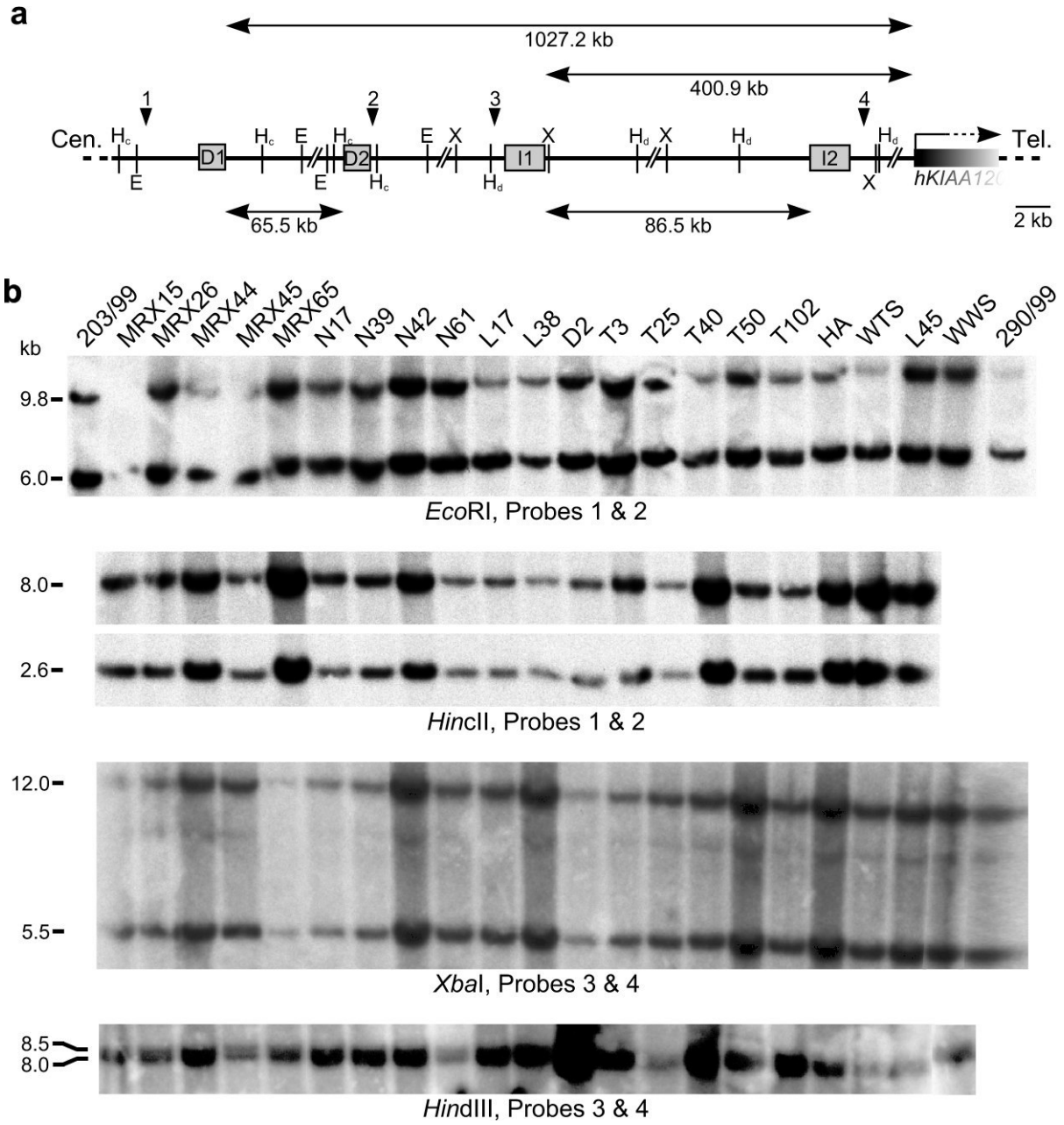


Fig. III-22 | **Screening *hKIAA1202* and its surrounding genomic region for genomic rearrangements.**

a. A restriction map (H_c, *Hinc*II; E, *Eco*RI; X, *Xba*I; H_d, *Hin*dIII) shows the sequence flanking the direct (D1, D2) and inverted (I1, I2) repeat copies found 5' of *hKIAA1202* (fading rectangle) in a 1900 kb region encompassing *hKIAA1202* from centromere (Cen.) to telomere (Tel.). Black lines represent genomic sequences, which are only partially shown (double slash). Arrowheads symbolise the Southern blot probes employed in panel b.

b. Two male control DNA samples (203/99 and 290/99) and 21 patient DNA preparations (mentioned at the top) are digested, size-separated, transferred and hybridised with two pooled probes, as indicated. Patient WWS is missing from the *Hinc*II digest and controls were not included in all hybridisations. No apparent rearrangements are observed. Expected sizes are 9406 bp (Probe 1) and 5714 bp (Probe 2) for the *Eco*RI digest, 8296 bp (Probe 1) and 2430 bp (Probe 2) for the *Hinc*II digest, 5325 bp (Probe 3) and 12053 bp (Probe 4) for the *Xba*I digest, and 8414 bp (Probe 3) and 7985 bp (Probe 4) for the *Hin*dIII digest. Size markers on the left.

The copies of the inverted repeat and of the direct repeat lie 86.5 kb and 65.5 kb apart, respectively. The most distal copy of the inverted repeat and of the direct repeat are located 400.9 kb and 1027.2 kb from the 5' end of *hKIAA1202*, respectively (Fig. III-22a). Both copies of each repeat are unique to the ~1900 kb interval under study. Genomic DNA from 21 Euro-MRX families, all linked to Xp11, and from two male controls were digested with either *Eco*RI, *Hin*cII, *Hin*dIII or *Xba*I. The digested DNA was screened for chromosomal rearrangements at the sites of the direct and inverted repeats using Southern hybridisation with pools of two PCR-amplified and RA labelled probes. No rearrangements were apparent (Fig. III-22b).

It should be noted that A. Dubos and Dr. A. Hanauer from the University of Strasbourg, France found that the X-chromosomal BP of a t(X;19) translocation in a mildly mentally retarded female patient disrupts *hKIAA1202* in intron 2a (Table IV-2)[199].

C.1.3.3. Variable repeats

Three patient-specific genotypes within *hKIAA1202* variable repeats were observed.

Apart from global rearrangements at the chromosome level, many neurodegenerative diseases are known to occur through expansions of simple repeats[773]. Using a modified primer prediction algorithm (courtesy of Dr. S. Haas, MPI-MG, Berlin, Germany), we scanned the ~1900 kb genomic region (described in the previous section), including *hKIAA1202*, for stretches of simple repeats. Nine such stretches were investigated by PCR, but in pilot studies, only four repeats proved to show variability in length.

Length of the first repeat, an imperfect GA repeat in intron 1 (position 71259 – 71340 on BAC RP11-554P16), was correspondingly variable between patients and controls.

The second repeat, a short, perfect AC repeat between exons 2a and 3 (position 52521 – 52570 on BAC RP11-119E20), contained 25, 29 and 31 dimers in the genomic sequence, the control presenting with the longest repeat length and a patient (T50), respectively (Table IV-2).

Investigation of the third repeat, an exonic repetitive sequence (position 3381 – 3515, *hKIAA1202* isoform I) translating into an imperfect poly-Q and poly-E repeat, led to the identification of the following three haplotypes (variability is highlighted in bold face):

- $(CAG)_3(AAG)(CAG)(CAA)\mathbf{(CAG)_3(CAA)}(CAG)_4(AAG)(CAA)(CAG)\mathbf{(GAG)_7}(GAA)_5(GAG)(GAA)(GAG)(GAA)(GAG)_5(GCA)\mathbf{(GAG)_6}$
- $(CAG)_3(AAG)(CAG)(CAA)(CAG)_4(AAG)(CAA)(CAG)\mathbf{(GAG)_6}(GAA)_5(GAG)(GAA)(GAG)(GAA)(GAG)_5(GCA)\mathbf{(GAG)_3(GAA)(GAG)_2}$

- $(CAG)_3(AAG)(CAG)(CAA)\mathbf{(CAG)_3(CAA)}(CAG)_4(AAG)(CAA)(CAG)\mathbf{(GAG)_7}(GAA)_5(GAG)(GAA)(GAG)(GAA)(GAG)_5(GCA)\mathbf{(GAG)_3(GAA)(GAG)_2}$

coding for $Q_3KQ_{10}KQ_2E_{21}AE_6$, $Q_3KQ_6KQ_2E_{20}AE_6$ and $Q_3KQ_{10}KQ_2E_{21}AE_6$, respectively.

Direct sequencing revealed the first haplotype in 33.3% of patients (9/27) and in 13.8% of controls (4/29). The second haplotype was found in 63.0% of patients (17/27) and in 86.2% of controls (25/29). The third haplotype was observed in one patient (N61) and in none of the controls. It should be noted that patient N61 also carries the silent c.3426A>G exchange (Table IV-2) located within the **$(GAG)_7$** repeat and, therefore, the patient's haplotype actually is ...(CAG)**$(GAG)_8(GAA)_4$**(GAG)...

For the fourth repeat, a short, nearly perfect stretch of AAAG tetramers, we directly sequenced PCR products amplified from genomic DNAs of 81 European male controls and from one patient (N61) who was an outlier upon PAGE analysis (Table IV-2). Interestingly, this repeat is contained within the *hKIAA1202* mRNA, forming part of the AluY element at the ultimate 3' end of the *hKIAA1202* message (position 5932 – 5988, *hKIAA1202* isoform I). Including the *hKIAA1202* mRNA and the genomic clone containing most of the *hKIAA1202* gene (RP11-119E20), we found several haplotypes which can be denoted as $(AAAG)_\alpha(AAAAG)(AAAG)_\beta$ with a relative frequency of 69.9% (58/83), 1.2% (1/83), 24.1% (20/83), 3.6% (3/83) and 1.2% (1/83) for $\alpha - \beta$ alleles 7 – 5, 7 – 6, 8 – 5, 9 – 5 and 14 – 5, respectively (Table IV-2). Patient N61 carries a unique 14 – 4 allele which is only similar to the rare 14 – 5 haplotype and varies from all those found among healthy controls in the least variable bit of the repetitive sequence, that is, 'β'.

Even though we found patient-specific genotypes in the last three variable repeats discussed (Table IV-2), differences between patients and controls were small, so we did not pursue the possibility of involvement of these repeats in XLMR any further.

C.1.3.4. Attempt to generate an *mKiaa1202* knock-out

Collaboration with B. Greber, MPI-MG, Berlin, Germany.

ENU-treated ES cell clones carrying an *mKiaa1202* mutation were not implanted, due to the hypomorphic nature of the mutation.

To model the absence of hKIAA1202 in the t(X;8) patient, an attempt was made to knock *Kiaa1202* out in the mouse. An *mKiaa1202* KO would not only provide a handle on Kiaa1202 function, it would also allow for studies on the whole-organism level.

A library containing ~40000 ENU-mutagenised E14.1 ES cell clones derived from 129/Ola mice was screened as described by Greber *et al.*[774]. Essentially, ENU-induced muta-

tions causing aberrant splicing likely to disrupt proper functioning of the gene product were identified by nested exon-skipping RT-PCR (Fig. III-23). The library was screened for *mKiaa1202* exons 2, 3, 4, 5, 7 and 8. Primers are listed in Appendix E. Aberrant exon 4 splicing was initially detected in 6 pools of ES cell clones and could subsequently be confirmed in 11 single ES cell clones. However, apart from the aberrant splicing, PCR amplification across exon 4 with standard primers also detected correct splicing of exon 4. Comparison of the amount of correctly and aberrantly spliced transcripts indicated that an estimated > 90% of transcripts were correctly spliced in the selected ES cell clones. Due to the hypomorphic nature of the mutation and the fact that *mKiaa1202* is a single copy X-chromosomal gene, injection of mutant ES cell clones into C57BL/6 blastocysts followed by embryo transfer to oviducts of pseudopregnant NMRI females was deemed unreasonable.

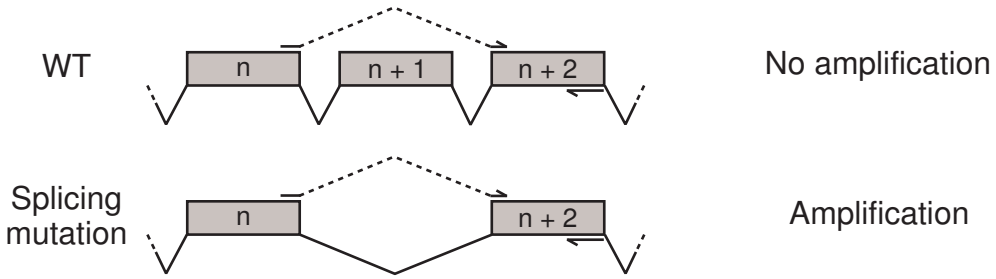


Fig. III-23 | **Principle of the exon-skipping reverse transcription polymerase chain reaction.**

The top panel depicts part of a transcript, including three of its exons (grey rectangles). The polymerase chain reaction using an exon *n + 2*-specific primer (arrow) and an exon-skipping primer (dashed arrow), of which the 5' end is specific to exon *n* and the 3' end is specific to exon *n + 2*, selectively amplifies transcripts lacking exon *n + 1*. When such transcripts do not normally occur, but are the consequence of a splicing mutation (bottom panel), this method allows for the rapid identification of aberrantly spliced mRNAs.

C.2. Computational studies

We performed *in silico* analyses on the hKIAA1202 sequence to identify putative hKIAA1202 homologues and to gain insight into the domain structure of the protein.

C.2.1. Identification of hKIAA1202 homologues

hKIAA1202 has putative homologues among vertebrates.

The conceptually translated hKIAA1202 protein sequence contains an N-terminal PDZ domain[775], a putative EVH1-BS[776] situated around position 1100, an ASD2 domain[777] in its C-terminus and a consensus PDZ-BS[778] (Fig. III-20c). Based on protein predictions and (partial) mRNA sequences, cross-species global alignment analyses of the hKIAA1202 protein against all available sequence data identified putative homologues among vertebrates

(Fig. III-24a). As the protostomia and their subranks were evolutionarily too distant from *H. sapiens* to distinguish unambiguously between true hKIAA1202 homologues and homologues of sequences similar to hKIAA1202, such as other members of the APX/Shrm protein family (see next section), protostomia sequences were omitted from the analysis. The *P. troglodytes* prediction was also excluded from the alignment because it was clearly incorrect. Also, the *T. nigroviridis* and *D. rerio* predictions were left out due to their low overall homology, which would upset the multiple alignment, thereby abolishing the visual representation of overall similarity between the other Kiaa1202 homologues. However, the *P. troglodytes*, *T. nigroviridis* and *D. rerio* sequences were included in the ASD2 alignment. Owing to incomplete genome sequences, putative *M. mulatta* and *M. domestica* Kiaa1202 were truncated at their N-terminus. Similarly, the *M. mulatta* positions indicated with 'X' in Fig. III-24a are caused by unfinished genome sequencing. Three predictions needed refinement: to extend the prediction up to the first methionine residue encountered in the 5' upstream sequence, seven and eight N-terminal residues were added to *X. tropicalis* and *F. rubripes* Kiaa1202, respectively. In addition, fifteen C-terminal amino acids were added to the *F. rubripes* prediction, extending it down to the first termination codon in the 3' flanking sequence. A 43 AA imperfect poly-Q stretch needed to be inserted manually in putative *C. familiaris* Kiaa1202, as it was overlooked by the prediction algorithm, probably because of its low complexity. Even though overall sequence identity among Kiaa1202 homologues varied from 91.7% (*R. norvegicus* vs. *M. musculus*) to 22.2% (*M. domestica* vs. *F. rubripes*) (Fig. III-24b), a consistent high degree of conservation existed at the N-terminus, comprised of the PDZ domain. Remarkably, the *H. sapiens*, *R. norvegicus*, *M. musculus* and *C. familiaris* Kiaa1202 PDZ domains are 100% identical (Fig. III-25). Similarly, conservation is high at the C-terminal end of the sequence corresponding to the ASD2 domain (Fig. III-26). While the putative proline-rich EVH1-BS is highly conserved among mammals, its conservation is less pronounced in *X. tropicalis* and *F. rubripes*. Similarly, the extreme C-terminus is well conserved among mammals, but that of *X. tropicalis* and *F. rubripes* Kiaa1202 diverges considerably from the consensus. Whereas the former is shortened by seven residues, the latter has a unique twelve amino acid insertion. Although shortening of the *X. tropicalis* C-terminus may appear as an inaccuracy in the prediction, this does not seem to be the case, as the last amino acid is immediately followed by a TAG stop codon. The last four residues of the eutherian homologues comprise a predicted PDZ-BS that putatively binds class I ligands. *M. domestica* and *F. rubripes* seem to be lacking such a PDZ-BS. The situation is ambivalent for *X. tropicalis* Kiaa1202: on the one hand, its four final amino acids are predicted to constitute

a PDZ-BS that binds class III ligands, on the other hand this may well be a chance event, as the requirements for a consensus PDZ-BS are not very specific[778]. Two arguments support the idea of a chance event. First, if the predicted PDZ-BS would represent an actual PDZ-BS, then the ligand binding class would differ between eutherian and frog Kiaa1202, which is unlikely. Second, as *X. tropicalis* is located in evolution between *M. domestica* and *F. rubripes*, both lacking a PDZ-BS, it seems improbable that the frog homologue would bind PDZ domains.

C.2.2. Analysis of protein organisation

hKIAA1202 belongs to the APX/Shrm protein family of cytoskeleton-associated proteins.

In silico translated hKIAA1202 shares the domain structure of the novel APX/Shrm protein family. This protein family is characterised by a PDZ domain, two ASD domains (ASD1 and ASD2), and putative EVH1 and PDZ BSs[777]. All family members characterised to date associate with the cytoskeleton and some of them were demonstrated to be crucial to cytoarchitecture. See IV.F.3 for a more detailed explanation.

The APX/Shrm protein family consists of four founding members and their homologues[200]. The founding members are xApx[779], the human proteins hAPXL[780] and hKIAA1202[199], and mouse mShrm[777]. From the *X. tropicalis* genome project, we now know that hAPXL is actually the homologue of xAPXL and that hAPXL2[781] is likely the xApx homologue (Dr. J. Wallingford, personal communication).

xApx lacks the PDZ domain and the EVH1-BS, hAPXL lacks a PDZ-BS, and hKIAA1202 does not contain any obvious ASD1 domain (Fig. III-20c). Therefore, the ASD2 domain seems to be the common denominator among family members. Both *mShrm* and *hKIAA1202* code for a protein containing the PDZ domain (mShrmL and isoform I, respectively) and for a variant missing this domain (mShrmS and isoform IV, respectively). mShrmS lacks the first 177 residues (i.e. 8.9% of mShrmL) and hKIAA1202 isoform IV lacks the first 130 residues (i.e. 8.7% of isoform I).

Global multiple alignments of genomic sequences clearly show that the founding proteins are not simply encoded by homologous genes. In fact, they are four different proteins, showing similarity in their domains. The hKIAA1202 PDZ and ASD2 domains are ~60% identical to those of the other APX/Shrm family members and, although homology between the different ASD1 domains is consistently lower than that between the ASD2 domains, sequence identity still ranges from 30 to 45% (Fig. III-27). Bioinformatics-based searches identified

Shrm-related proteins in all chordates. In addition, insect genomes, including those from *D. melanogaster*, *A. gambiae* and *A. mellifera*, encode a partially-related protein containing an ASD2 domain. Finally, BLAST searches of the deposited sequences from invertebrate genome projects identify what can be considered Shrm orthologues in both *C. intestinalis* (Dr. J. Hildebrand, personal communication) and *S. purpuratus*. All members of the APX/Shrm family identified to date, including their recently adopted new names that eliminate the confusing existing nomenclature, are listed in Table III-10.

