Biallelic Variants in *UBA5* Link Dysfunctional UFM1 Ubiquitin-like Modifier Pathway to Severe Infantile-Onset Encephalopathy

Mikko Muona,[1,2,3,4,27,28] Ryosuke Ishimura,[5,6,28] Anni Laari,[2,3,4,28] Yoshinobu Ichimura,[5] Tarja Linnankivi,[7] Riikka Keski-Filppula,[8,9,10] Riitta Herva,[11] Heikki Rantala,[8,9,12] Anders Paetau,[13] Minna Pöyhönen,[14] Miki Obata,[5] Takefumi Uemura,[15] Thomas Karhu,[2,3,4] Norihisa Bizen,[16] Hirohide Takebayashi,[16] Shane McKee,[17] Michael J. Parker,[18] Nadia Akawi,[19] Jeremy McRae,[19] Matthew E. Hurles,[19] the DDD Study, Outi Kuismin,[8,9,10] Mitja I. Kurki,[1,20,21,22] Anna-Kaisa Anttonen,[2,3,4,14] Keiji Tanaka,[6] Aarno Palotie,[1,19,21,22,23,24,25,26] Satoshi Waguri,[15] Anna-Elina Lehesjoki,[2,3,4,29,*] and Masaaki Komatsu[5,29,*]

The ubiquitin fold modifier 1 (UFM1) cascade is a recently identified evolutionarily conserved ubiquitin-like modification system whose function and link to human disease have remained largely uncharacterized. By using exome sequencing in Finnish individuals with severe epileptic syndromes, we identified pathogenic compound heterozygous variants in *UBA5*, encoding an activating enzyme for UFM1, in two unrelated families. Two additional individuals with biallelic *UBA5* variants were identified from the UK-based Deciphering Developmental Disorders study and one from the Northern Finland Intellectual Disability cohort. The affected individuals (n = 9) presented in early infancy with severe irritability, followed by dystonia and stagnation of development. Furthermore, the majority of individuals display postnatal microcephaly and epilepsy and develop spasticity. The affected individuals were compound heterozygous for a missense substitution, c.1111G>A (p.Ala371Thr; allele frequency of 0.28% in Europeans), and a nonsense variant or c.164G>A that encodes an amino acid substitution p.Arg55His, but also affects splicing by facilitating exon 2 skipping, thus also being in effect a loss-of-function allele. Using an in vitro thioester formation assay and cellular analyses, we show that the p.Ala371Thr variant is hypomorphic with attenuated ability to transfer the activated UFM1 to UFC1. Finally, we show that the CNS-specific knockout of *Ufm1* in mice causes neonatal death accompanied by microcephaly and apoptosis in specific neurons, further suggesting that the UFM1 system is essential for CNS development and function. Taken together, our data imply that the combination of a hypomorphic p.Ala371Thr variant in *trans* with a loss-of-function allele in *UBA5* underlies a severe infantile-onset encephalopathy.

Post-translational modifications through attachment of ubiquitin or ubiquitin-like proteins (UBLs) are involved in various biological processes.[1] There are eight human UBL-conjugating systems, in which each UBL is attached to specific proteins or lipids usually through three-step cascades involving E1-, E2-, and E3-like enzymes.[2] UBA5 is an E1-like (activating) enzyme for the most recently identified UBL, UFM1 (Figure 1A).[4] All molecules (i.e., UFM1, UBA5, UFC1, UFL1, and UFSP2) involved in conjugation of UFM1 to intracellular proteins (ufmylation) are conserved in metazoa and plants, but not in yeast, suggesting important roles in multicellular organisms. In fact, the ubiquitously expressed UFM1 system has an essential role in erythroid differentiation in mice,[7,8] plays a crucial role in breast