Table III-10 | **New nomenclature for Shrm-related proteins**

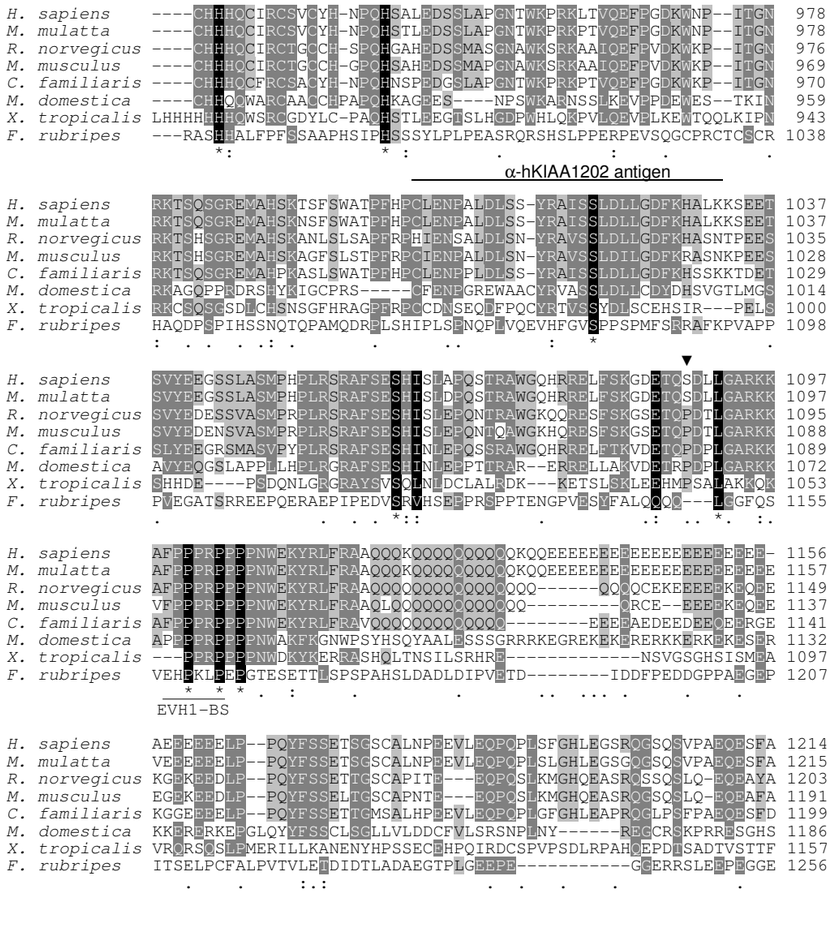
Accession #	Previous name	New name
CAA78718	*X. laevis* Apx	xShroom 1
NP_597713	*H. sapiens* APXL2	hShroom 1
CAA58534	*H. sapiens* APXL	hShroom 2
ABD19518	*M. musculus* Apxl	mShroom 2
AAF13269	*M. musculus* ShrmL	mShroom 3a
AAF13270	*M. musculus* ShrmS	mShroom 3b
NP_065910	*H. sapiens* Shrm	hShroom 3
ABD59319	*X. laevis* Shrm-like	xShroom 3
NP_065768	*H. sapiens* KIAA1202	hShroom 4a
AAK95579	*H. sapiens* SHAP-A	hShroom 4b
DQ435686	*M. musculus* KIAA1202	mShroom 4
ABA81834	*D. melanogaster* Shrm	dmShroom
EAA12598	*A. gambiae* Shrm	agShroom
XP_392427	*A. mellifera* Shrm	amShroom
XP_783573	*S. purpuratus* Shrm	spShroom

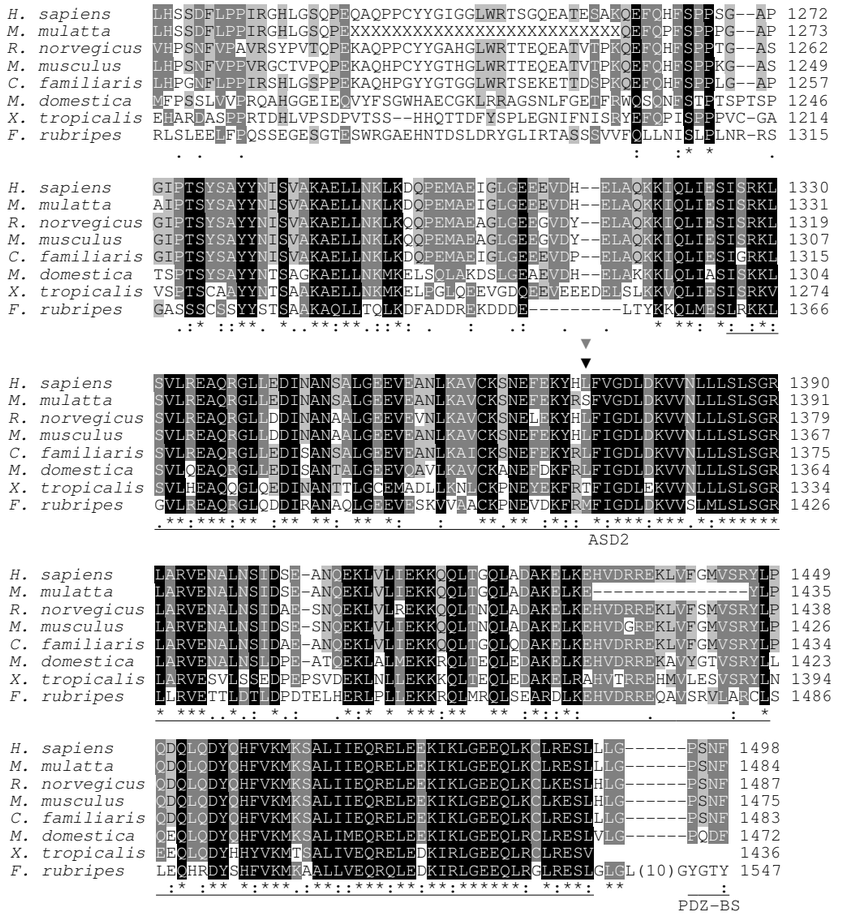
Fig. III-24 | *Next page.* **Cross-species global multiple sequence alignment of hKIAA1202 homologues.**

a. hKIAA1202 is aligned with several of its vertebrate homologues. See text for details on the manual refinement made to the *X. tropicalis*, *F. rubripes* and *C. familiaris* Kiaa1202 predictions. hKIAA1202's PSD-95/Dlg/ZO-1 (PDZ) domain, Enabled/Vasodilator-Stimulated Phosphoprotein homology domain 1 binding site (EVH1-BS), APX Shroom domain 2 (ASD2) and PDZ-BS are underlined. Multiple sequence alignments for the PDZ and ASD2 domains are presented in Figs. III-25 and III-26, respectively. The α-hKIAA1202 antigenic site is indicated. Black arrowheads represent amino acid exchanges published in Hagens *et al.*[199], the grey arrowhead represents the amino acid exchange recovered in this study. Please note that additional exchanges at the DNA level, as well as variants of the polyQ/E repeat (position 1121 – 1163), have been recovered. An asterisk marks identities, a colon conserved substitutions and a dot semi-conserved substitutions. Shades of grey highlight similarities within subsets of sequences.
b. Sequence identity matrix based on the alignment shown in panel a. The first 82 amino acids were omitted from the analysis, as they are missing from the *M. domestica* prediction. Hs, *H. sapiens*; Mm, *M. mulatta*; Rn, *R. norvegicus*; Mmu, *M. musculus*; Cf, *C. familiaris*; Md, *M. domestica*; Xt, *X. tropicalis*; Fr, *F. rubripes*.

a

```
H. sapiens      MENRE--GSFQYVPVQLQGGAPWGFTLKGGLEHGEPLTVSKIEDGGKAALSQKMRTGDEL   58
M. mulatta                                    IEDGGKAALSQKMRTGDEL   58
R. norvegicus   MENRE--GSFQYVPVQLQGGAPWGFTLKGGLEHGEPLTVSKIEDGGKAALSQKMRTGDEL   58
M. musculus     MESRE--GSFQYVPVQLQGGAPWGFTLKGGLEHGEPLTVSKIEDGGKAALSQKMRTGDEL   58
C. familiaris   MENPE--GSFQYVPVQLQGGAPWGFTLKGGLEHGEPLTVSKIENGGKAASQKMRTGDEL   58
M. domestica                                                              58
X. tropicalis   MDPQEADGSSQGIHVQLQGGAPWGFTLKGGLEHGEPLISKIENGGKASVGEKMEVGDEL   60
F. rubripes     METVEQLVSFHHIQVQLSGGAPWGFTLKGGLEHGEPLITKIEDGGKAAVGEKLKVGDEL   60
                *:      * : : ***.*************** *** ::***:****::.:*:..****
                                                  PDZ

H. sapiens      VNINGTPLYGSRQEALILIKGSFRILKLIVRRRNAPVSRPHSWHVAKLLE----------  108
M. mulatta      VNINGTPLYGSRQEALILIKGSFRILKLIVRRRNAPVSRPHSWHVAKLLE----------  108
R. norvegicus   VNINGTPLYGSRQEALILIKGSFRILKLIVRRRNTPVSRPHSWHVAKLLE----------  108
M. musculus     VNINGTPLYGSRQEALILIKGSFRILKLIVRRRNTPVSRPHSWHVAKLLE----------  108
C. familiaris   VNINGTPLYGSRQEALILIKGSFRILKLIVRRRNAPVSRPHSWHVAKLLE----------  108
M. domestica                              FLPILVGRRTTPLARPHSWHMAKLLE----------  108
X. tropicalis   VNINGTPLYGSRQEALILIKGSYKILRMIVRRRNLSVIRPHSWHLAKLTE----------  110
F. rubripes     ININGSTLYGSRQEALILIKGSYRILKLTVRRRSVEVIRPHSWHPAKLSEPPLPPPPPPP  120
                :****:.**************:::* : * **. .: ****** *** *

H. sapiens      ------------GCE--EAATTVHFPSEAFSLSWHSGCNTSDVCVQWCPLSR-HGSTEKS  153
M. mulatta      ------------XXX--XXXXXXXXXXXXXXXXXXXXXXXXDVCVQWCPLSR-HGSTEKS  153
R. norvegicus   ------------GCE--DVATTVHFPSEAFSLSWHSGCNTSDVSVQWCPLSR-HGSTEKS  153
M. musculus     ------------GCE--DVATTVHFPSEAFSLSWHSGCNTSDVSVQWCPLSR-HGSTEKS  153
C. familiaris   ------------GCE--EAATTVHFPSEAFSLSWHSGCNTGDVCVQWCPLSR-HGSTEKS  153
M. domestica    ------------GREVAAAAAVNFPEDPFSLSWHSGCDTSDPCVQWCPLSR-HGSTEKS  155
X. tropicalis   ------------VHP---DVASVQYPTDAFSLSWHSGCENSELPMQWNPLSR-HGSTDKS  154
F. rubripes     SPPPCLPPADSPPLEPPPPLSAVQLHPGPYTLPWHT-TDNSDFSLQWGQLSRPYSSTDRS  179
                                                :  :**  *** : **::*

H. sapiens      SSIGSMESLEQPG---QATYESHLLPIDQNMYPNQRDSAYSSFSASSNASDCALS-LRPE  209
M. mulatta      SSIGSMESLEQPG---QATYESHLLPIDQNMYPNQRDSAYSSFSASSNASDCALS-LRPE  209
R. norvegicus   SSIGSMESLEQPG---QATYEGHLLPIEQNMYPSQRDSAYSSFSASSNASDCALS-LKPE  209
M. musculus     SSIGSMESLEQPG---QPTYEGHLLPIDQNMYPSQRDSAYSSFSASSNASDCALS-LKPE  209
C. familiaris   SSIGSMESLEQPG---QATYEGHLLPIDQNMYPNQRDSAYSSFSASSNASDCALS-LRPE  209
M. domestica    SSIGSMESLEQPS---QVCQEGHVSPVEQSTYQSQRDSAYSSFSAGSSASDYTLP-LRLE  211
X. tropicalis   SSIGSMESLDQPG---HNYYEGTLSPIDPSMYQNKRDSAYSSFSASSNASDYTLS-ARAE  210
F. rubripes     SSLGSMESLDTETPTTQPYSDSCNSPMDPAAENNKRDSAYSSFSASSNTSDYAMVPLRTE  239
                **:******: *    :  :.   *::   : .:********.*.:** ::   : *

H. sapiens      EPASTDCIMQCPGPTKAPSGRPNVAETSGG---SRRTNGGHLTPSSQMSSRPQEGYQSGP  266
M. mulatta      EAASTDCIMQCPGPTKAPSGRPNVAETSGG---SRRTNGGHLTPSSQMSSRPQEGHQSGP  266
R. norvegicus   EPPSTDCVMSCPGPTKVTDDQANVSETSGS---SHSTSGGHMTSTSHASSCSEEGHHSGP  266
M. musculus     EPPSTDCVMPCPGPIKVTDDQANVSENSGS---SHSTSEDHVLSTSHASSYSDEGHHSGP  266
C. familiaris   EPGSADCVMQCSGPAKAPNGRPSVAETSGG---SRRPNGSHPLLSSQMSSCPEEDHHSEP  266
M. domestica    ETTASDCSPLCLGPAKGQDGIYLHTGPDCEETLGAAGPSGHQLPRSQPSPSPQEGPLSGS  271
X. tropicalis   ESSQMDCVAD---SSKPCDGRYLHTGQGAIEIQQETSSLSPSEHQQRPSSFPFDANHLSF  267
F. rubripes     DGCSMDNLLQSLGPSPHGYSGAPTLGGSSVEGTQLLMQKSRSLTRPRPRPVEVKERPSSC  299
                :   *    . ..    .      ..       .     :  .  .

H. sapiens      ---------------------AKAVRGPQPPVRRDSLQASRAQLLNGEQRRASEPVVPPP  306
M. mulatta      ---------------------AKAVRGPQPPVRRDSLQASRAQLLNGEQRRASEPVVPPP  306
R. norvegicus   ---------------------AKMVRGPEPPVRRDSLQASRAQLLNGEQHRASEPVDSPQ  306
M. musculus     ---------------------AKMARGPEPPVRSDSLPASRAQLLNGEQHRASEPVDSPP  306
C. familiaris   ---------------------AKAARGPQPPVRRDSLQASRAQLLNGEQRRASEPVDSPQ  306
M. domestica    SQV-------------------AVASTGPPQPPVRRDSLRASKTQFRQTEKRRASVEGGAPY  314
X. tropicalis   ------------------------IKSPPQPPIRRDSLRASKNQICHGERRRASAPGDSLQ  304
F. rubripes     CYDNELKGGDFELLRNGESGAEGRKKAPQPPARKDSFRATRSRPNVADKRCVSAPVEIPS  359
                                        :    .*:** * **: *:: :    ::: .* *

H. sapiens      QKEKLSLE-PVLPARNPNRECCLSCHDQVTSEGHQNCEFSQPP------ESSQQGSE-HL  358
M. mulatta      QKEKLSLE-TVLPARNPNRECCLSCHDQVTSEGHQNCEFSQPP------ESSQQGSE-HL  358
R. norvegicus   QKEKPALE-TMLPPRSSNQECCISCQDQVTGEDHQNCELSKPS------ESSQDDCD-HL  358
M. musculus     QKEKPGLE-TVLPPRSSNQECCLSCQDQVTDEDHQNCELSKPS------ESSQDDCE-HL  358
C. familiaris   Q-EKVSLD-TVLSPRNPNRECCLSCQDQVTNEGHQNCELSQAP------ESSQQDSE-HL  357
M. domestica    TKERWNSE-TSLSSDVPPGPGCPCSQVLEPHEEHWKDDLAQEQYCVLSHPIGLDHCESRL  373
X. tropicalis   ISGMWSSE-NQQ-HKNSDTSQCKCGIEFCT--VHLKNGLSSDQ---YYMLSSQTDGGNQS  357
F. rubripes     VPSYDDGNNPPLNAGIHNGLSGPGDKDRTRDASNSLTNSGFDK------DRRSQTPLDLF  413
                          :                .       :       .             :

H. sapiens      LMQASTKAVGSE---KACDRASSVDSNPLNEASAELAKAS-FGRPPHLICPTGHRHSAPE  414
M. mulatta      LMQASTKAVGSE---KACDKASSTDSNPLNEASAELAKAS-FGRPPHLIGPTGHRHSAPE  414
R. norvegicus   LIEGSSKALDSE---KAHDKGSNKEFGLLKEASADLANTLSFGAIPHLRGSIEHRHSAPE  415
M. musculus     LIEDSSKALDSE---KAHDKGSNKEFGLLKEASADLANTLNFGAIPHLRGTMEHRHSAPE  415
C. familiaris   IMEASAKAAGFE---KACDKASSIDYSPLNKASAELAKAS-FGGPPHLTGPTGHRHSAPE  413
M. domestica    LLKDCPQRTGWEGQDQIREESPPSHSPGRGPEAKEEAKAALTTIPAHLTGLTGHRHSAPE  433
X. tropicalis   TDQLALSDREME---SCCTEMQSRWPRDGRNIKQTLNKEMENSSYHSAQTVKTVKDSLSK  414
F. rubripes     PDREIASAAHLR-------SSCSGSLSPEPSLEAQGPQLNIRRAKTSRASVGLHRHSAPE  466
                  .   .        .                    :                :.* .:

H. sapiens      QLLASHLQHVHLDTRG---SKGMELPPVQDCHQWTLSPLHSSHKCKSPCPETGCTHDQS  471
M. mulatta      QLLASHLQHVHLDTRG---SKGMELPPGQDCQQWTLSPLHSSHKCKSPCPEAGCTHDQS  471
R. norvegicus   QLLASHLQQMHLDSRG---NKGMELPTGQDCHQWTMSPLHNNPKCKKSPGLETGCTQDQT  472
M. musculus     QLLASHLQQVHLDSRG---SKGMELPIGQDCHQWTVSPLHNNPKCKKSPSLETGCTQDQT  472

C. familiaris   QLLASHLGYGHLDTRG---SKGTELPAGHDGHEWTLSPLHSSHTGKKSPCPETGGTQDQH  470
M. domestica    QLLTSQIRSLHLSVEGDGNSEETELQAKQEGHCWTLSPLHSAHVGPKSPCHEGECAQNKP  493
X. tropicalis   HPSCLHMKSSSLPQSE---QEEVSVQTKFHRKEWRNTLLQENFQCQTSECNGISSQELKC  471
F. rubripes     KLLATQLQLLQFASDGS-SPEPRGQDDANDGCGRAAQPTGADHQGASHSHLHANKWEGSR  525
                :  :::  :     :          .                .       . .

H. sapiens      SKERKTRQVDDRSLVLG-HQSQSSPPHGEADGHPSPKGFLDPNRTSRAASELANQQPSAS  530
M. mulatta      SKERKTRQMDDRSLVLG-HQSQSSPPHGEADGHPSPKGLLDPNRTSRAGNELANQQPSAS  530
R. norvegicus   SKERKIKPTDDKLMVSA-NQSQSDALLGEVDGHPSPRTYLDSNRASRASSDLTGQQPSAT  531
M. musculus     RKERKTTPLDDKLMASV-HQSQSDVLLGEVDGHP--------NRAGRASSDLTSQQPSAT  523
C. familiaris   SKERKARQVDDRPLCSG-HQSPSSSPLGEADGHPPPKGFQDPNRASGAGSESASQQPSAS  529
M. domestica    GEKWMEGKAGSPALASE-SSGQSSSPTPGGTSCSHQP---------ESAPSEPVTVQLSGL  543
X. tropicalis   NINRDSEIYQVDTLTDTGDFSQLSGKNKEQSQHITT---------LERSVSEPNEVRESFP  523
F. rubripes     SSTPGSVFLEGDEDISSEKVGGLSINGSCHPPVEHQHPWGRSVSVPGDAVDMAAPGRSEP  585
                                          . .          .       :       *

H. sapiens      GSLVQQATDCSSTTKAASGTEAGEEGDSEPKECSRMGGRRSGGTRGRSIQNRRKSERFAT  590
M. mulatta      GSLVQKARDCSSTTKVAGGTEAGEEGDSEHKECSRMGGRRSGGTRGRSIQNRRKSERFAT  590
R. norvegicus   CSPVQQTRDFFSANKVVGHTEATEEGDNEPKEGGRLCGRRSGGPRGRSIQNRRKSERFAT  591
M. musculus     CSSVQQTRDFLSAHKIVDHTEASEEGDNEPKECGRLCGRRSGGPRGRSIQNRRRSERFAT  583
C. familiaris   GSLIQQTRDCSSTTKVADSTEATEEGDSEPKECGWVGGRRHRGPRGRSIQNRRKSERFAT  590
M. domestica    PDVPSQA-----ISKPASHVDTTLEEGTEAAKDKGSENTRRSCQRYRSAQQRRRSERFAT  598
X. tropicalis   VLLPKHSVG---GMRSSCSSETLLESHEQEESQGP-TKKPGSSRHRSAQMRRRSDRFAT  579
F. rubripes     IAERDFEPLSSAASMDTLLEEQKAAEREKSQFRKEEGQANVRRSSRSHFRNRRRSERFAT  645
                      .     :       :       :       .       : **:*:****

H. sapiens      NLRNEIQRRKAQLQKSKGPLSQLCDTKDP--VEETQEPPESPPLTASNTSLLSSCKKPE  647
M. mulatta      NLRNEIQRRKAQLQKSKGPLSQLCDTKDP--VEETQEPPESPPLTASNTSLLSSCKKPE  647
R. norvegicus   NLRNEIQRRKAQLQKSKGPLSQLCDTKED--VEETQEPPESPPLPASNSSLLPSYKNVP  648
M. musculus     NLRNEIQRRKAQLQKSKGPLSQLCDTNEA--VEETQEPPESPPLSASNASLLPSYKNVP  640
C. familiaris   NLRNEIQRRKAQLQKSKNSLSQLCDTQDP--VEETEEPLESPPPTSNSSLLSSYKKPE  646
M. domestica    NLRNEIQRRKAQLQKIKGSLALVGELEDPEPVEETAAPEPSPPLPAPPQVPTLAAHQRLA  658
X. tropicalis   NLRNEIQRRKAQLQKNKGSSVLLGGEPP--VEERREPPTPSQSPPRP------MPPPPPE  630
F. rubripes     NLRNEIQRKKAQLQRSCGPGGLLCSGETVQ-EPEGPPLNPEEADPDP-----DPGAQAVK  699
                ********:*****:  ..   : .:    **       . . . . .

H. sapiens      SPRDKLFNKS----MMLRARSSECLSQAPESHESRTGLEGRISPGQRPGQSSLGLNTWWK  703
M. mulatta      SPRDKLFNKS----MMLRARSSECLSQAPESHESRTGLEGRISPGQRLGQSSLVLNTWWK  703
R. norvegicus   SHGDQLFNKS----MILRARSSECLSQASESSKARVGLEGRMSPGQRSGQSSLALNTWWK  704
M. musculus     SPGDKVFNKS----MILRARSSECLSQASESSKARGGVEGRMSPGQRSGQSSLALNTWWK  696
C. familiaris   SPRDKLFTKS----MILRARSSECLSQAPESHESRTGLEGQISPGQRPGQPSLGLNTWWK  702
M. domestica    SPGERLSDQPQQNGNMLRTRSSECLSQTTESRELWKPLKG---PQQNPWQPIAGLHPWCK  715
X. tropicalis   KNKSRLLELK--------RANAEQISCLDNEEPMQRGREMMFSEPGRSSDHTSLFAAQWR-  682
F. rubripes     VWNAKVVPTP-------PAAPTENSSIPVQSNQPQTATHPETSGRSRSVQILDPGVPSTG  752
                          ::              : . : .  :.        .     :

H. sapiens      APDPSSSDPPKAHAHCGVRGCHWRWSPEFNSQFLVAAAMEGPSNP--------------G  749
M. mulatta      APDPSSSDPEKARAHCGVHGCHWRWSREFNSQFLVAVAMEGPSNP--------------G  749
R. norvegicus   ASDSSTLDTEKANAHHGVCRCHWRWSPEFNAQFQVALSTEGPSNP--------------D  750
M. musculus     ASDSSTLDTEKANAHHGVCRCHWRWSPEFNAQFQVALSTEAPSNP--------------D  742
C. familiaris   VSDSSPSDSEKTNVHRGVRGCHWRWSPEFNLQFHVALAMESPSNL--------------G  748
M. domestica    APG---HDPSKAGSQPGVRGCRWRWSPERKLQFQLMPMSSGVLNP--------------R  758
X. tropicalis   -SRHSSSDFFDPHVQQMPNGGRWKWSPEFKLLEHPQLSKGSPPEVSV------------V  729
F. rubripes     VGIRVVEEPAPAG-----KARRWRWTPEFKLQEEPEPERRCGLSKVQGVTGSRHGVCAFT  807
                      :    .          :*:*: *:: *     .

H. sapiens      DNKELKASTAQAGEDAILLPFADRKFFEESSKSLSTSHLFGLTTHSNKTFTQRPKPID-  808
M. mulatta      DNKELKASTAQAGEDAVLLPFADRKFFEESSKSLSTSHLFGLTTHSNKTFTQRPKPID-  808
R. norvegicus   DNKELKIPTPQACDEAILMPFADRKFFEESSKSLSTSHLFGLTTHN-KPFIQRPKPID-  808
M. musculus     DSKELKTSTPQAGEEAVLMPFADRKFFEESSKSLSTSHLFGLTTHNNKPFIQRQKPID-  801
C. familiaris   DSKELKASTAPAGEEAILLPFADRKKLFEEGSKSLSTSHLFGLTTHSSKTFTQRPKPVD-  807
M. domestica    EE-----MSVTSGEETTFVPFADRKKVFEDTSKSLPPAAS---SLGHSKPFALRPKAVDT  810
X. tropicalis   HAEGASLPNRVASSENVLMPFADRRFFENSSKVHNVSHIELQIKSNKNNYCLSFPDPP-  788
F. rubripes     SSSYGRSSSCSRMEETDLLPFADRVKFFEETCKSVSASHRQRRPQGQPEVQDDPGWRRE-  866
                          ::  ::***** :.**: .*    :         :

H. sapiens      QNFQPMSSSCRELRHPMDQSYHSADQPYHATDQSYHSMSP-LQSETPTYSECFASKGLE  867
M. mulatta      QNFQPVSSNYRELRHPMDQSYHSADQPYHATDQSYHSMSP-LQSETPTYSECFASKGLE  867
R. norvegicus   QNFQSVS--YRELRCHPMDQSYHSADQPYHAADQSYHSLSP-LQSEAPTYPDCFAPKGVD  865
M. musculus     QNFQSVS--YRDLRCHPLDQSYHSADQSYHAADQSYHSLSP-LQSETPTYPECFATKGRD  858
C. familiaris   PNFQPLSSSYRELRHHLMDQSMH-------ATDQSYHSVSL-LQSETPPYSECFASKGLE  859
M. domestica    PAFQPMSTSYRDLRRH--------------SVDQAYHSSAS-LAPEPATYPECFMNKELE  855
X. tropicalis   LSQKVVS----ALRRH--------------SVDHTVHPSSENRQDSALPYSDYCVNHTVD  830
F. rubripes     ---PAPVEGAHQHPQHPPQAQWGPRQHDGYSAPRNGHAFAE-PPAALSLRGRAMSENDLR  922
                      .       *  . .  .  .   :.*:  *. :                .

H. sapiens      NSMCCKPL-HCGDFDYHRTCSYSCSVQCALVHDPCIYCSGEICPALLKRNMMENCYNCR-  925
M. mulatta      NSMCCKPL-HCGDFDYHRTCSYSCSVQCALVHDPCIYCSGEICPALLKRNMMENCYNCR-  925
R. norvegicus   NSLYCKPV-HHGDCDYHRTCSHPCSAQCTVMHDPCICCSGEICPALLKRNLLEKCHNCR-  923
M. musculus     NSLCCKPV-HHGDCDYHRTCSHPCSAQCTVRHDPCICCSGEICPALLKRNLLEKCHNCR-  916
C. familiaris   QSMCCKPL-HCGDFDYHRTCSYSCSVQCAIVHDPCMYCSGEICPALRKRNMMENCYNCW-  917
M. domestica    QPVCYQPLGHCCEFECLSPCSYACGIPCAVGHDTCTYC-----PVLLKRSMLEAHRSYR-  909
X. tropicalis   PLLCCSQGGHAAEYIHHPTG------YCCRVHESCHCCSSDVCPALVKRNMPMSHLSCHF  884
F. rubripes     FDGSHHWSPSVCASTLSEVEEGAGGVGAGEVGSGAPPGRKKTPPPRPPPPKWEQFHRR-  981
                          .       . . .  .  . .    .  *
```





b

Sequence Identity Matrix for Kiaa1202 Vertebrate Homologues

	Hs	Mm	Rn	Mmu	Cf	Md	Xt	Fr
Hs	100	90.1	77.5	77.4	85.2	45.2	37.3	27.0
Mm		100	70.8	70.9	78.2	44.7	34.3	24.3
Rn			100	91.7	74.5	41.6	37.2	26.1
Mmu				100	74.4	41.7	37.3	26.7
Cf					100	44.6	38.2	27.3
Md						100	32.1	22.2
Xt							100	24.7
Fr								100

a

```
H. sapiens      LQGGAPWGFTLKGGLEHCEPLTVSKIEDGGKAALSQKMRTGDELVNINGTPLYGSRQEAL 74
R. norvegicus   LQGGAPWGFTLKGGLEHCEPLTVSKIEDGGKAALSQKMRTGDELVNINGTPLYGSRQEAL 74
M. musculus     LQGGAPWGFTLKGGLEHCEPLTVSKIEDGGKAALSQKMRTGDELVNINGTPLYGSRQEAL 74
C. familiaris   LQGGAPWGFTLKGGLEHCEPLTVSKIEDGGKAALSQKMRTGDELVNINGTPLYGSRQEAL 74
B. taurus       LQGGAPWGFTLKGGLEHCEPLTVSKIEDGGKAALSQKMRTGDELVNINGTPLYGSHQEAL 60
X. tropicalis   LQGGAPWGFTLKGGLEHGEPLIISKIENGGKASMCEKMEVGDELVNINGTPLYGSRQEAL 76
F. rubripes     LSGGAPWGFTLKGGLEHGEPLIITKIEDGGKAAVCEKLKVGDELININGSTLYGSRQEAL 76
                *.***************  ***  ::***:****::.:*:..****:****:.****:****

H. sapiens      ILIKGSFRILKLIVRR 90
R. norvegicus   ILIKGSFRILKLIVRR 90
M. musculus     ILIKGSFRILKLIVRR 90
C. familiaris   ILIKGSFRILKLIVRR 90
B. taurus       ILIKGSFRTLKLILLF 76
X. tropicalis   ILIKGSYKILRMIVRR 92
F. rubripes     ILIKGSYRILKLTVRR 92
                ******:: *:: :
```

b

Sequence Identity Matrix for PDZ Domains							
	Hs	**Rn**	**Mmu**	**Cf**	**Bt**	**Xt**	**Fr**
Hs	100	100	100	100	93.4	81.5	78.9
Rn		100	100	100	93.4	81.5	78.9
Mmu			100	100	93.4	81.5	78.9
Cf				100	93.4	81.5	78.9
Bt					100	75.0	72.3
Xt						100	81.5
Fr							100

Fig. III-25 | **Cross-species global multiple sequence alignment of PSD-95/Dlg/ZO-1 domains from hKIAA1202 homologues.**

a. Apart from *B. taurus*, where the Kiaa1202 homologue is not completely known, numbers refer to positions in the respective Kiaa1202 homologues. An asterisk marks identities, a colon conserved substitutions and a dot semi-conserved substitutions. Shades of grey highlight similarities within subsets of sequences.
b. Sequence identity matrix for the PSD-95/Dlg/ZO-1 (PDZ) domains of several different hKIAA1202 homologues. Bt, *B. taurus*; all other abbreviations as in Fig. III-24b.

C.3. Protein studies

We cloned the hKIAA1202 ORF and generated an α-hKIAA1202 antibody, which enabled us to investigate the over-expressed, as well as the endogenous, protein. We studied hKIAA1202's subcellular localisation and set out to identify some of its interaction partners.

C.3.1. Cloning of the hKIAA1202 open reading frame

Collaboration with J. Ruschmann, MPI-MG, Berlin, Germany.

hKIAA1202 ORF I has been cloned in several mammalian expression vectors.

The 5' end of a full-length *hKIAA1202* cDNA clone (courtesy of Dr. T. Nagase, Kazusa DNA Research Institute, Chiba, Japan), which consists of 6029 bp including a 15 bp poly-A tail (AB033028), was ligated to a *Sal*I adaptor and inserted at the *Sal*I site of the pBluescript II SK+ vector. The 3'end was inserted at the *Not*I site of the vector[757].
In order to clone the hKIAA1202 ORF I in mammalian expression vectors, a 1110 bp fragment containing a *Bam*HI site at its 3' end was PCR-amplified from the *hKIAA1202* cDNA, introducing an *Eco*RI restriction site in front of the start codon, then TA-cloned into pGEM-T Easy and sequence-verified. The subsequently isolated 1107 bp *Eco*RI/*Bam*HI fragment was

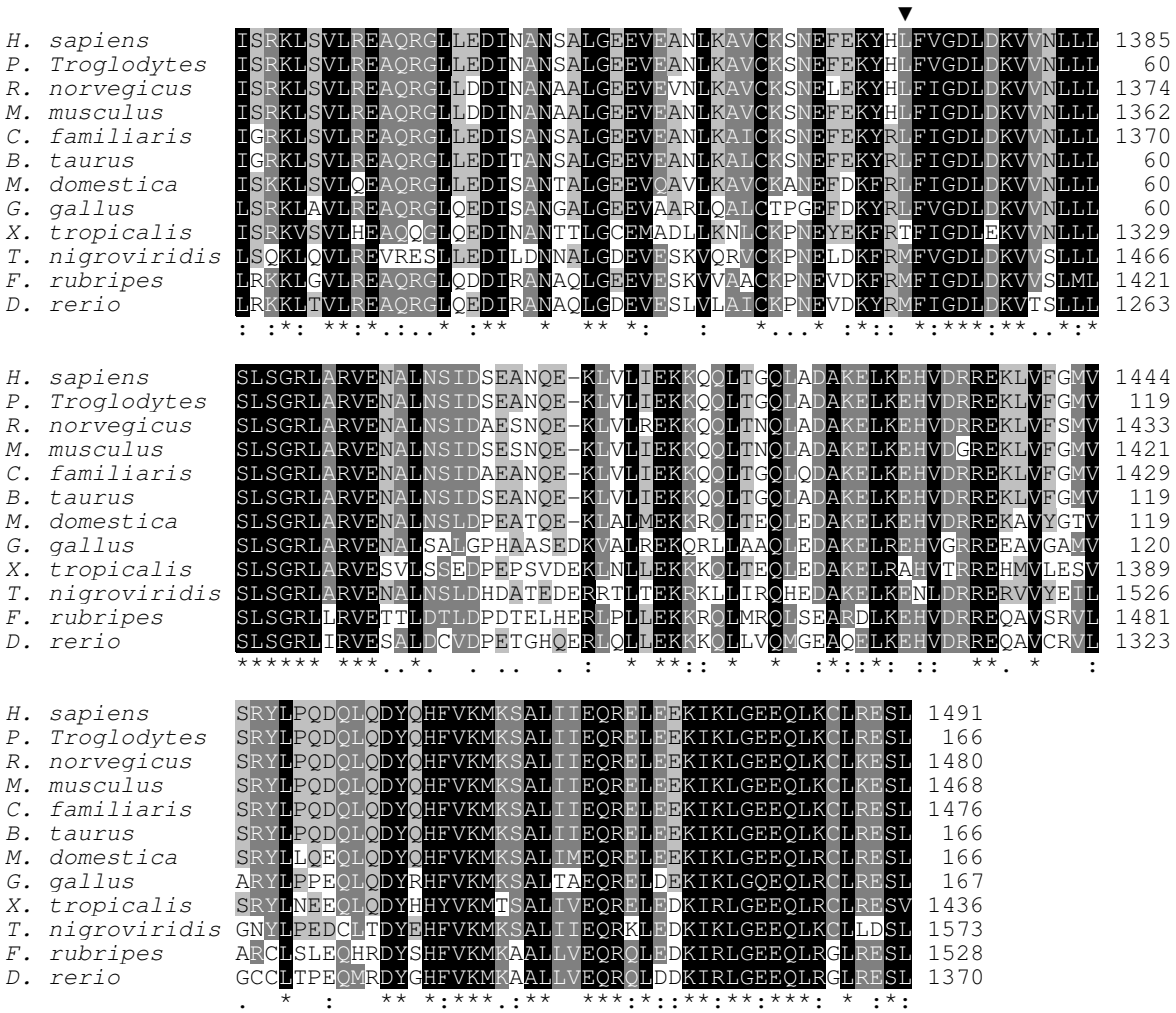
ligated to the 3100 bp *Bam*HI/*Cla*I fragment from the *hKIAA1202* cDNA. The resulting *Eco*RI/*Cla*I fragment was introduced into the pBluescript II KS+ cloning vector. Next, a 326 bp fragment encompassing a *Cla*I site was PCR-amplified from the *hKIAA1202* cDNA, introducing a *Sal*I restriction site behind the last codon. This amplicon was TA-cloned into the pGEM-T Easy vector and sequence-verified prior to isolation of the 296 bp *Cla*I/*Sal*I fragment. This fragment was fused to the *Eco*RI/*Cla*I fragment in the pBluescript II KS+ vector, resulting in the 4494 bp hKIAA1202 ORF I. This construct was then used as a starting point to generate a set of differently tagged versions of the hKIAA1202 ORF I.

- The complete ORF I was subcloned between the *Eco*RI and *Sal*I sites of the pEGFP-N3 and pEGFP-C1 mammalian expression vectors in order to tag the ORF with C- and N-terminal EGFP, respectively.
- The ORF was cut from the pEGFP-N3 vector and ligated into the pcDNA/V5-HisB vector using the *Hin*dIII and *Kpn*I restriction sites from the pEGFP-N3 MCS, allowing the addition of a C-terminal V5-tag to the ORF.
- To tag the ORF with an N-terminal HA-tag, the ORF was shuttled into the pTL1-HA3 vector using the *Hin*dIII and *Kpn*I restriction sites.
- A C-terminal HA-tag was introduced by exchanging it with the V5-tag from the hKIAA1202-V5 construct using *Not*I and *Age*I restriction sites. The HA-tag itself was PCR-amplified from the pTL1-HA3 vector, introducing *Not*I and *Age*I restriction sites, TA-cloned into pGEM-T Easy and sequence-verified.
- Using pcDNA/V5-HisB as a template, the MCS was PCR-amplified, introducing an *Nhe*I restriction site and a FLAG-tag 5' of the MCS, TA-cloned into pGEM-T Easy and sequence-verified. The MCS of the pcDNA/V5-HisB vector was then exchanged for the modified MCS using *Nhe*I and *Sac*II endonucleases. Similarly, a pcDNA/V5-HisB vector with a cMyc-tag upstream of the MCS was engineered. In both modified vectors, the hKIAA1202 ORF was inserted between the *Hin*dIII and *Kpn*I restriction sites.

Fig. III-26 | *Next page.* **Cross-species global multiple sequence alignment of APX Shroom domains 2 from hKIAA1202 homologues.**

a. Apart from *P. Troglodytes*, *B. taurus*, *M. domestica* and *G. gallus*, where the entire Kiaa1202 homologue is unknown, numbers refer to positions in the respective Kiaa1202 homologues. An arrowhead indicates the p.L1372F exchange. An asterisk marks identities, a colon conserved substitutions and a dot semi-conserved substitutions. Shades of grey highlight similarities within subsets of sequences.
b. Sequence identity matrix for the APX Shroom domain 2 (ASD2) of several different hKIAA1202 homologues. Pt, *P. Troglodytes*; Bt, *B. taurus*; Gg, *G. gallus*; Tn, *T. nigroviridis*; Dr, *D. rerio*; all other abbreviations as in Fig. III-24b.

a



b

Sequence Identity Matrix for ASD2 Domains

	Hs	Pt	Rn	Mmu	Cf	Bt	Md	Gg	Xt	Tn	Fr	Dr
Hs	100	100	93.3	95.7	95.7	96.9	84.3	71.8	67.6	64.6	59.2	59.8
Pt		100	93.3	95.7	95.7	96.9	84.3	71.8	67.6	64.6	59.2	59.8
Rn			100	96.3	91.5	91.5	81.3	68.8	67.0	63.4	60.4	59.8
Mmu				100	92.7	93.9	82.5	68.8	66.4	62.2	59.8	59.2
Cf					100	97.5	84.9	71.8	67.6	63.4	60.4	61.6
Bt						100	84.3	71.8	68.2	63.4	60.4	61.0
Md							100	74.8	71.2	64.6	64.6	62.8
Gg								100	63.4	59.8	59.8	59.8
Xt									100	56.2	58.6	59.8
Tn										100	59.2	59.2
Fr											100	79.6
Dr												100

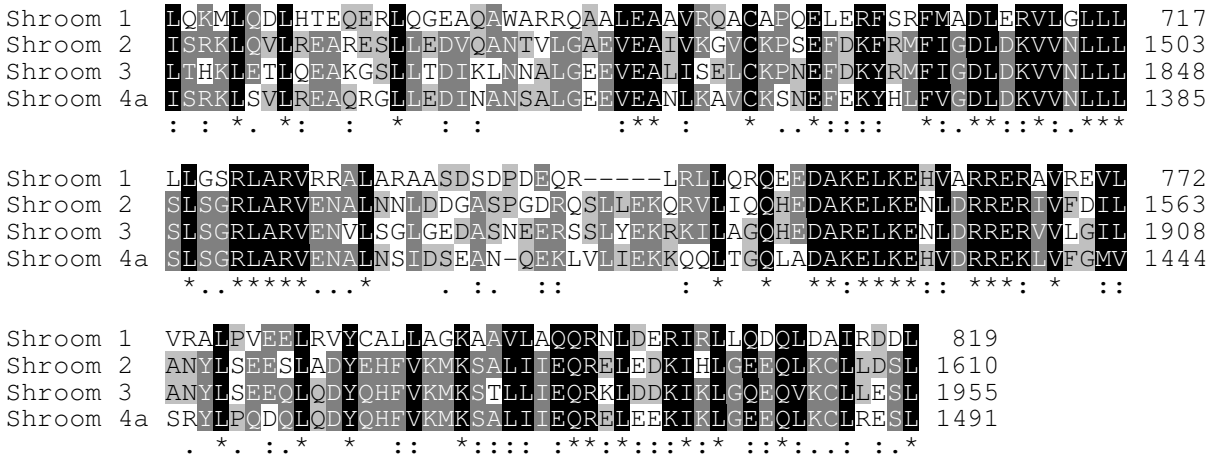
a

Sequence Identity Matrix for hShroom PDZ Domains			
	2	**3**	**4a**
2	100	67.5	61.0
3		100	63.6
4a			100

b

Sequence Identity Matrix for hShroom ASD1 Domains			
	1	**2**	**3**
1	100	32.9	29.8
2		100	44.6
3			100

c



Sequence Identity Matrix for hShroom ASD2 Domains				
	1	**2**	**3**	**4a**
1	100	37.7	32.9	35.3
2		100	68.2	65.8
3			100	61.6
4a				100

Immunoblotting of lysates of cells over-expressing tagged forms of hKIAA1202, followed by detection with antibodies raised against the tags, resulted in bands of approximately 220 kDa (Fig. III-29a). These bands were also recognised by an α-hKIAA1202 antibody (see III.C.3.2.2). The calculated MW of the proteins translated from the *hKIAA1202* constructs is ~170 kDa, suggesting post-translational modification(s) of hKIAA1202. A similar difference between predicted and actual MW has recently been observed for mApxl[782], another member of the Apx/Shrm protein family. Glycosylation affecting protein mobility in SDS-PAGE gels by > 200% have been reported[783]. But much smaller post-translational modifications, such as phosphorylation, apparently increasing MW by > 20% are also known[609].

C.3.2. Generation and characterisation of an α-hKIAA1202 antibody

To investigate endogenous rather than over-expressed hKIAA1202, we generated and characterised a peptide α-hKIAA1202 antibody.

C.3.2.1. Generation of an α-hKIAA1202 antibody

A polyclonal antibody raised against a hKIAA1202 peptide was generated.

In an attempt to investigate the endogenous hKIAA1202 protein, we generated an α-hKIAA1202 antibody. We selected a 28 AA antigenic peptide (CLENPALDLSSYRAIS-SLDLLGDFKHAL, hKIAA1202 AA 1004 – 1031) based on investigation of the complexity, helicity and hydropathy of the hKIAA1202 sequence, aided by antigenicity prediction software, and expertise (courtesy of Dr. R. Schneider, University of Innsbruck, Innsbruck, Austria). Moreover, this sequence is well conserved in rodents (Fig. III-24a), which may be of use in subsequent cytological experiments, as some of the most commonly used cell lines in molecular biology are derived from mouse and rat. The peptide was synthesised commercially, coupled to LPH and used for rabbit immunisation. At regular intervals, the titre of the antisera was measured using dot blotting and ELISA. BioGenes, the company executing the

Fig. III-27 | *Previous page.* **Multiple sequence alignments of domains found in APX/Shroom family members.**

Multiple sequence alignments and sequence identity matrices are presented for PSD-95/Dlg/ZO-1 domains (PDZ, panel a), and APX Shroom domains 1 (ASD1, panel b) and 2 (ASD2, panel c) of APX/Shroom family members. To avoid evolution-based dissimilarity, sequences of the human proteins are used in the analysis (hShroom 1, hAPXL2; hShroom 2, hAPXL; hShroom 3, hShroom; hShroom 4a, hKIAA1202; see Table III-10). hAPXL2 does not contain a PDZ domain and an ASD1 is absent from hKIAA1202. Numbers refer to positions in the respective protein sequence. An asterisk marks identities, a colon conserved substitutions and a dot semi-conserved substitutions. Shades of grey highlight similarities within subsets of sequences.

immunisation protocol, also performed the ELISA. While the antiserum titre reached 1: > 200000 after immunisation and four booster injections over a period of 40 days, the pre-immune serum had a titre of 1:300, as assayed by ELISA (Fig. III-28a). Dot blotting qualitatively confirmed the ELISA analysis; whereas a 1:50000 antiserum dilution readily recognised the spotted peptide, no signal was obtained with a 1:100 dilution of the pre-immune serum (Fig. III-28b).

After completion of the immunisation protocol, we collected the serum and isolated the α-hKIAA1202 antibody by immunopurification. To this end, we bound our antigenic peptide to a column via sulfhydryl chemistry using its N-terminal cysteine residue. Elution was monitored by dot blotting. Aliquots of elution fractions with the highest titre were stored at 4°C, -20°C and -80°C. After a week, they were assayed by dot blot analysis to determine the most appropriate storage conditions. As no differences were observed (Fig. III-28c), we preferred -80°C for long-term storage, keeping working stocks at 4°C.