[1]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki 00290, Finland; [2]Folkhälsan Institute of Genetics, Helsinki 00290, Finland; [3]Neuroscience Center, University of Helsinki, Helsinki 00290, Finland; [4]Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki 00290, Finland; [5]Department of Biochemistry, Niigata University Graduate School of Medical and Dental Sciences, Chuo-ku, Niigata 951-8510, Japan; [6]Laboratory of Protein Metabolism, The Tokyo Metropolitan Institute of Medical Science, Setagaya-ku, Tokyo 156-8506, Japan; [7]Department of Child Neurology, Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki 00290, Finland; [8]PEDEGO Research Unit, University of Oulu, Oulu 90014, Finland; [9]Medical Research Center Oulu, University of Oulu, Oulu 90014, Finland; [10]Department of Clinical Genetics, Oulu University Hospital, Oulu 90029, Finland; [11]Department of Pathology, Cancer and Translational Medicine Research Unit, Medical Research Center Oulu (MRC Oulu), Oulu University Hospital and University of Oulu, Oulu 90014, Finland; [12]Department of Children and Adolescents, Division of Paediatric Neurology, Oulu University Hospital, Oulu 90029, Finland; [13]Department of Pathology, University of Helsinki and Helsinki University Central Hospital, Helsinki 00290, Finland; [14]Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki 00290, Finland; [15]Department of Anatomy and Histology, Fukushima Medical University School of Medicine, Hikarigaoka, Fukushima 960-1295, Japan; [16]Division of Neurobiology and Anatomy, Niigata University Graduate School of Medical and Dental Sciences, Chuo-ku, Niigata 951-8510, Japan; [17]Department of Genetic Medicine, Belfast City Hospital, Belfast BT9 7AB, UK; [18]Sheffield Children's Hospital NHS Foundation Trust, Western Bank, Sheffield S10 2TH, UK; [19]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK; [20]Neurosurgery of NeuroCenter, Kuopio University Hospital, Kuopio 70029, Finland; [21]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA; [22]Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02141, USA; [23]Stanley Center for Psychiatric Research, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02141, USA; [24]Program in Genetics and Genomics, Biological and Biomedical Sciences, Harvard Medical School, Boston, MA 02114, USA; [25]Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA 02114, USA; [26]Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA
[27]Present address: Blueprint Genetics, Helsinki 00290, Finland
[28]These authors contributed equally to this work
[29]These authors contributed equally to this work
*Correspondence: anna-elina.lehesjoki@helsinki.fi (A.-E.L.), komatsu-ms@med.niigata-u.ac.jp (M.K.)
http://dx.doi.org/10.1016/j.ajhg.2016.06.020.