C.3.2.2. Characterisation of the α-hKIAA1202 antibody

The α-KIAA1202 antibody specifically recognises several protein species in western blot applications.

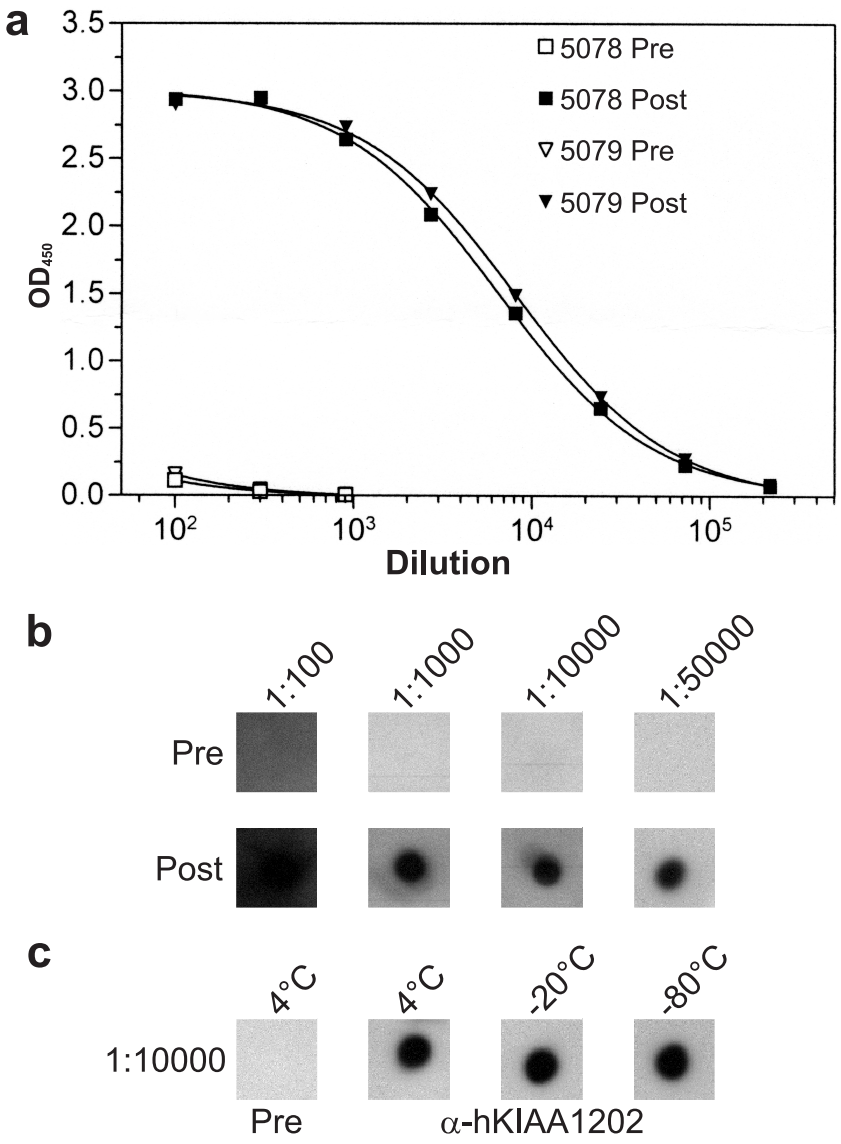
Elution fractions with the highest α-hKIAA1202 titre were used in a series of experiments to characterise the antibody for use in western hybridisation. Consecutive α-V5 and α-hKIAA1202 immunodetection of size-separated and transferred lysates from HeLa cells over-expressing V5-tagged hKIAA1202 isoform I showed that α-KIAA1202 specifically recognises over-expressed hKIAA1202 (Fig. III-29a).

Lysates of several mammalian cell lines were investigated for expression of hKIAA1202 protein. A number of protein species with MWs between > 220 – ~43 kDa were observed upon α-hKIAA1202 immunodetection (Fig. III-29b). All but a ~47 kDa fibroblast-specific

Fig. III-28 | *Next page.* **Titration of an α-hKIAA1202 antibody.**

a. Titration using an enzyme-linked immunosorbent assay measuring absorbance at 450 nm (OD_{450}) shows that dilutions of > 10^5 of post-immune sera (Post) of two injected rabbits (5078 and 5079) still recognise the antigen, whereas dilutions of > 10^2 of pre-immune sera (Pre) do not.
b. Spotted antigen, diluted as indicated at the top, is incubated with pre-immune (Pre) and post-immune (Post) sera, and detected with horseradish peroxidase-conjugated secondary antibodies. Results for both rabbits were essentially identical; those for rabbit 5079 are shown.
c. Same experimental procedure as in panel b, except that immunopurified antibody is used instead of post-immune sera. Prior to hybridisation of the spotted antigen, antibody aliquots were stored for one week at 4°C, -20°C or -80°C. No observable differences were apparent. Hybridisation with pre-immune serum (Pre) is included as a control.

Graph courtesy of BioGenes, Berlin, Germany (a).



band were undetectable after immunodetection with (i) pre-immune sera (Fig. III-29c), (ii) α-hKIAA1202 antibody saturated with antigen (Fig. III-29d) and (iii) secondary antibody alone. In addition, ten µg of LPH adjuvant was not recognised by the α-hKIAA1202 antibody (Fig. III-29e). Taken together, these observations could suggest that all protein species identified with the α-hKIAA1202 antibody, apart from the fibroblast-specific ~47 kDa band, are Kiaa1202 isoforms. Densitometric analysis indicated that the ~43 kDa band is > ten-fold more intense than this ~47 kDa band.

As would be expected from the sequence homology between the human antigen and the corresponding mouse sequence (78.6% ID and 96.4% similarity, Fig. III-24a), α-hKIAA1202 specifically recognises several protein species in the murine Neuro-2A neuroblastoma cell

line. The band migrating at ~200 kDa in human cell lines runs at a slightly larger MW in Neuro-2A cells (Fig. III-29b).

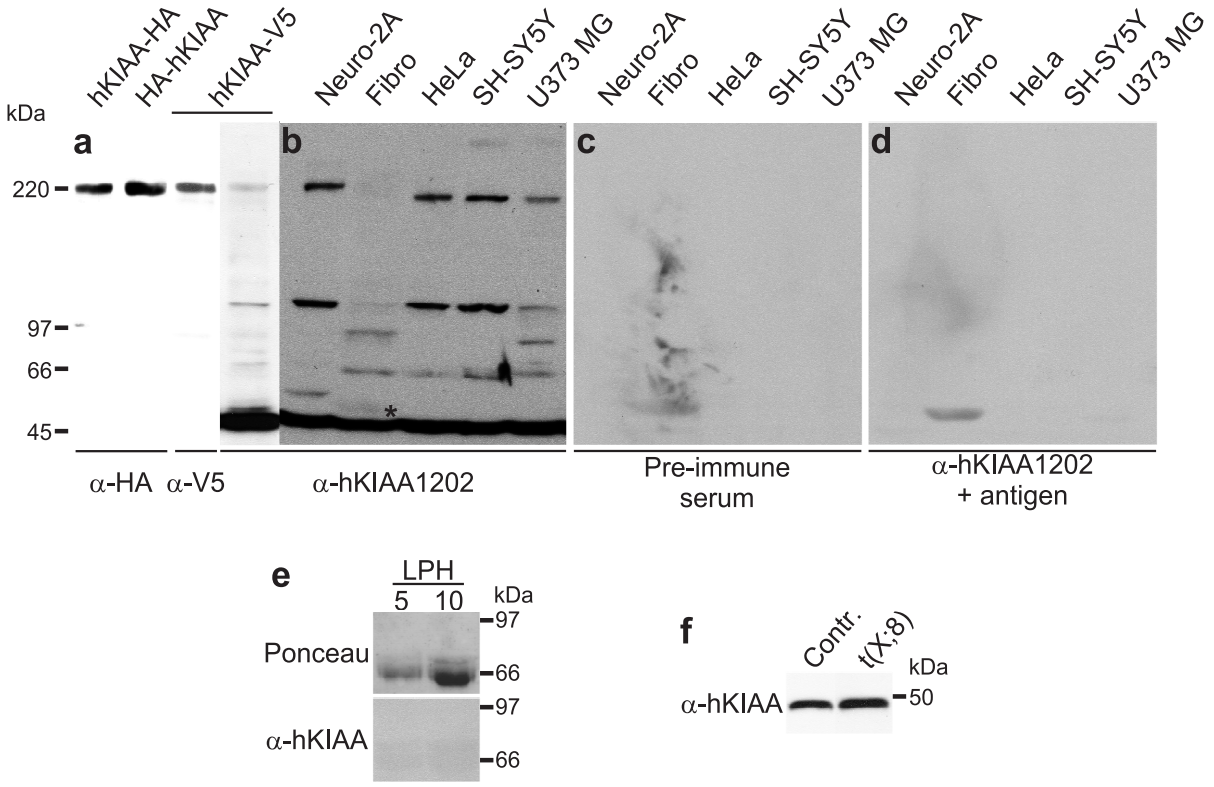


Fig. III-29 | **α-hKIAA1202 specifically recognises several protein species in human and mouse cells.**

a. Tagged versions of hKIAA1202 are over-expressed in HeLa cells and the lysates size-separated, transferred and hybridised with α-HA or α-V5 antibodies. The hKIAA1202-V5 membrane is subsequently incubated with α-hKIAA1202 antibody.
b. Lysates from different cell lines, as indicated at the top, are size-separated, transferred and incubated with α-hKIAA1202 antibody. 'Fibro', lysate from primary fibroblasts. The asterisk points out a non-specific band, only seen in fibroblasts. The ~220 kDa isoform from murine Neuro-2A cells migrates slightly slower compared to that from the human HeLa, SH-SY5Y and U373 MG cell lines.
c, d. α-hKIAA1202 signals from the membrane depicted in panel b are removed and the membrane is subsequently incubated with pre-immune serum (panel c) or α-hKIAA1202 antibody that is pre-associated with antigen prior to hybridisation (panel d). Only a single band in fibroblasts (asterisk in panel b) is apparent.
e. Five and ten micrograms of *L. polyphemus* hemocyanin (LPH) are loaded onto a denaturing polyacrylamide gel, electrophoresed, transferred and stained with Ponceau S Staining Solution (Ponceau, top panel). After removal of the staining, the membrane is hybridised with α-hKIAA1202 antibody (α-hKIAA, bottom panel). HeLa lysate (not shown) is included as a control.
f. After α-hKIAA1202 hybridisation of size-separated lysates from the lymphoblastoid cell lines employed in Fig. III-21e, the 43 kDa isoform becomes apparent. This result is representative for four control cell lines. Whether other protein species are also expressed in these cell lines needs further investigation.

α-hKIAA1202 immunoblotting of lysates from those lymphoblastoid cell lines used in the *hKIAA1202* expression analysis (Fig. III-21e), including the patient cell line, yielded the

~43 kDa band. Whether other protein species are also expressed in lymphoblastoid cell lines needs further delineation.

C.3.3. Subcellular localisation of Kiaa1202

We studied the cellular distribution of over-expressed and endogenous Kiaa1202 by IF and subsequent confocal microscopy.

C.3.3.1. Subcellular localisation of over-expressed hKIAA1202-V5

Over-expressed hKIAA1202-V5 localises to the cytosolic side of the cell membrane in mammalian cells.

To investigate the subcellular localisation of the hKIAA1202 protein, a tagged version of hKIAA1202 ORF I was expressed in mammalian cell lines. To exclude a possible cell-specific effect on the localisation of hKIAA1202, we transfected HeLa, COS-7, SH-SY5Y, Neuro-2A and U373 MG cells with all tested expression constructs. To rule out an influence of an N- versus C-terminal tag, we fused hKIAA1202 N- and C-terminally with EGFP. We also tagged hKIAA1202 with a small C-terminal V5 tag to circumvent possible steric hindrance of the bulky EGFP. Finally, we performed live imaging of HeLa cells transiently expressing EGFP-hKIAA1202 and hKIAA1202-EGFP to circumvent any artefacts that might arise from methodologies inherent to the IF protocol, such as cell permeabilisation and fixation. In all cells and for all expression constructs, we observed a localisation of over-expressed hKIAA1202 confined to small, localised patches at the cell membrane (Fig. III-30).

To distinguish between TM topology and submembranous localisation, we subjected the hKIAA1202 sequence to *in silico* analysis. None of six TM prediction algorithms (TMpred, PRED-TMR, Split35, TMHMM, TMAP and Moment) indicated a TM region. Next, we plotted the hydropathy profile of three ion channels involved in human disease (SCN1B, CFTR, ATPase6; Table III-11) according to Kyte and Doolittle[726] and compared the graphs with that of hKIAA1202. While the ion channels have between two and six regions surpassing the 1.6 TM threshold, hKIAA1202 has a maximal hydropathy score of 1.0 (Fig. III-31a). We also compared percentages of positions with a hydropathy score > 0 in the three ion channels with that of hKIAA1202 and that of the average for 18 ion channels implicated in human pathologies (ACCN1, ATPase6, CACNA1A, CACNA1C, CACNA1D, CACNA1F, CACNA1S, CFTR, CNGA1, CNGB3, HCN1, KCNA1, SCN1A, SCN1B, SCN3A, SCN4A, SCN5A and SCN7A). Whereas only 7.2% of positions in hKIAA1202 have a positive hydropathy score,

the mean value for positions from the 18 ion channels is 45.5%, ranging from 33.8% (HCN1) to 86.0% (ATPase6) (Fig. III-31b).

Table III-11 | **Overview of ion channels linked to human disease that were subjected to *in silico* analysis**

Ion channel	Gates	Associated disease	Positions with hydropathy scores > 0 (%)	Ref.
SCN1B	Na^+	GEFS+	34.5	784
CFTR	Cl^-	Cystic fibrosis	48.3	785
ATPase6	H^+	MR and ataxia	86.0	786

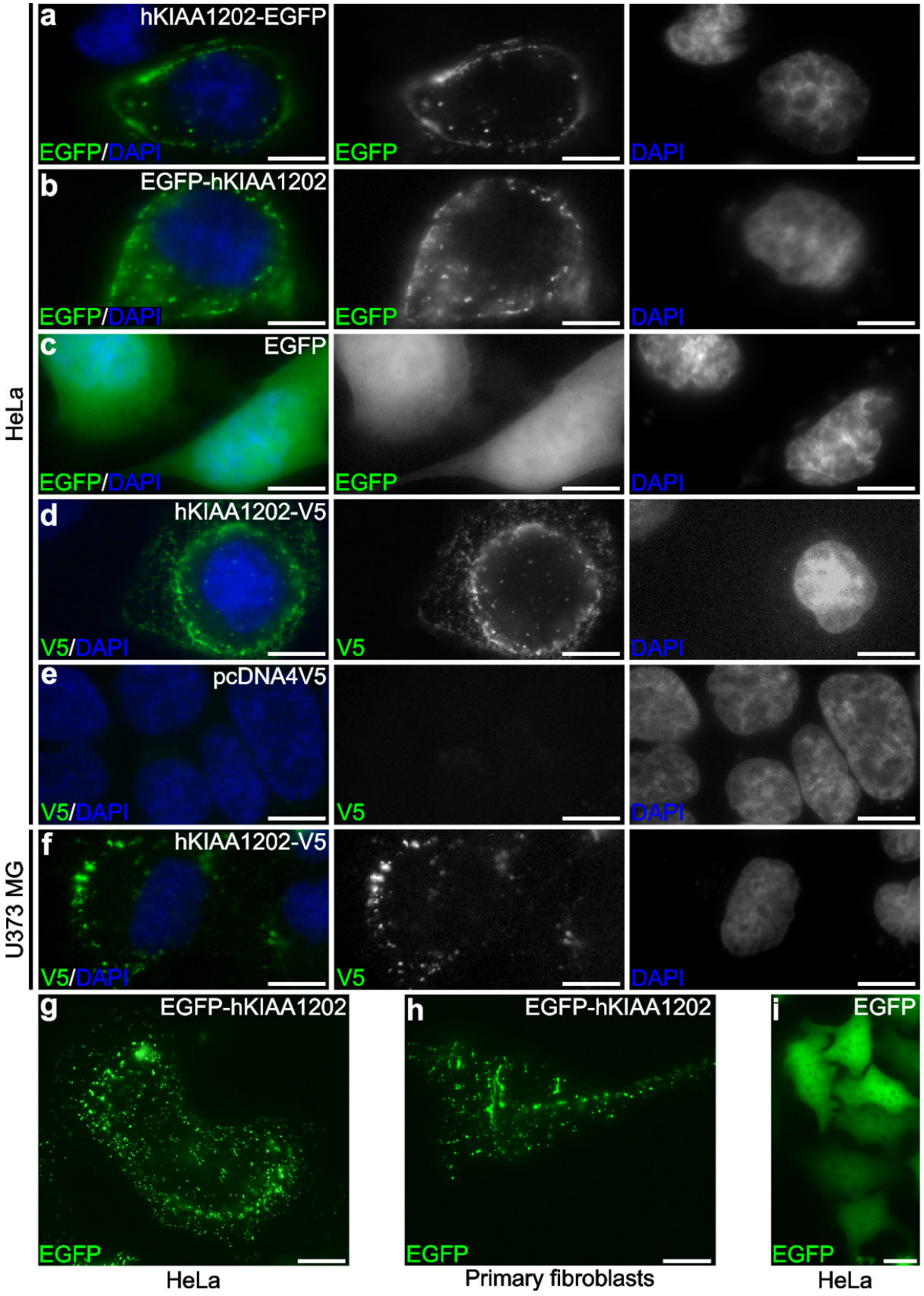
Experimental evidence supports the outcome of our computational analyses. Over-expressed hKIAA1202-V5 does not co-localise with Golgin (Fig. III-31c), a protein specific to the Golgi apparatus[787]. TM proteins and proteins of the extracellular matrix are known to be processed through the Golgi apparatus[788]. In addition, we showed partial co-localisation of hKIAA1202-V5 with F-actin in HeLa cells (see III.C.3.4.2.1).

Taken together, we conclude that over-expressed hKIAA1202 localises on the cytosolic side of the cell membrane.

Especially when proteins seem to be confined to specific areas in the cell, subcellular localisation of a higher resolution is required, as it may provide insights into the protein's cellular function and, in the case of hKIAA1202, would provide a selection criteria for choosing which of the putative interaction partners recovered in a Y2H screen (see III.C.3.4.3) would be subjected to further, time- and cost-intensive functional analyses. To address this requirement, EM was performed on cells over-expressing hKIAA1202-V5 (collaboration with Dr. R. Lurz and G. Lüder, MPI-MG, Berlin, Germany). Cells were embedded in LR White and LR Gold. While the former displayed extensive polymerisation shrinkage rupturing most of the cells, the latter maintained excellent cell morphology. Cells were incubated with monoclonal α-V5 (1:50 – 1:100), polyclonal α-V5 (1:20 – 1:300) and α-hKIAA1202 (1:3 – 1:10) antibodies.

Fig. III-30 | *Next page.* **Subcellular localisation of hKIAA1202 open reading frame I in mammalian cells.**

Indicated tagged versions of hKIAA1202 open reading frame I and expression vectors are transiently expressed in HeLa cells (panels a – e, g and i), U373 MG cells (panel f) and primary fibroblasts (panel h). Cells are subjected to permeabilisation and fixation prior to recording using direct fluorescence (enhanced green fluorescent protein, EGFP; panel a – c) or indirect immunofluorescence (V5; panel d – f), or they are visualised by live imaging (panels g – i). DNA is counterstained with 4'6-diamidino-2-phenylindole·2 HCl (DAPI). No difference in subcellular localisation is observed between N- and C-terminally tagged constructs, EGFP- and V5-tagged clones, and transfection in HeLa or U373 MG cells. In addition, the live imaging experiments show that the punctuated expression pattern is not an artefact of the (immuno)fluorescence protocol. Scale bars, 5 μm (panels a – f) or 20 μm (panels g – i).



Unfortunately, no significant difference between cells transfected with empty vector or V5-tagged *hKIAA1202* was observed in any of these experiments.

C.3.3.2. Testing of the α-hKIAA1202 antibody for use in immunofluorescence

The α-hKIAA1202 antibody identifies over-expressed hKIAA1202-V5 in IF.

To verify that α-hKIAA1202 could recognise the native full-length hKIAA1202 protein in IF, we transfected HeLa cells with hKIAA1202-V5, and performed IF with α-V5 and α-hKIAA1202 antibodies simultaneously. The V5-signal showed a complete overlap with the α-hKIAA1202 staining. Prior saturation of the α-KIAA1202 antibody with its antigen abolished the α-hKIAA1202 signal, but retained that of the α-V5 antibody (Fig. III-32). These experiments indicate that the α-hKIAA1202 antibody recognises over-expressed hKIAA1202 in IF applications. It should be noted that additional α-hKIAA1202 staining was observed which did not correspond to α-V5 signal, likely reflecting expression of endogenous hKIAA1202. To make sure that the observed V5 – hKIAA1202 co-localisation was not due to antibody cross-reactivity, all controls, as detailed under IIB.A.6.6.2, were performed. No cross-reactivity was apparent. Moreover, cross-channel bleeding of Cy3 and FITC fluorescence, putatively resulting in artifactual co-localisation signals, was eliminated as described under IIB.A.6.6.2.

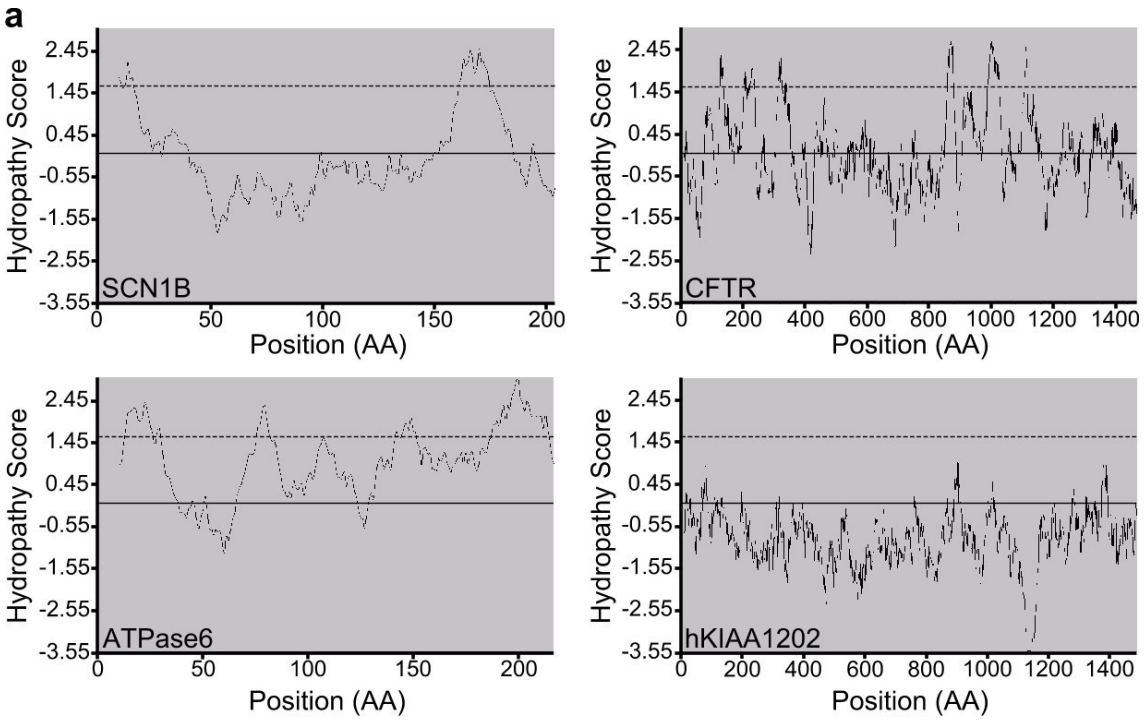
C.3.3.3. Subcellular localisation of endogenous Kiaa1202

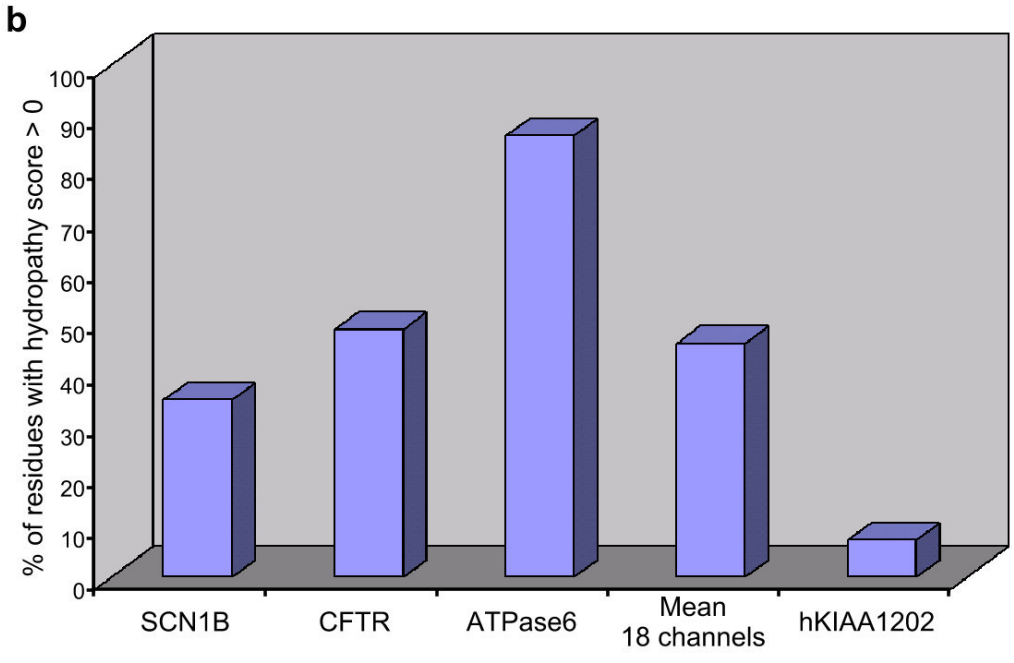
Endogenous α-hKIAA1202 shows a general cytoplasmic localisation and, in addition, localises to the leading edge of fibroblasts and to the neurites of neuronal cells.

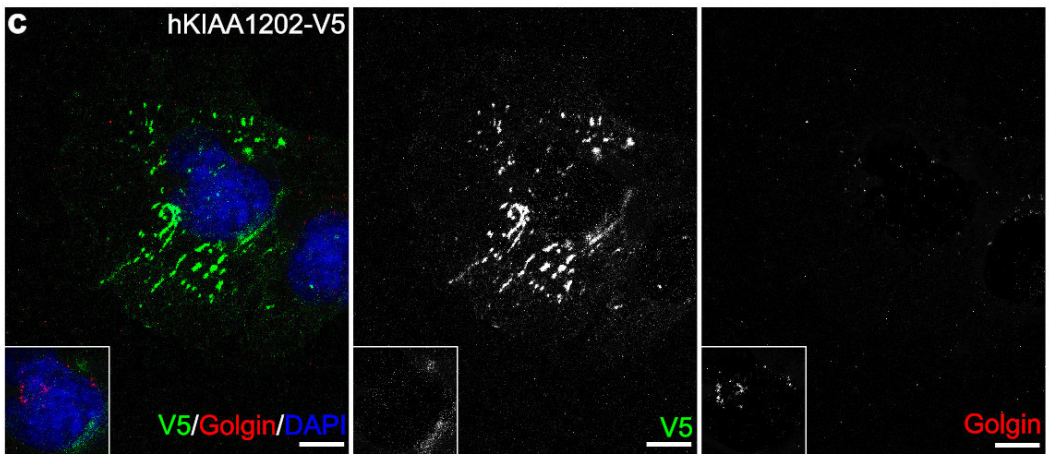
To investigate the subcellular localisation of endogenous Kiaa1202, we performed α-hKIAA1202 IF on several different mammalian cell lines. The endogenous protein is expressed throughout the cytoplasm. Specific staining was also detected in the neurites of the neuronal cell lines SH-SY5Y, Neuro-2A and U373 MG. Moreover, an α-hKIAA1202 signal was observed in the leading edge of human primary fibroblasts (Fig. III-33).The discrepancy in subcellular localisation between over-expressed and endogenous hKIAA1202 is probably

Fig. III-31 | *Next page.* **hKIAA1202 is not a transmembrane protein.**

a. Kyte – Doolittle hydropathy plots of three ion channels [sodium channel, voltage-gated, type I, β-subunit (SCN1B), cystic fibrosis transmembrane (TM) conductance regulator (CFTR) and ATPase6] and hKIAA1202 show that the latter has no regions surpassing the 1.6 TM threshold (dashed line). A solid line denotes a hydropathy score of 0. The scanning window spanned 19 amino acids (AA).
b. The normalised number of residues with a hydropathy score > 0 is plotted for SCN1B, CFTR, ATPase6, the mean of 18 ion channels (see text) and hKIAA1202.
c. Confocal recording of C-terminally V5-tagged hKIAA1202 open reading frame I in HeLa cells. Simultaneous detection of hKIAA1202-V5 (α-V5 antibody) and the Golgi apparatus (α-Golgin antibody) reveals that hKIAA1202 does not co-localise with the Golgi apparatus, suggesting that it is not a TM protein. The inset shows part of the focal plane 3 μm below that of the main image. DNA is counterstained with 4'6-diamidino-2-phenylindole·2 HCl (DAPI). Scale bar, 10 μm.


a
Hydropathy Score
2.45
1.45
0.45
-0.55
-1.55
-2.55
-3.55
SCN1B
0
50
100
150
200
Position (AA)
CFTR
0
200
400
600
800
1000
1200
1400
ATPase6
hKIAA1202


b
% of residues with hydropathy score > 0
100
90
80
70
60
50
40
30
20
10
0
SCN1B
CFTR
ATPase6
Mean
18 channels
hKIAA1202


c
hKIAA1202-V5
V5/Golgin/DAPI
V5
Golgin

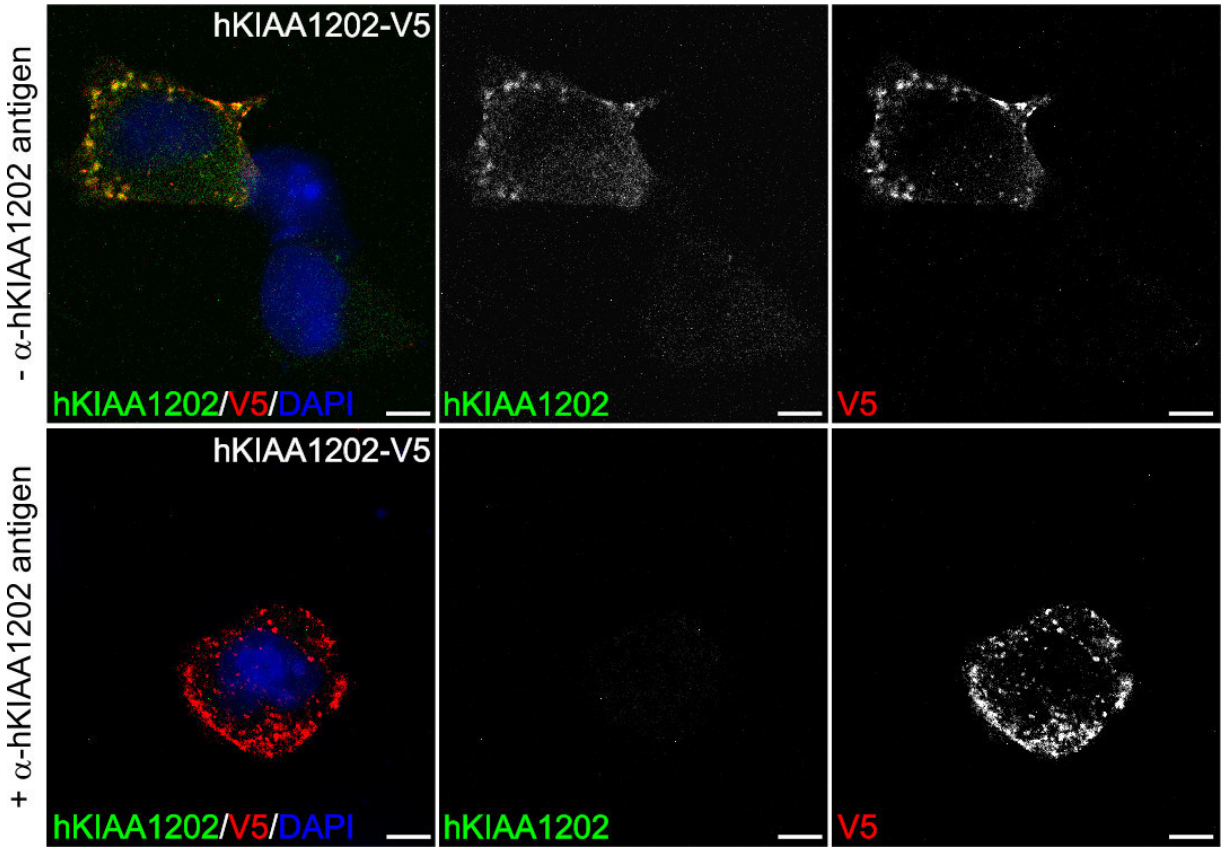


Fig. III-32 | **α-hKIAA1202 recognises native hKIAA1202-V5.**

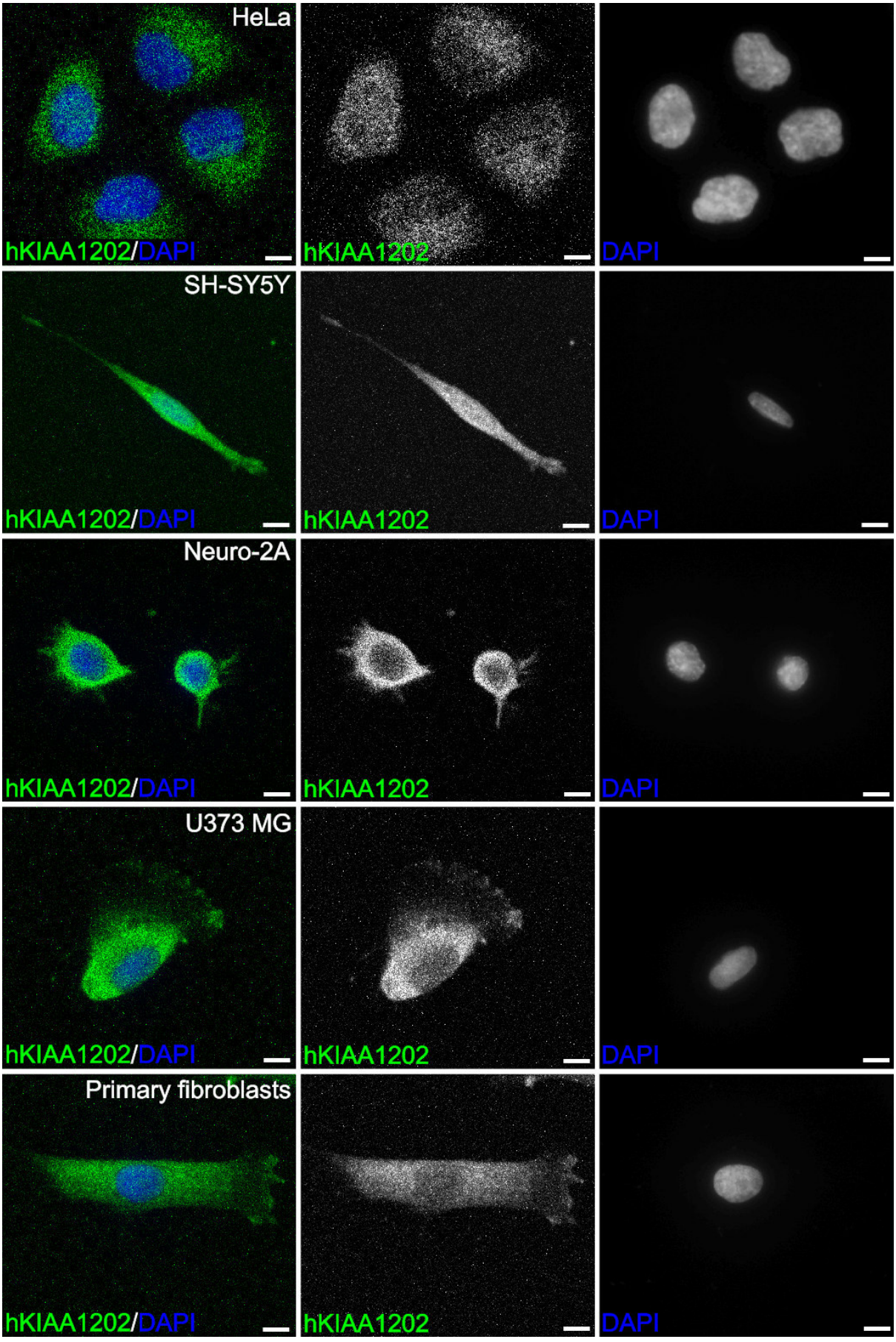
Without prior incubation of the α-hKIAA1202 antibody with its antigen, simultaneous detection of hKIAA1202-V5 in HeLa cells with α-hKIAA1202 and α-V5 antibodies results in overlapping signals (top panel). Non-overlapping α-hKIAA1202 signals probably derive from endogenous hKIAA1202. When the α-hKIAA1202 antibody is pre-associated with antigen prior to immunofluorescence, only the α-V5 antibody recognises hKIAA1202-V5, demonstrating the specificity of the α-hKIAA1202 antibody (bottom panel). DNA is counterstained with 4'6-diamidino-2-phenylindole·2 HCl (DAPI). Scale bars, 10 μm.

due to the fact that the α-hKIAA1202 antibody recognises multiple protein species. In addition, the protein isoform corresponding to ORF I, used in the over-expression study, is expressed at significantly lower levels as some of the other protein species (Fig. III-29b).

Control experiments validating these data were performed as described under IIB.A.6.3.2. As with over-expressed hKIAA1202-V5, different attempts to refine the localisation of endogenous hKIAA1202 using EM were unsuccessful (see III.C.3.3.1).

Fig. III-33 | *Next page.* **Subcellular localisation of endogenous Kiaa1202.**

The image shows stacked confocal recordings of indicated cells stained with α-hKIAA1202 antibody. DNA is counterstained with 4'6-diamidino-2-phenylindole·2 HCl (DAPI). Scale bars, 10 μm.


HeLa
hKIAA1202/DAPI
hKIAA1202
DAPI
SH-SY5Y
hKIAA1202/DAPI
hKIAA1202
DAPI
Neuro-2A
hKIAA1202/DAPI
hKIAA1202
DAPI
U373 MG
hKIAA1202/DAPI
hKIAA1202
DAPI
Primary fibroblasts
hKIAA1202/DAPI
hKIAA1202
DAPI

C.3.4. Investigation of putative hKIAA1202 interactions

To investigate hKIAA1202's molecular environment, we conducted a series of co-localisation and co-IP studies, as well as a relocalisation experiment, and a Y2H screen. To confirm the obtained results, we established a cell line stably expressing tagged hKIAA1202.

C.3.4.1. Generation of a cell line stably expressing hKIAA1202-V5

Collaboration with J. Ruschmann, MPI-MG, Berlin, Germany.

A U373 MG glioblastoma astrocytoma cell line, stably expressing V5-tagged hKIAA1202, was established to confirm putative hKIAA1202 interactors.

To verify interaction of hKIAA1202 with several of the putative binding partners that would be recovered in a Y2H screen, we set out to establish a human cell line that stably expresses a tagged version of hKIAA1202. We reasoned that a cell line endogenously expressing hKIAA1202 contains the molecular machinery to regulate expression of tagged hKIAA1202. Moreover, such a cell line is likely to express all hKIAA1202 interaction partners. In this way, some of the advantages of over-expression studies, such as the possibility to use commercially available antibodies, were retained without introducing over-expression-induced artefacts. Western hybridisation experiments showed that the human glioblastoma astrocytoma cell line U373 MG expresses hKIAA1202 endogenously (Fig. III-29b). Therefore, we established a monoclonal U373 MG cell line stably expressing hKIAA1202-V5.

To secure intact hKIAA1202-V5 insertion into the U373 MG genome, we linearised the hKIAA1202-V5 construct using *Bgl*II. This endonuclease digests the vector upstream of the CMV promoter driving hKIAA1202-V5 expression. As a control, we transfected similarly linearised empty pcDNA4/V5-HisB vector. We employed the Zeocin resistance gene present in the pcDNA4/V5-HisB vector as a marker to select stably transfected cells. We generated monoclonal cell lines by picking single cell clones and expanding them under Zeocin selective pressure. Apart from the occasional enlarged cell due to lack of Zeocin resistance, cells showed a normal morphology (Fig. III-34a). Cells expressing hKIAA1202-V5 grew approximately 50% slower than those transfected with empty vector or those that were not transfected. PCR reactions showed an integration of the hKIAA1202-V5 construct and empty vector DNA in the U373 MG genome (Fig. III-34b). Immunoblotting of lysates of such U373 MG cell lines, followed by α-V5 immunodetection, showed a specific band of the expected size in the hKIAA1202-V5 cell line, but not in the vector control (Fig. III-34c).

C.3.4.2. Putative interaction between Kiaa1202 and filamentous Actin

Three observations point towards a putative interaction between hKIAA1202 and F-actin: distribution of over-expressed hKIAA1202 at the cell membrane, localisation of endogenous Kiaa1202 to the leading edge of fibroblasts and neurites of neuronal cells, and cytoskeletal involvement of members of the Shrm protein family to which hKIAA1202 belongs.

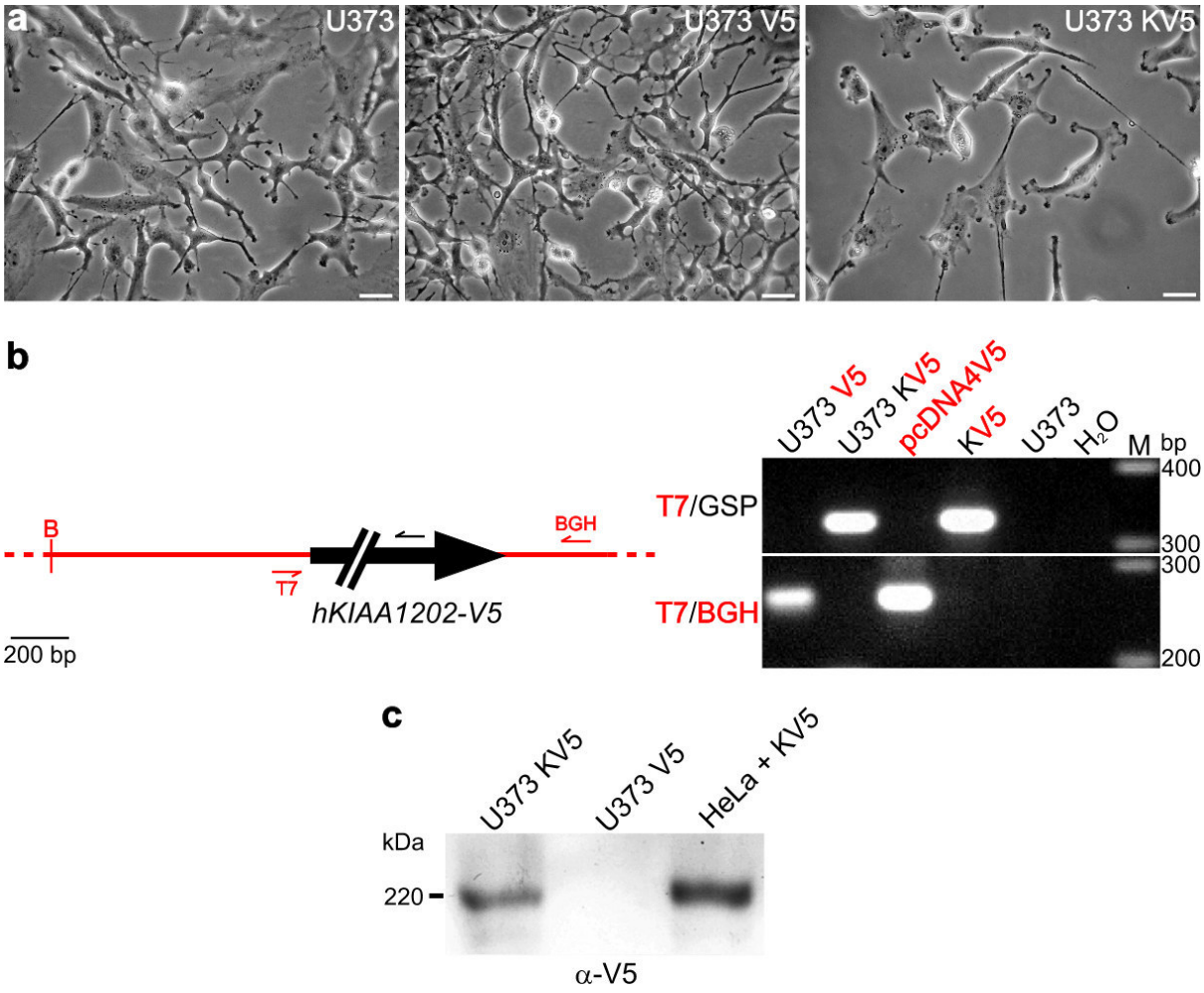


Fig. III-34 | **Generation of a U373 MG cell line stably expressing hKIAA1202-V5.**

a. Morphologies of U373 MG cell lines stably transfected with vector (middle panel, U373 V5) or *hKIAA1202-V5* (right panel, U373 KV5) are not observably different from that of the non-transfected cell line (left panel, U373). Scale bars, 30 μm.
b. The left panel represents a schematic drawing of the *Bgl*II (B) linearised vector (red line) containing the *hKIAA1202-V5* sequence (big black arrow; not drawn to scale, double slash). Red arrows indicate the positions of the vector-specific T7 and BGH primers, the small black arrow the *hKIAA1202* gene-specific primer (GSP). The right panel shows size-separated polymerase chain reaction products obtained with the primer pairs denoted on the left. Templates, specified at the top, are: genomic DNA from the U373 MG cell lines stably transfected with vector (U373 V5) or *hKIAA1202-V5* (U373 KV5), vector DNA (pcDNA4V5), cloned *hKIAA1202-V5* (KV5), genomic U373 MG DNA (U373) and water (H_2O). Size markers (M) on the right.
c. Cell lysates from the stably transfected cells shown in panel a (U373 KV5 and U373 V5) are size-separated and immunoblotted with α-V5 antibody. Lysate from HeLa cells transiently transfected with *hKIAA1202-V5* (HeLa + KV5) is included as a positive control.

C.3.4.2.1. Co-localisation studies of over-expressed hKIAA1202-V5 and filamentous Actin

Over-expressed hKIAA1202-V5 partially co-localises with F-actin.

As described under III.C.2.2, it is likely that hKIAA1202 is a member of the APX/Shrm protein family. For most of the members of this family, an association with the cytoskeleton has been reported[777,781,789,790]. For example, mShrm and mApxl have been shown to bind F-actin[777,782]. Therefore, we investigated the putative co-localisation of F-actin and over-expressed hKIAA1202-V5 in HeLa cells using IF and confocal microscopy. Over-expressed hKIAA1202-V5 showed a partial co-localisation with F-actin, a major component of the cytoskeleton[266] (Fig. III-35a). Upon disruption of F-actin with the Actin polymerisation inhibitor latrunculin B, hKIAA1202-V5 expression significantly overlapped with the localisation of F-actin remnants (Fig. III-35b). Such a co-localisation pattern was not observed after latrunculin B treatment of HeLa cells over-expressing hFBXO4-V5 (Figs. III-35c – d), an FBP reported to be distributed both in the cytoplasm and nucleus, but not at the membrane[765]. These data are consistent with our presumption that hKIAA1202 localises at the cytosolic side of the cell membrane without spanning it (see III.C.3.3.1).

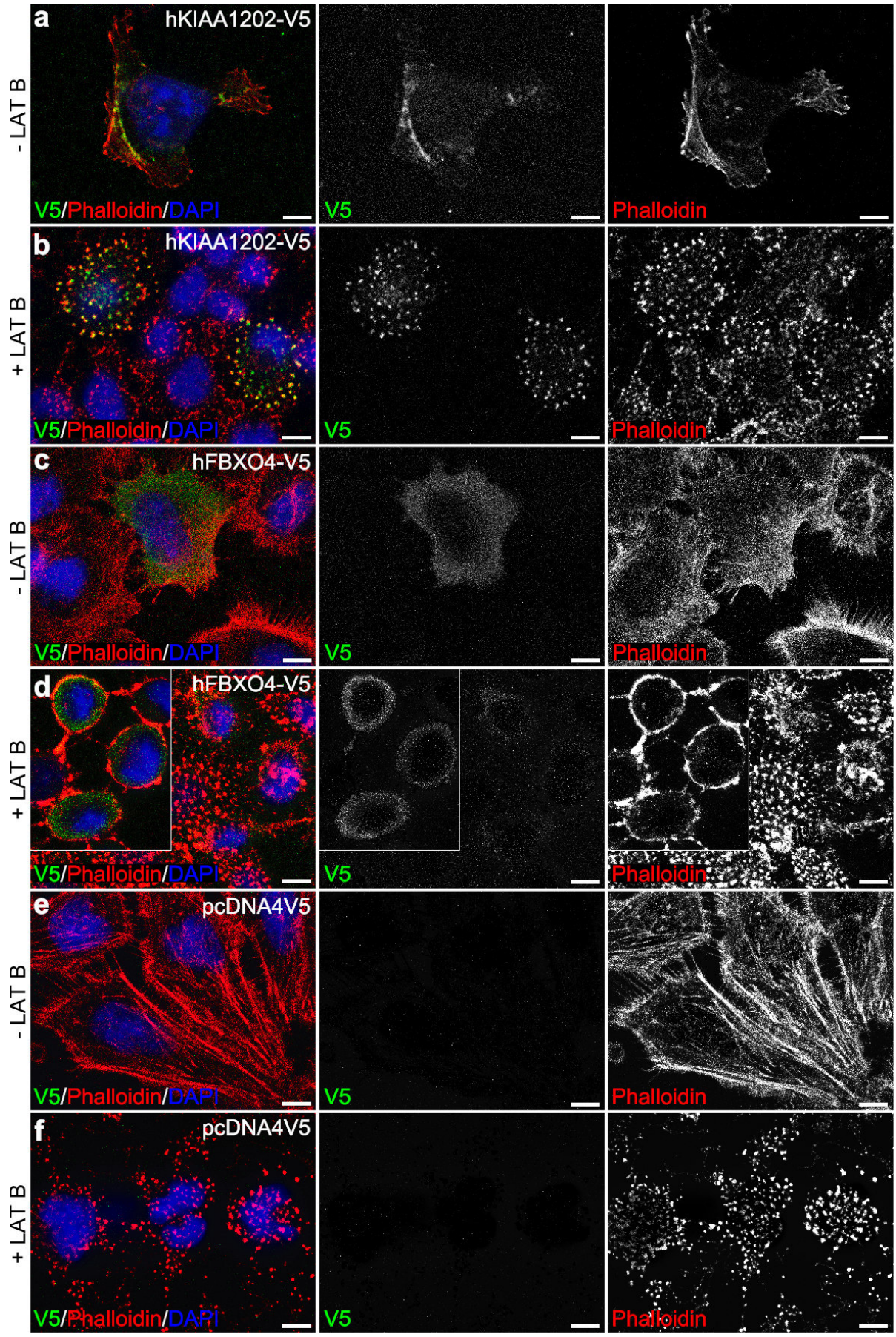
C.3.4.2.2. Co-localisation studies of endogenous Kiaa1202 and filamentous Actin

Endogenous Kiaa1202 co-localises with F-actin in the leading edge of fibroblasts and in neurites of differentiating neuronal cells.

Double staining of several mammalian cell lines with α-hKIAA1202 antibody and TRITC-conjugated phalloidin, an F-actin specific dye, followed by confocal analysis, showed that Kiaa1202 localises with F-actin in the leading edge of primary fibroblasts and in the protrusions of neuronal cell lines (Fig. III-36). As F-actin is essential to the rapid growth

Fig. III-35 | *Next page.* **hKIAA1202-V5 partially co-localises with filamentous Actin.**

Indicated C-terminally V5-tagged constructs (panels a – d) or expression vector (panels e and f) are ectopically expressed in HeLa cells with (panels b, d and f) or without (panels a, c and e) treatment with the Actin polymerisation inhibitor latrunculin B (LAT B) prior to protein fixation. Constructs are detected with α-V5 antibody, filamentous Actin (F-actin) with tetramethylrhodamine isothiocyanate-conjugated phalloidin and DNA with 4'6-diamidino-2-phenylindole·2 HCl (DAPI). The images show optical 0.50 μm sections for each of these experiments. Scale bars, 10 μm.
a. hKIAA1202-V5 partially co-localises with the F-actin based cytoskeleton.
b. hKIAA1202-V5 partially co-localises with the F-actin remnants when the F-actin cytoskeleton is disrupted by latrunculin B.
c. hFBXO4-V5 does not co-localise with F-actin.
d. Upon latrunculin B treatment, the F-actin cytoskeleton is disrupted without affecting the hFBXO4-V5 expression pattern. The inset shows part of the focal plane 7 μm below that of the main image.
e, f. Vector control.


a
hKIAA1202-V5
V5/Phalloidin/DAPI
V5
Phalloidin
- LAT B
b
hKIAA1202-V5
V5/Phalloidin/DAPI
V5
Phalloidin
+ LAT B
c
hFBXO4-V5
V5/Phalloidin/DAPI
V5
Phalloidin
- LAT B
d
hFBXO4-V5
V5/Phalloidin/DAPI
V5
Phalloidin
+ LAT B
e
pcDNA4V5
V5/Phalloidin/DAPI
V5
Phalloidin
- LAT B
f
pcDNA4V5
V5/Phalloidin/DAPI
V5
Phalloidin
+ LAT B

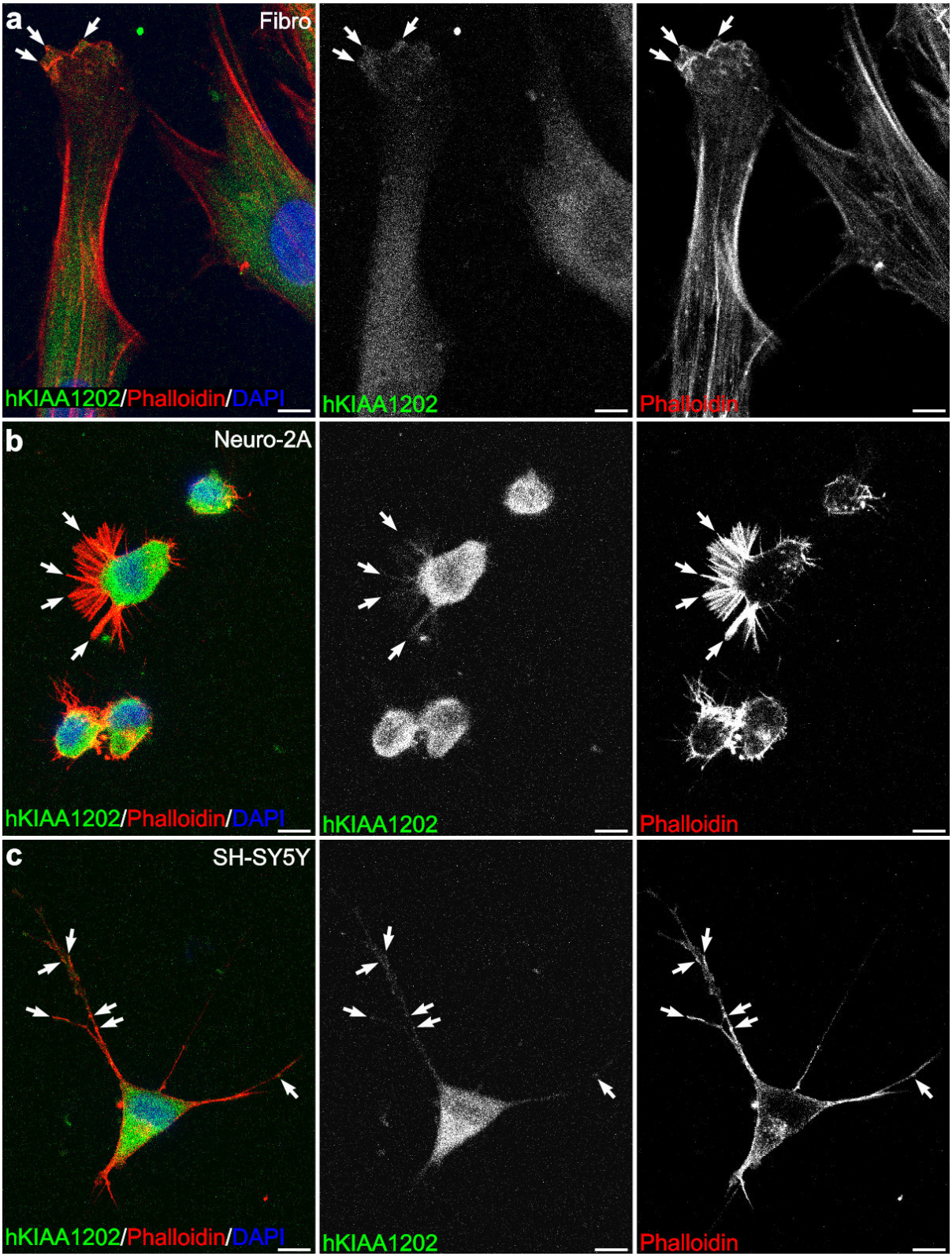


Fig. III-36 | **Endogenous Kiaa1202 localises at cellular sites of rapid Actin remodelling.**

Simultaneous detection of Kiaa1202 (with α-hKIAA1202 antibody) and filamentous Actin (F-actin) shows overlapping signals at the leading edge of fibroblasts (panel a, arrows), and in the neurites of Neuro-2A (panel b, arrows) and SH-SY5Y cells (panel c, arrows). The images show optical 0.50 μm sections for each experiment. F-actin and DNA staining as in Fig. III-35; scale bars, 10 μm.

and retraction of neuronal cell protrusions[791], we differentiated Neuro-2A cells and again performed double staining of mKiaa1202 and F-actin. Even though observations were made close to the threshold of detection, confocal recordings show an increased mKiaa1202 signal coinciding with the most prominent spots of F-actin staining along the cellular projections (Fig. III-37). To verify these results, the experiment was repeated on differentiated SH-SY5Y cells. However, the protrusions, which are barely detectable using confocal microscopy, become markedly thinner upon differentiation, such that the level of hKIAA1202 signal is below the threshold of detection in confocal analyses. Although fine structures still could not be visualised with conventional epifluorescence microscopy, we were able to record staining of networks of interlinked differentiated cells to a certain extent (Fig. III-38). These observations, though not conclusive, seem to be consistent with the situation observed in Neuro-2A cells and with that of the over-expression studies.

C.3.4.2.3. Mitochondrial outer membrane recruitment assay

To employ the MOM recruitment assay, which has been used to establish proteins' interaction with F-actin by assessing their subcellular relocalisation, we cloned *hKIAA1202-V5* 3' of the topogenic scTOM70 MOM sequence. MOM-hKIAA1202-V5 localises to the mitochondria and redirects F-actin's subcellular distribution.

To interact, proteins must reside at the same subcellular location, so co-localisation is a condition *sine qua non* for protein interactions. It is clear, however, that the co-localisation studies described earlier are only a first approach towards the delineation of a putative hKIAA1202 – F-actin interaction. To find out whether hKIAA1202 actually interacts with F-actin, we employed the MOM recruitment assay published by Kaufmann *et al.*[792]. We fused hKIAA1202-V5 to the 29 AA sequence (MKSFITRNKTAILATVAATGTAIGAYY-YY) targeting scTOM70, part of a supramolecular complex accelerating the import of mitochondrial precursor proteins[793,794], to the MOM[795]. N-terminal fusions of proteins to this topogenic peptide were shown to locate specifically to the MOM and, thus, over-expression of a *MOM-hKIAA1202-V5* construct should anchor hKIAA1202 to the mitochondria.

The MOM anchor sequence was PCR-amplified from a pcDNA3 clone containing the topogenic encoding sequence (courtesy of Dr. M. Rothkegel, TU Braunschweig, Braunschweig, Germany), introducing *Nhe*I and *Hin*dIII restriction sites. The amplicon was TA-cloned into pGEM-T Easy and sequence-verified prior to isolation of the 107 bp *Nhe*I/*Hin*dIII fragment. The *hKIAA1202-V5* construct was digested with *Nhe*I and *Hin*dIII, resulting in 8405 bp *Nhe*I/*Nhe*I, 1183 bp *Nhe*I/*Hin*dIII and 16 bp *Hin*dIII/*Nhe*I fragments. Whereas the

latter was discarded, the *Nhe*I/*Hin*dIII fragment was ligated to the 107 bp *Nhe*I/*Hin*dIII fragment, followed by *Nhe*I digestion to avoid concatenates and *Hin*dIII/*Hin*dIII fragments. The resulting 1290 bp *Nhe*I/*Nhe*I fragment was fused to the 8405 bp *Nhe*I/*Nhe*I fragment yielding the 9695 bp *MOM-hKIAA1202-V5* plasmid.

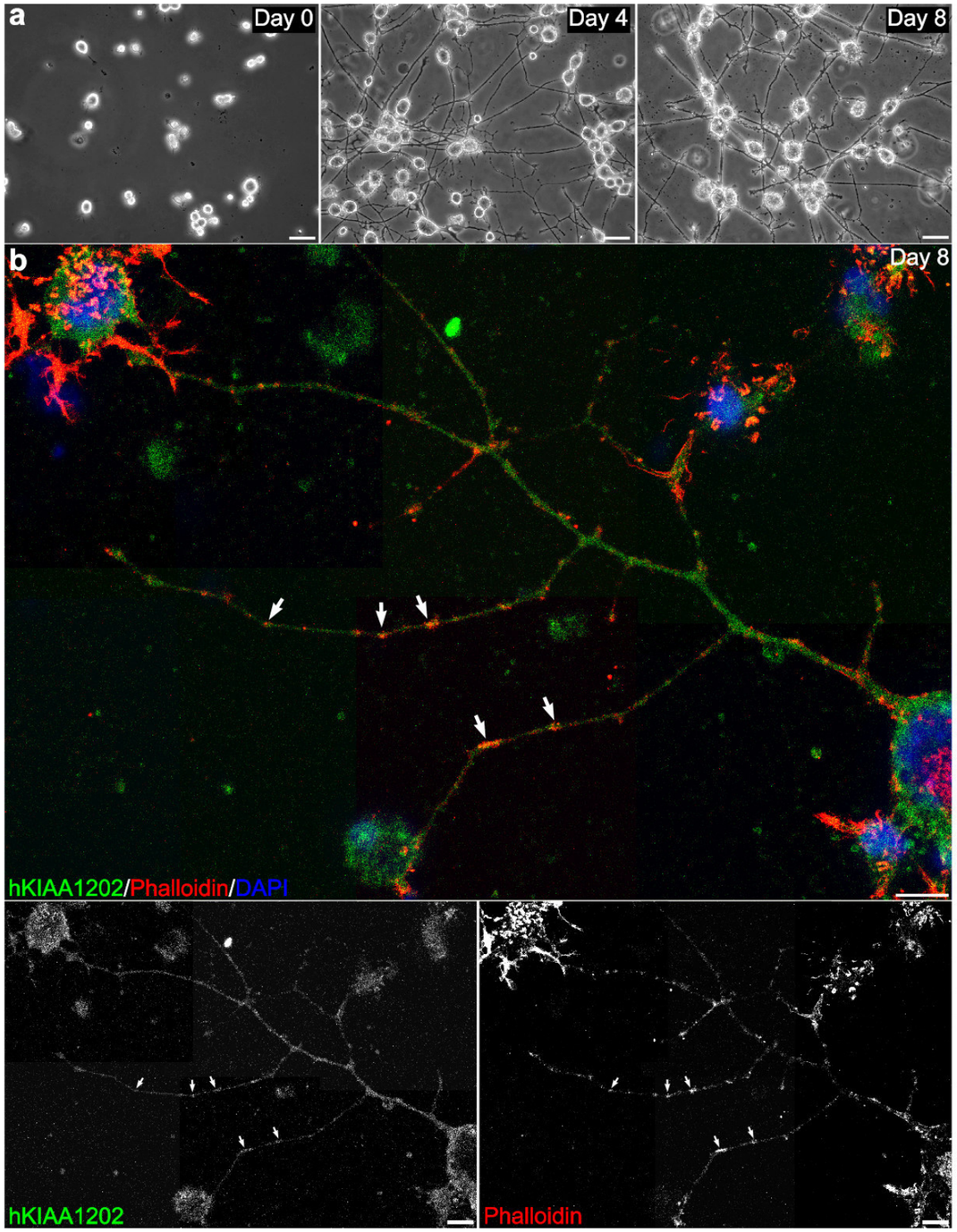
Confocal studies of cells transfected with *MOM-hKIAA1202-V5* and subsequently immunolabelled with α-V5 and the mitochondrial marker α-Prohibitin[796] showed that MOM-hKIAA1202-V5 is, indeed, targeted to the mitochondria (Figs. III-39a – b). Cells expressing the MOM-hKIAA1202-V5 showed mitochondrial clumping (Fig. III-39b), which has been shown to be characteristic for F-actin binding proteins fused to mitochondrial anchor sequences[797], and is a first indication for the interaction of hKIAA1202 with F-actin. Staining of those cells with TRITC-conjugated phalloidin showed that F-actin is sequestered to the clumped mitochondria. In non-transfected cells, F-actin was never observed to co-localise with the mitochondria (Figs. III-39c – e). Taken together, our results show that hKIAA1202 is able to direct the subcellular distribution of F-actin. This is strong evidence for an *in vivo* hKIAA1202-V5 – F-actin interaction.

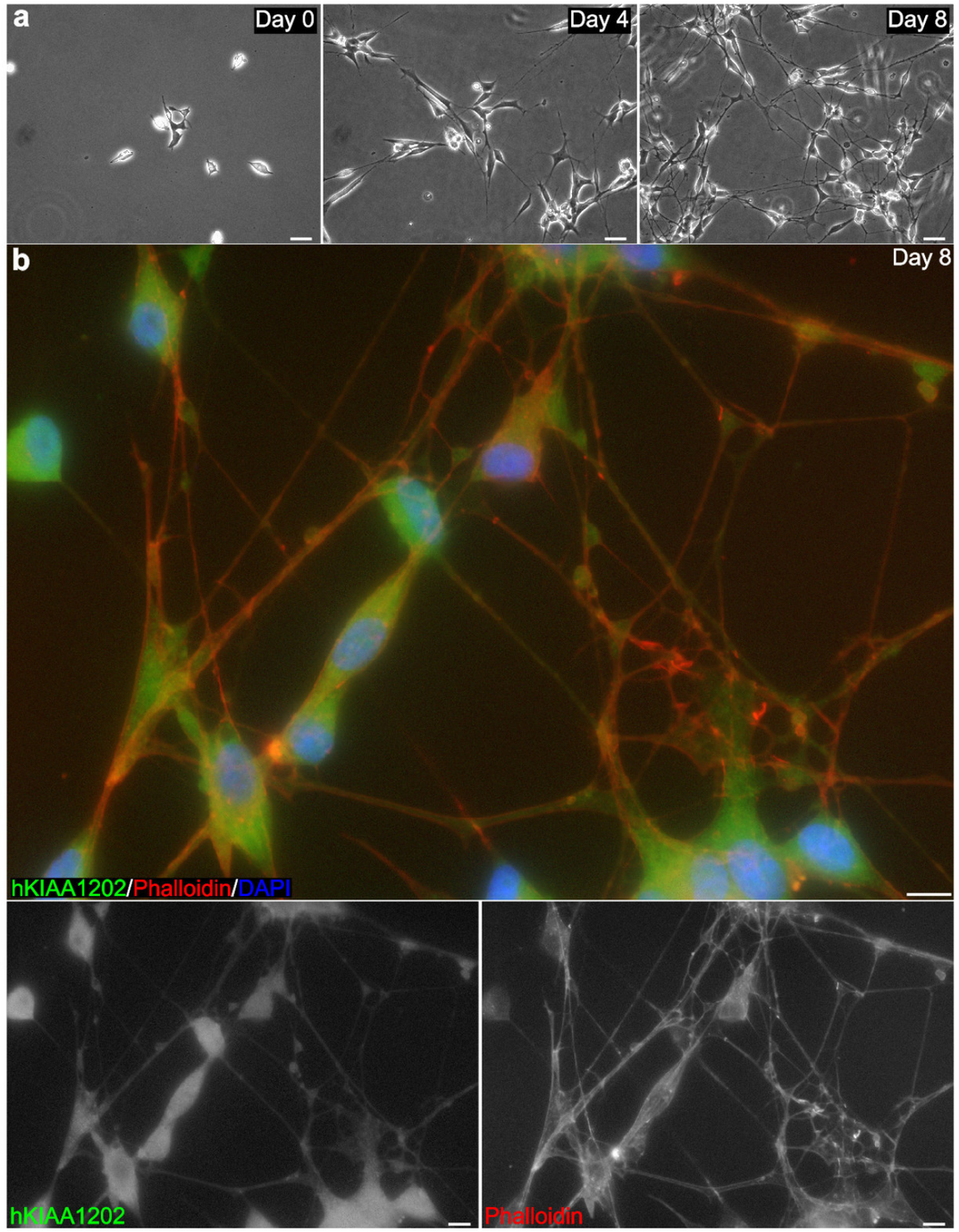
Fig. III-37 | *Next page, left.* **Endogenous mKiaa1202 co-localises with the most prominent spots of filamentous Actin staining in the neurites of differentiating Neuro-2A cells.**

a. Light microscopic image of plated, differentiating Neuro-2A cells at days 0, 4 and 8. Scale bars, 30 μm.
b. Neuro-2A cells are differentiated for eight days prior to fixation. Simultaneous detection of mKiaa1202 (with α-hKIAA1202 antibody; bottom panel, left) and filamentous Actin (F-actin; bottom panel, right) shows partial overlap of both signals (arrows). The image is composed of overlapping confocal recordings. F-actin and DNA staining as in Fig. III-35; scale bar, 20 μm.

Fig. III-38 | *Next page, right.* **Endogenous hKIAA1202 seems to co-localise with filamentous Actin in networks of interlinked differentiating SH-SY5Y cells.**

a. Light microscopic image of plated, differentiating SH-SY5Y cells at days 0, 4 and 8. Scale bars, 30 μm.
b. SH-SY5Y cells are differentiated for eight days prior to fixation. The image shows simultaneous detection of hKIAA1202 (with α-hKIAA1202 antibody; bottom panel, left) and filamentous Actin (F-actin; bottom panel, right). Fine structures are not recorded as epifluorescence photographs were made at the threshold of detection. F-actin and DNA staining as in Fig. III-35; scale bar, 20 μm.


a
Day 0
Day 4
Day 8
b
Day 8
hKIAA1202/Phalloidin/DAPI
hKIAA1202
Phalloidin


a
Day 0
Day 4
Day 8
b
Day 8
hKIAA1202/Phalloidin/DAPI
hKIAA1202
Phalloidin

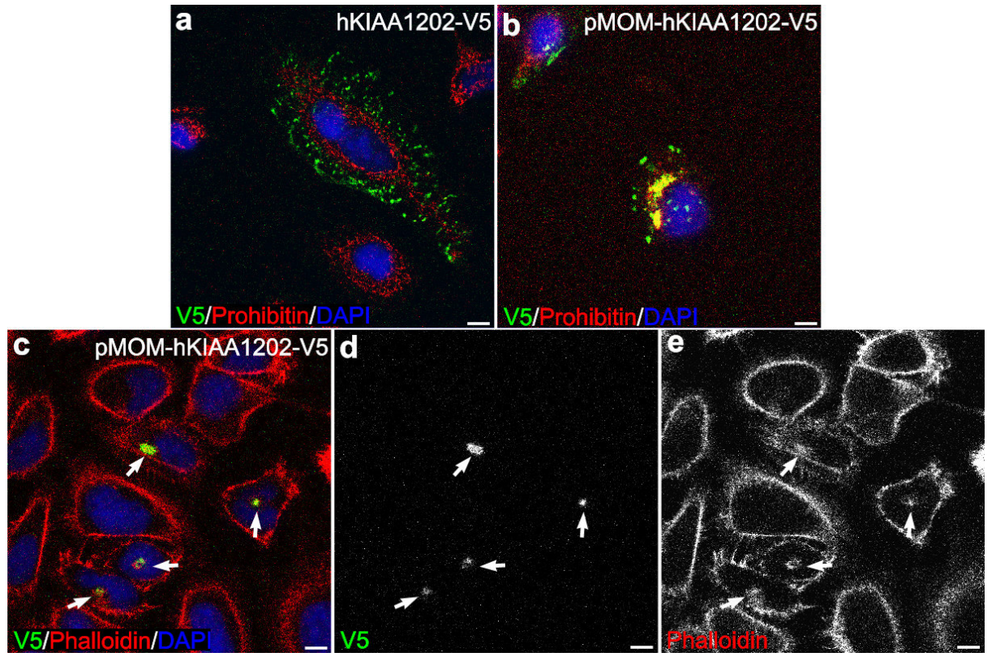


Fig. III-39 | **hKIAA1202 can redirect the subcellular localisation of filamentous Actin.**

Indicated C-terminally V5-tagged constructs are ectopically expressed in HeLa cells and detected with α-V5 antibody. Prohibitin is stained with α-Prohibitin antibody. The images show optical 0.50 μm sections for each experiment. Filamentous Actin (F-actin) and DNA staining as in Fig. III-35; scale bars, 10 μm.
a. In HeLa cells, hKIAA1202-V5 does not co-localise with the mitochondrial marker protein Prohibitin.
b. MOM-hKIAA1202-V5 is targeted to the mitochondria. Note the extensive mitochondrial clumping (yellow).
c, d, e. In transfected cells, F-actin forms aggregates (arrows) that co-localise with MOM-hKIAA1202-V5. Such F-actin aggregation is not seen in non-transfected cells.

C.3.4.2.4. Co-immunoprecipitation experiments with hKIAA1202 and filamentous Actin

Pilot co-IP experiments suggest a specific interaction between hKIAA1202 and Actin, although these results need further delineation.

To further confirm the notion that hKIAA1202 interacts with F-actin, a series of co-IP experiments was performed.

Using 2 μg of α-Actin antibody raised in rabbit or 2 μg of rIgG, co-IPs were performed on 800 μg lysate of HeLa cells over-expressing hKIAA1202-V5. Reactions were washed with Pagano Lysis Buffer containing 1 M NaCl, size-separated on 8% SDS-PAGE gels, and immunoblotted with α-Vimentin (see III.C.3.4.3.4), α-V5 and α-Actin antibodies. Under these conditions, specific interaction between over-expressed hKIAA1202-V5 and endogenous Actin and Vimentin was observed (Fig. III-40a).

To circumvent possible over-expression-induced artefacts, α-Actin co-IPs were performed on lysates from U373 MG cells stably transfected with hKIAA1202-V5 or pcDNA4/V5-HisB. Using 1 µg of α-Actin antibody, co-IPs were performed on 1 mg lysate. Reactions were washed with Pagano Lysis Buffer, size-separated on 8% SDS-PAGE gels and immunoblotted with α-V5 and α-Actin antibodies. Under these conditions, specific interaction between physiologically relevant amounts of V5-tagged hKIAA1202 and endogenous Actin was observed (Fig. III-40b). Even though these results are in agreement with those of the over-expression study, it should be noted that the IgG control described for the over-expression studies is much more stringent than the vector control applied in the experiments using U373 MG lysates.

Finally, 2.5 µg of α-Actin antibody raised in goat or 2.5 µg of gIgG were used in co-IPs on 2.5 mg of SH-SY5Y lysate. Reactions were washed with Pagano Lysis Buffer containing 1 M NaCl, size-separated on 8% SDS-PAGE gels and immunoblotted with α-hKIAA1202 and α-Actin antibodies. Specific interaction between the endogenous ~200 kDa hKIAA1202 species (Fig. III-29b, lane 'SY-SY5Y') and endogenous Actin was observed, which is again in line with the earlier results. A specific interaction between other hKIAA1202 isoforms and Actin was not observed (Fig. III-40c).

Although all observations were consistent with the idea of hKIAA1202 interacting with Actin, it should be stressed that these results are preliminary. Further optimisation is indispensable in order to confirm and delineate the hKIAA1202 – Actin interaction in detail. For example, the over-expression data should be validated with differently tagged versions of hKIAA1202. To this end, hKIAA1202 has already been tagged with (i) N-terminal HA, (ii) C-terminal HA, (iii) N-terminal FLAG/C-terminal V5 and (iv) N-terminal cMyc/C-terminal V5. Also, the endogenous experiments need to be repeated in different cell lines. With regard to the situation in the brain, the most stringent validation would be to replicate the observations in primary mammalian neurons.

C.3.4.3. Putative hKIAA1202 interactions identified by yeast two-hybrid screening

Collaboration with Dr. U. Stelzl, MDC, Berlin, Germany.

To determine hKIAA1202 interacting partners other than F-actin, we made use of automated Y2H methodology.

C.3.4.3.1. Rationale and overview of the yeast two-hybrid system

Y2H methodology has been used frequently to screen for protein – protein interactions.

To determine possible hKIAA1202 interaction partners other than F-actin, we decided to employ Y2H methodology to screen a set of ~5500 non-redundant human cDNAs, described under IIB.A.5.7.

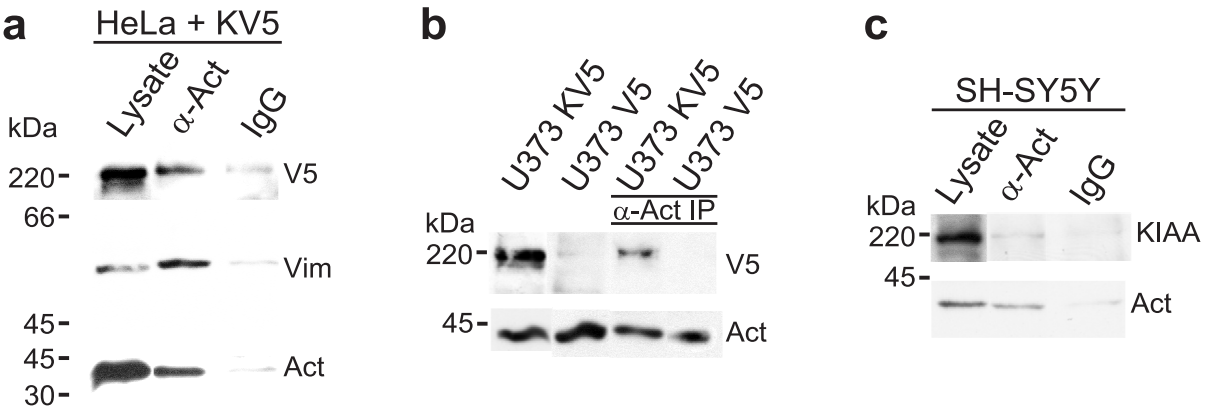


Fig. III-40 | **Putative interaction between hKIAA1202, and Actin and Vimentin.**

a. HeLa cells are transiently transfected with hKIAA1202-V5 (KV5). Cells are lysed and, from these lysates, Actin is immunoprecipitated with α-Actin antibody (α-Act), and precipitates are size-separated and immunoblotted with α-V5 (V5), α-Vimentin (Vim) and α-Actin (Act) antibodies. An IgG immunoprecipitation is included as a control. Size marker on the left.
b. U373 MG cells stably transfected with hKIAA1202-V5 (U373 KV5) or vector (U373 V5) are lysed and Actin is immunoprecipitated (IP) from the lysates with α-Actin antibody. Precipitates are size-separated and immunoblotted with α-V5 and α-Actin antibodies. Size marker on the left.
c. SH-SY5Y cells are lysed and Actin is immunoprecipitated from the lysate with α-Actin antibody. The precipitate is size-separated and immunoblotted with α-hKIAA1202 (KIAA) and α-Actin antibodies. An IgG immunoprecipitation is included as a control. Size marker on the left.

Essentially, the Y2H screen was conducted by mating L40ccua yeast cells expressing the hKIAA1202 baits (Y2H-1 – Y2H-6) with L40ccα cells expressing the preys (each of the ~5500 cDNA clones). The L40 strain is Trp-, Leu- and His-deficient. As the bait vector contains the Trp CDS, haploid cells transformed with bait constructs are no longer Trp-deficient. Similarly, haploid cells transformed with prey constructs are no longer Leu-deficient, as the prey vector encodes the Leu CDS. After mating, diploid cells containing both bait and prey constructs are selected for on SDII medium lacking Leu and Trp. When the bait protein fused to the lexA DNA BD interacts with the prey protein fused to the GAL4 AD, this results in transcription of the *his3* and *lacZ* genes, which are both under the control of lexA promoters. The interaction-induced His rescue can be selected for on SDIV medium lacking Leu, Trp, His and Ura, whereby the latter was omitted from the medium to sustain general selective

pressure throughout the screening procedure. As specified below, transcription of the *lacZ* reporter gene allows for β-galactosidase assaying.

To avoid large numbers of false positives, which is one of the major drawbacks of Y2H technology, the following measures were taken:

- Prior to the actual screen, cells expressing Y2H-1 – Y2H-6 bait constructs were mated with cells expressing the empty prey vector and cells expressing Y2H-1 – Y2H-6 cloned into the prey vector. All matings were grown on SDII plates lacking His and Ura. Baits growing after mating with the prey vector and/or with all preys were discarded from the screen, as they are auto-activating. Similarly, auto-activation of preys was tested by growing matings on SDII plates lacking Leu and Ura.
- Positives after a first round of screening were confirmed by reciprocal mating; baits and preys were swapped.
- Interactions that were positive after reciprocal mating were further confirmed by growing matings on filters placed onto SDIV plates lacking Leu, Trp, His and Ura. This SDIV filter assay selected for those matings leading to an expression rate of the *his3* marker gene, a measure for the strength of interaction between bait and prey, that was high enough to enable growth of diploid cells not only on the SDIV plate, but also on the filter. Interactions testing positive in this assay are termed 'medium-confidence interactions'.
- Finally, a β-galactosidase assay was performed on matings growing on filters, which were placed on SDIV plates. Only interacting bait and prey proteins would enable transcription of the *lacZ* reporter gene, which encodes β-galactosidase, which, in turn, hydrolyses the colourless β-galactoside X-gal. Eventually, this hydrolysis leads to the formation of an indigo dye, allowing straightforward identification of interaction partners. Interactions testing positive in this assay are termed 'high-confidence interactions'.

C.3.4.3.2. Cloning of hKIAA1202 constructs for use in yeast two-hybrid

Courtesy of J. Ruschmann, MPI-MG, Berlin, Germany.

Six overlapping hKIAA1202 fragments were cloned into Y2H bait and prey vectors.

To screen for putative hKIAA1202 interactions by Y2H, we divided the sequence encoding AA 5 – 1489 of hKIAA1202 isoform I into six overlapping fragments: Y2H-1, AA 5 – 295; Y2H-2, AA 253 – 513; Y2H-3, AA 501 – 752; Y2H-4, AA 722 – 983; Y2H-5, AA 977 – 1232 and Y2H-6, AA 1226 – 1489. We cloned each fragment into the pGAD426

('prey vector') and pBTM117c ('bait vector') vectors, generating N-terminal in-frame fusions with the GAL4 activation and lexA DNA binding domains, respectively. Fragments were PCR-amplified from the hKIAA1202-V5 clone, introducing *Sal*I and *Not*I restriction sites. Primers were designed so that sequences encoding protein domains were contained in single amplicons (Fig. III-20c). PCR products were digested with *Sal*I and *Not*I, and directionally inserted into both Y2H vectors. All constructs were sequence-verified.

C.3.4.3.3. Auto-activation and homo-oligomerisation

Auto-activation assays reveal a possible homo-oligomerisation of hKIAA1202.

When mating all hKIAA1202 Y2H constructs pairwise with one another, strong auto-activation was seen for Y2H-1, and relatively weak auto-activation was observed for Y2H-5 and Y2H-6 (Fig. III-41). Therefore, these three constructs were omitted from the actual Y2H screen. Unfortunately, this implied that a number of putative binding partners may have been missed. As empirical evidence showed that the addition of N-terminal amino acids often abolishes auto-activation (Dr. U. Stelzl, personal communication), Y2H-1, Y2H-5 and Y2H-6 could be replaced as follows: Y2H-4, Y2H-5 and Y2H-6 could be combined in a single construct (Y2H-4 – 6, AA 722 – 1489) and Y2H-1 could be incorporated into constructs Y2H-1 – 3 (AA 5 – 752) and Y2H-1 – 6 (AA 5 – 1489) with the insertion of additional 5' sequence. The latter could easily be obtained using PCR technology. Introduction of *Sal*I and *Not*I restriction sites would again allow insertion between the *Sal*I and *Not*I sites of the pGAD426 and pBTM117c vectors. While cloning of Y2H-4 – 6 is complete, clones Y2H-1 – 3 and Y2H-1 – 6 await completion.

Interestingly, cells expressing Y2H-4 as a bait grew after mating with cells expressing Y2H-4 as a prey, indicating a possible oligomerisation domain within Y2H-4 (Fig. III-41). This clone encodes hKIAA1202 AA 722 – 983, including a region containing 21% Cys-residues (14/67 AA). The observation that the hKIAA1202 banding pattern of α-hKIAA1202-immunoblotted cell lysates did not change by adding up to 80 mM DTT precluded the possibility that Y2H-4's self-activation was due to sulphide bonding between the Cys-rich regions. To verify the ability of hKIAA1202 to form multimers, we immunoprecipitated HA-tagged hKIAA1202 that was over-expressed in a U373MG cell line stably expressing hKIAA1202-V5. Immunoprecipitates were size-separated and immunoblotted with α-V5 and α-HA antibodies, suggesting binding between hKIAA1202-HA and hKIAA1202-V5 (Fig. III-42a). This result suggests the existence of hKIAA1202 oligomers *in vivo*, which is in line with the original Y2H-4 auto-activation observation.

To further confirm the ability of hKIAA1202 to form oligomers and to establish the number of subunits of such oligomers, we performed chemical cross-linking using GA, which is a recognised strategy to study multimerisation[798]. To this end, we cloned a V5/His_6-tagged version of Y2H-4. The Y2H-4 insert was PCR-amplified from the *hKIAA1202-V5* clone, with

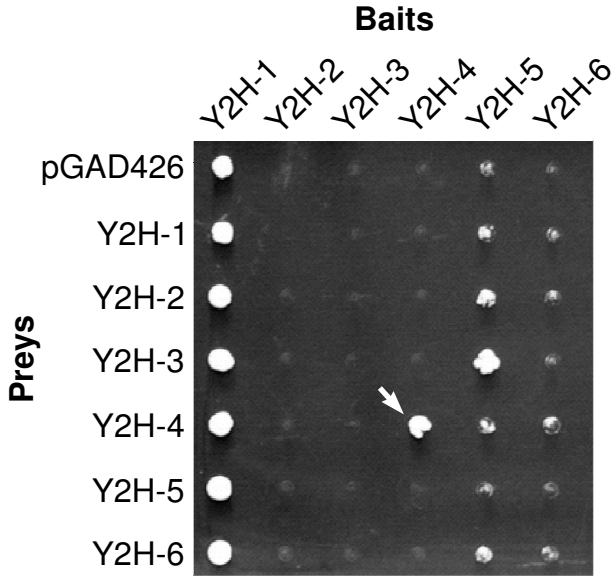


Fig. III-41 | **Auto-activation assay of hKIAA1202 yeast two-hybrid constructs.**

One-to-one mating of yeast cells expressing parts of hKIAA1202 open reading frame (ORF) I fused to the LexA DNA binding domain (baits) with cells transfected with the pGAD426 vector or parts of the hKIAA1202 ORF I fused to the GAL4 activation domain (preys) showed strong auto-activation for the yeast two-hybrid 1 (Y2H-1) clone. Constructs Y2H-5 and Y2H-6 weakly auto-activate. Clone Y2H-4 reveals homo-oligomerisation (arrow).

Photograph courtesy of Dr. U. Stelzl, MDC, Berlin, Germany.

addition of *Hin*dIII and *Apa*I restriction sites. The amplicon was TA-cloned in pGEM-T Easy and sequence-verified before being directionally inserted between the *Hin*dIII and *Apa*I sites of pcDNA4/V5-HisB. Immunoblotting of lysate from HeLa cells over-expressing Y2H-4-V5/His_6, followed by detection with an α-V5 antibody, showed a single band with a MW of ~34 kDa (Fig. III-42b), which corresponds with the calculated MW of 33 kDa. Y2H-4-V5/His_6, over-expressed in HeLa cells was affinity-purified from the lysate by Ni-NTA chromatography, concentrated using a centrifugal filter device with the appropriate MW cut-off and added to a series of GA cross-linking reactions. Reaction conditions were varied as described under IIB.A.5.5. Upon immunoblotting of the reactions with α-V5 antibody, two clear shifts in the MW of Y2H-4-V5/His_6 became apparent with as little as 1.0 mM GA added. While the shift from B (~34 kDa) to A (~30 kDa) is likely due to intramolecular cross-linking, the shift from A to C (~85 kDa) is a consequence of intermolecular cross-linking

events (Fig. III-42b). Although precise MW estimates are not evident, the A:C MW ratio suggests a Y2H-4-V5/His_6 trimerisation. As expected, the GA-dependent decrease in intensity of A is mirrored by an increased intensity of C, especially at GA concentrations of up to 10 mM. At higher GA concentrations, this correlation is no longer obvious as the high MW complexes (D), which hardly enter the running gel, increasingly scavenge Y2H-4-V5/His_6.

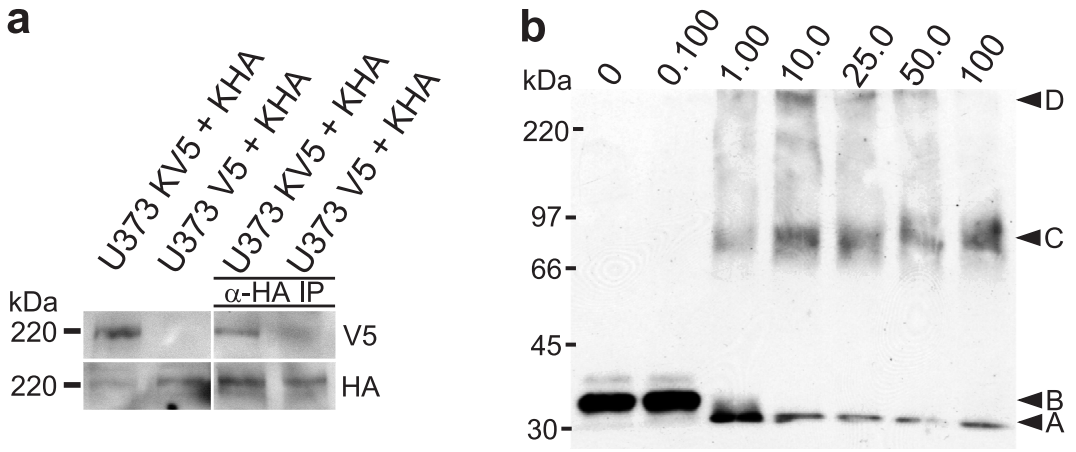


Fig. III-42 | **Putative homo-oligomerisation of hKIAA1202.**

a. A U373 MG cell line stably expressing hKIAA1202-V5 (U373 KV5) or vector (U373 V5) is transiently transfected with Hemagglutinin (HA)-tagged hKIAA1202 (KHA). Cell lysates are size-separated, and immunoblotted with α-V5 and α-HA antibodies (first two lanes). The same lysates are used in an α-HA immunoprecipitation (IP) assay and subsequent immunoblotting with α-V5 and α-HA antibodies (last two lanes) suggests binding between hKIAA1202-V5 and hKIAA1202-HA. Size marker on the left.
b. 330 ng purified and V5-tagged protein fragment consisting of hKIAA1202 amino acids 722 – 983 is incubated in Cross-linking – Buffer A at room temperature for 30 minutes with millimolar concentrations of glutaraldehyde as specified at the top. Amino acids 722 – 983 show self-activation in the yeast two-hybrid auto-activation assay. See text for a discussion of banding patterns A – D (arrowheads) that occur after size-separation and α-V5 immunoblotting of incubations. Size marker on the left.

Taken together, these data suggest that hKIAA1202 forms oligomers which may consist of three subunits. However, these results need to be confirmed, which may be complicated by a putative MW of ≥ ~600 kDa for the full-length hKIAA1202 oligomer. This problem might be overcome by over-expressing hKIAA1202-V5 in HeLa cells, followed by gel filtration of the lysate. Alternatively, over-expressed hKIAA1202-V5 may be purified, concentrated and cross-linked as described for Y2H-4-V5/His_6. However, in case of full-length hKIAA1202-V5 oligomers, the cross-linking reactions should be size-separated using SDS-VAGE followed by α-V5 immunoblotting. This system has been designed for the detection of proteins with a MW of up to 4000 kDa[799]. Ideally, endogenous full-length hKIAA1202 oligomers could be detected by establishing a non-denaturing variant of VAGE followed by α-hKIAA1202 immunoblotting.

C.3.4.3.4. Putative Kiaa1202 – Vimentin interaction

Automated Y2H screening recovered nine high- and eight medium-confidence hKIAA1202 interactions, including a possible hKIAA1202 – Vimentin interaction. Endogenous Kiaa1202 co-localises partially with Vimentin in mammalian cells and pilot co-IP experiments suggest an interaction between hKIAA1202-V5 and Vimentin.

Due to auto-activation, only constructs Y2H-2, Y2H-3 and Y2H-4 have been employed in the automated Y2H screen. After parsing of the results by applying all selection criteria described under III.C.3.4.3.1, seventeen putatively meaningful hKIAA1202 interaction partners were retained. Whereas nine high-confidence interactions tested positive for the β-galactosidase assay, which is the most stringent criterion, eight medium-confidence interactions tested positive for the SDIV filter assay (Table III-12). Intriguingly, all potential interaction partners showed binding with Y2H-4, which also harboured the putative homo-oligomerisation domain.

The intermediate filament Vimentin, one of the high-confidence interactors emerging from the hKIAA1202 Y2H screen (Table III-12), is a cytoskeletal component[800]. Due to several lines of evidence pointing towards an interaction between Kiaa1202 and the F-actin-based cytoskeleton (see III.C.3.4.2) and an involvement of Vimentin in neuritogenesis[801], Vimentin was considered a prime candidate for further interaction studies.

A first step towards the confirmation of a putative Kiaa1202 – Vimentin interaction was a series of co-localisation experiments. Double staining of several mammalian cell lines with α-hKIAA1202 and α-Vimentin antibodies, followed by confocal analysis, indicated that Kiaa1202 co-localised with Vimentin. As for F-actin, Vimentin co-localised with Kiaa1202 at the leading edge of primary fibroblasts and in the protrusions of neuronal cells. Moreover, Vimentin and Kiaa1202, at least in part, co-localised in the cytoplasm in all cell lines investigated (Fig. III-43). Due to the low level of Kiaa1202 expression, resulting in weak fluorescence signals, a complete overlap between Kiaa1202 and Vimentin staining was not observed consistently. To get a better impression of the situation in neurites, we studied localisation of both proteins simultaneously in differentiated Neuro-2A cells. Even though mKiaa1202 signals were weak in confocal analysis, a co-localisation between mKiaa1202 and Vimentin stainings along the cellular processes was detected, especially during the earlier stages of differentiation (Fig. III-44), which is in line with observations made by Shea *et al.*, showing a requirement for Vimentin in the initiation of neuritogenesis[802]. Despite the fact that this result

Table III-12 | **Putative biologically relevant hKIAA1202 interaction partners acquired through Y2H screening**

hKIAA1202 Y2H-2/3/4[§]	**Putative interaction partner (Accession Number)** [†]	**Function**	**Refs.**
		Group I – Transcriptional control	
2, **3**, 4	H3-K9 methyltransferase 2 (Q9H5I1)	H3-K9-specific methyltransferase participating in regulation of higher order chromatin organisation (during spermatogenesis). Recruiting HP1 to methylated H3-K9 represents a tag for transcriptional repression.	412,415, 803,804
4	H3-K9 methyltransferase 4 (Q15047)	H3-K9-specific methyltransferase contributing to euchromatic gene silencing by KRAB ZFPs.	443
4	HTATIP (Q9BWK7)	Catalytic subunit of the NuA4 histone acetyltransferase complex, which is involved in activation of transcriptional programs associated with oncogene mediated growth induction, tumour suppressor mediated growth arrest, replicative senescence, apoptosis, and DNA repair.	805,806
3, **4** (2×)	Chromodomain helicase-DNA-binding protein 3 (Q9Y4I0)	Central component of the NuRD complex with ATP-dependent chromatin-remodelling and histone deacetylase activity. Probable transcription regulator.	807,808
4	Transducin-like enhancer protein 1 (Q04724)	Transcriptional corepressor that binds to a number of TFs (e.g. inhibits NF-κB-regulated gene expression). Binds H3.	809-811

Table III-12 | **Putative biologically relevant hKIAA1202 interaction partners acquired through Y2H screening**

hKIAA1202 Y2H-2/3/4[§]	**Putative interaction partner (Accession Number) [†]**	**Function**	**Refs.**
		Group II – Protein modification	
4	COP9 signalosome complex, subunit 6 (Q7L5N1)	The COP9 signalosome complex is a regulator of the Ubiquitin conjugation pathway by mediating the deneddylation of the cullin subunits of SCF-type E3 ligases, thereby decreasing their activity.	812
4	ASB13 (Q8WXK3)	Ankyrin repeat and SOCS box proteins bridge between substrates (via the ankyrin protein-protein interaction module) and E3 ligases (via the SOCS box), thereby targeting proteins for poly-ubiquitinylation.	813-815
4 (2×)	SUMO 2 (precursor) (P61956)	Post-translational protein tag interfering with multiple biological processes such as the cell cycle, subcellular transport and transcription.	816,817
		Group III – Unknown function	
4	Hypothetical protein, 41.1 kDa (Q9H0V7)	NA	NA
4	Hypothetical protein, 72.7 kDa (Q9NTF9)	NA	NA
4	KIAA1377 (fragment) (Q9P2H0)	Was shown to interact with Huntingtin in a Y2H screen.	818
		Group IV – Other	
3, **4**	Vimentin (fragment) (Q15868)	Vimentins are class-III intermediate filaments.	800
4	Importin α1 subunit (P52294)	Functions as an adapter protein in nuclear protein import. Binds to substrates containing a simple or bipartite NLS.	819,820
4	Growth/differentiation factor 9 (precursor) (O60383)	Secreted protein belonging to the TGFβ family and required for ovarian folliculogenesis.	821

Table III-12 | **Putative biologically relevant hKIAA1202 interaction partners acquired through Y2H screening**

hKIAA1202 Y2H-2/3/4[§]	Putative interaction partner (Accession Number) [†]	Function	Refs.
4[‡]	EF-1γ (fragment) (Q96CU2)	Subunit of EF-1. EF-1 delivers aminoacyl-tRNAs to the A-site of the ribosome and bundles Actin. The γ subunit binds the dopamine D_3 receptor.	822-825
4	Pellino 1 (Q96SM0)	Scaffold protein involved in the IL-1 signaling pathway and required for NF-κB activation.	826
4	Tartrate-resistant acid phosphatase type 5 (precursor) (P13686)	This relatively minor lysosomal isozyme of acid phosphatase can become the dominant isozyme in certain pathologies such as Gaucher disease (OMIM 230800), a cerebroside lipidose.	827

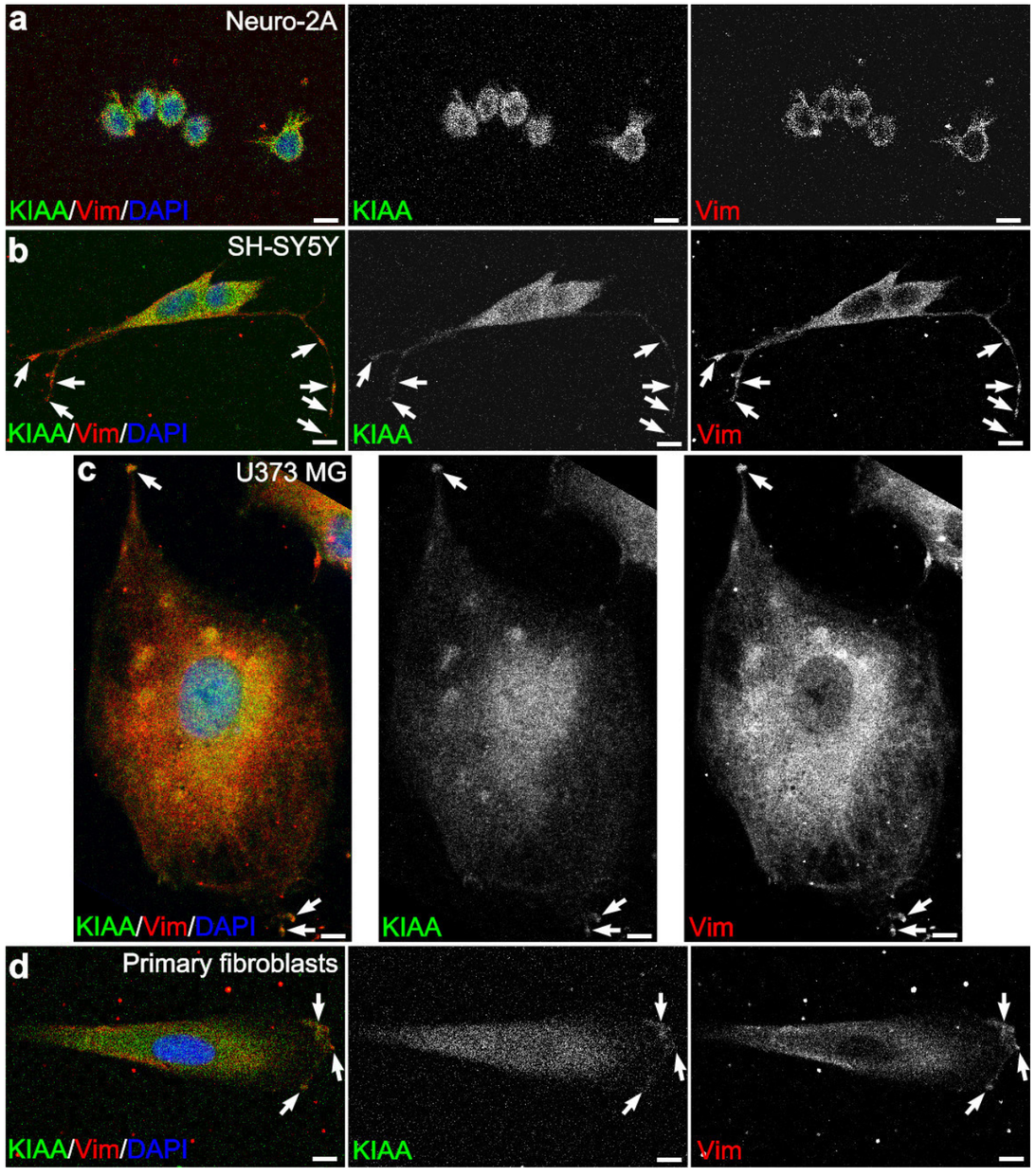
[§] High-confidence interactions are noted in bold.
[†] Swiss-prot accession numbers (http://www.expasy.ch/sprot/).
[‡] An additional medium-confidence interaction between Y2H-4 and full-length EF-1γ (P26641) was obtained.

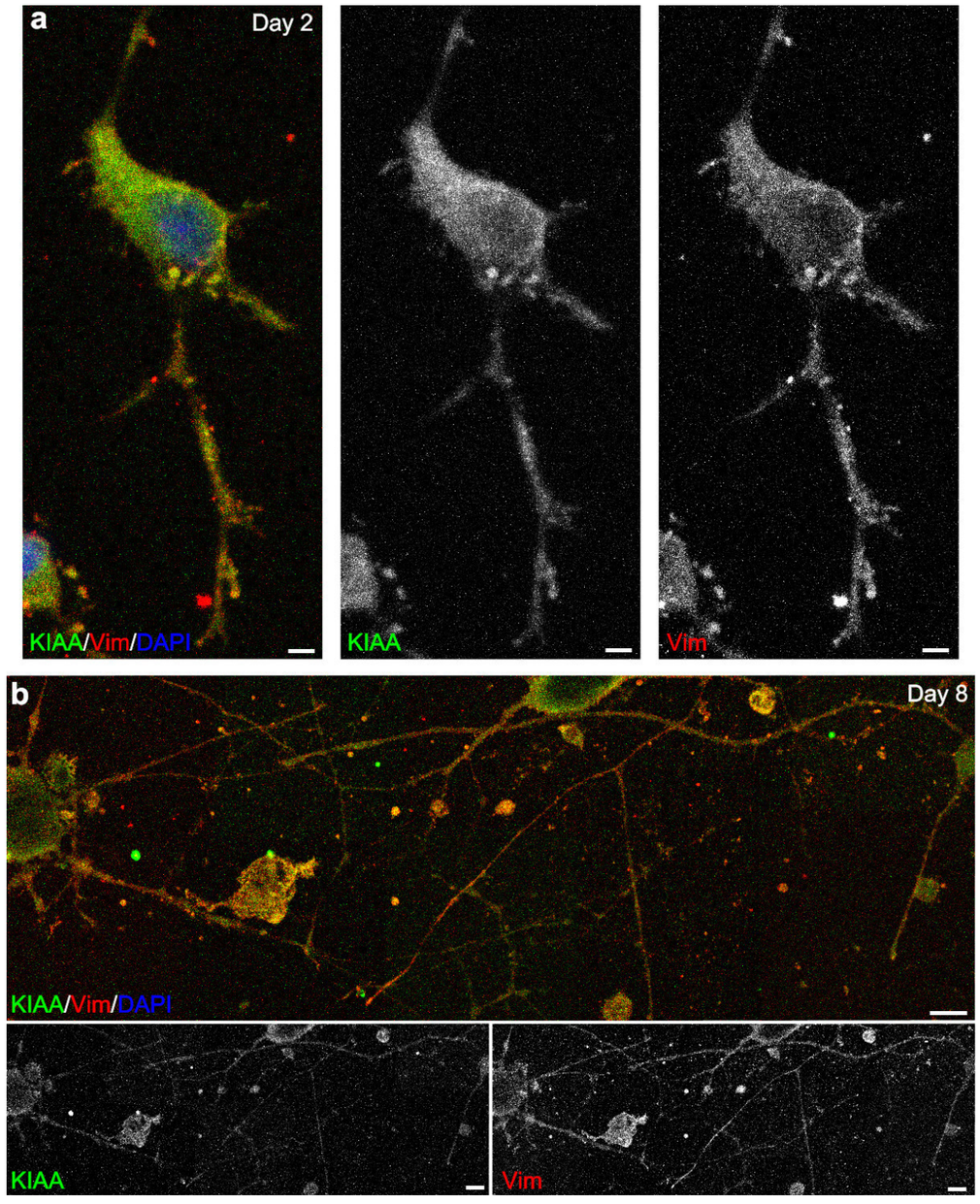
Fig. III-43 | *Next page, left.* **Endogenous Kiaa1202 partially co-localises with Vimentin.**

Simultaneous detection of Kiaa1202 and Vimentin shows partial co-localisation of both proteins in the cytoplasm of all investigated cell lines. Additional co-localisation was observed in the neurites of SH-SY5Y (arrows, panel b) and U373 MG (arrows, panel c) cells, and in the leading edge of primary fibroblasts (arrows, panel d). Kiaa1202 is detected with α-hKIAA1202 antibody (KIAA), Vimentin with α-Vimentin antibody (Vim) and DNA with 4'6-diamidino-2-phenylindole·2 HCl (DAPI). The images show optical 0.50 µm sections for each experiment. Scale bars, 10 µm.

Fig. III-44 | *Next page, right.* **Endogenous mKiaa1202 co-localises with Vimentin in differentiating Neuro-2A cells.**

Simultaneous detection of mKiaa1202 (with the α-hKIAA1202 antibody, KIAA) and Vimentin (with an α-Vimentin antibody, Vim) shows significant overlap of both signals. DNA is counterstained with 4'6-diamidino-2-phenylindole·2 HCl (DAPI). Light microscopic images of the Neuro-2A differentiation process are shown in Fig. III-37a.
a. Neuro-2A cells are differentiated for two days prior to fixation. Scale bar, 5 µm.
b. Neuro-2A cells are differentiated for eight days prior to fixation. The image is composed of overlapping confocal recordings. Scale bar, 25 µm.


a
Neuro-2A
KIAA/Vim/DAPI
KIAA
Vim
b
SH-SY5Y
KIAA/Vim/DAPI
KIAA
Vim
c
U373 MG
KIAA/Vim/DAPI
KIAA
Vim
d
Primary fibroblasts
KIAA/Vim/DAPI
KIAA
Vim


a
Day 2
KIAA/Vim/DAPI
KIAA
Vim
b
Day 8
KIAA/Vim/DAPI
KIAA
Vim

could not be reproduced in differentiated SH-SY5Y cells for reasons specified under III.C.3.4.2.2, a similar pattern of expression was seen in non-induced, slightly differentiated SH-SY5Y cells (Fig. III-43b). Taken together, Kiaa1202 seemed to co-localise with Vimentin in several different mammalian cell lines.

In line with these subcellular localisation data, initial co-IP experiments are suggestive of an interaction between hKIAA1202 and Vimentin. Using 5 µg of α-Vimentin antibody, co-IPs were performed on 1 mg lysate from U373 MG cells stably transfected with hKIAA1202-V5 or pcDNA4/V5-HisB. Reactions were washed with Buffer C, size-separated on 8% SDS-PAGE gels and immunoblotted with α-V5 and α-Vimentin antibodies. Under these conditions, specific interaction between hKIAA1202-V5 and endogenous Vimentin was observed (Fig. III-45). This result is in line with the Actin – hKIAA1202 – Vimentin complex suggested by the α-Actin co-IP experiments (see III.C.3.4.2.4 and Fig. III-40a). As has also been outlined for the initial α-Actin co-IPs, caution should be applied in interpreting the data from these pilot experiments.

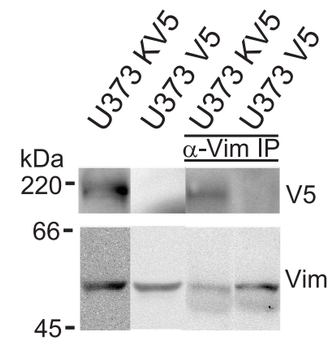


Fig. III-45 | **Putative interaction between hKIAA1202 and Vimentin.**

U373 MG cells stably transfected with hKIAA1202-V5 (U373 KV5) or vector (U373 V5) are lysed and Vimentin is immunoprecipitated (IP) from the lysates with α-Vimentin (α-Vim) antibody. Precipitates are size-separated and immunoblotted with α-V5 and α-Vimentin antibodies. Size marker on the left.