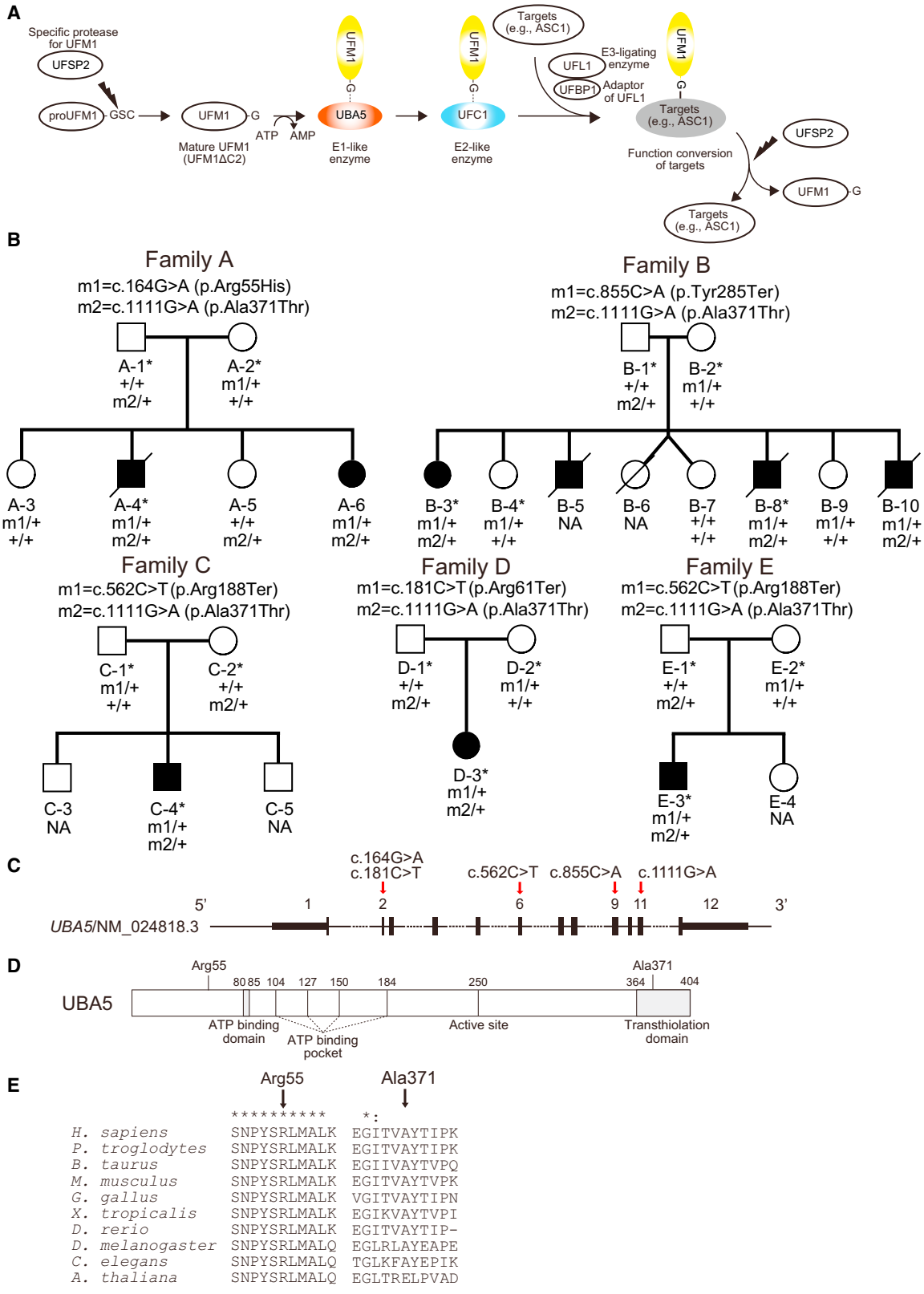

Figure 1. Compound Heterozygous Variants in *UBA5*

(A) A schematic of the UFM1 ubiquitin-like modifier cascade. UFM1 is synthesized in a precursor form and cleaved at the C terminus by specific protease, UFSP2.[3] The E1-like enzyme UBA5 activates mature UFM1 (UFM1ΔC2), forming a high-energy thioester bond. The activated UFM1 is then transferred to an E2-like (conjugating) enzyme, UFC1, through a similar thioester linkage.[4] Finally, UFM1 is covalently conjugated (ufmylated) with cellular proteins such as UFM1-binding protein 1 (UFBP1, official symbol DDRGK1) and a nuclear

(legend continued on next page)

cancer development,[5] and is implicated in cellular stress response[8–11] (UFM1 system is reviewed in Daniel and Liebau[12]). However, the role of the UFM1 system in the central nervous system (CNS) has not been studied, and the mechanism by which UFM1 system executes its functions is largely unknown.

In this study, we identified pathogenic compound heterozygous *UBA5* (MIM: 610552) variants in nine affected individuals who are from five unrelated families and show early infantile-onset encephalopathy. Functional analysis of the mutants suggests that reduced UBA5 activity leading to impaired UFM1 system underlies this syndrome.

As part of a study that aimed to identify genetic causes underlying severe infantile-onset epileptic syndromes in 30 Finnish individuals (A.L., unpublished data), we exome sequenced an index case (A-4) and his parents in family A (Figure 1B). Informed consent was obtained from the parents and the study was approved by an institutional review board at the Helsinki University Central Hospital. Whole-exome capture (Agilent SureSelect Human All Exon 50Mb V3), sequencing (Illumina HiSeq 2000; performed at the Wellcome Trust Sanger Institute), sequence read alignment to hs37d5 reference genome (based on GRCh37), and variant calling (Genome Analysis Tool Kit [GATK] HaplotypeCaller v. 3.3)[13–15] was done as described previously with minor modifications.[16] Given that the index case subject A-4 has an affected sister (A-6) in family A (Figure 1B), we analyzed the exome data of A-4 (see Table S1 for sequencing metrics) primarily for rare (<1% allele frequency) potentially deleterious autosomal-recessive variants including missense, nonsense, splice site, in-frame insertion and deletion, and frameshift variants based on Variant Effect Predictor[17] annotations in CCDS genes (Ensembl release 78). We also assessed the possibility of parental mosaicism by calling de novo variants using DeNovoGear,[18] and additionally, we analyzed heterozygous, potentially deleterious variants absent from population variant databases separately without using the de novo variant caller. The following databases were used to determine population allele frequencies: the ~60,000 exomes of the Exome Aggregation Consortium (ExAC; v.0.3),[19] phase 3 release of the 1000 Genomes project[20] (2,535 individuals), and Exome Variant Server (EVS) of the NHLBI GO Exome Sequencing Project (v.0.0.25; 6,503 individuals). All candidate variants were confirmed by bidirectional Sanger sequencing, and segregation analysis was performed on available family members.

The only variants passing the filtering under recessive model in affected individual A-4 were compound heterozygous missense variants c.164G>A (p.Arg55His; affects also splicing, see below) and c.1111G>A (p.Ala371Thr) in *UBA5* (GenBank: NM_024818.3, Ensembl: ENST00000356232; Figures 1B–1E and S1A, Tables S2 and S3). Sanger sequencing of additional family members revealed that these variants were compound heterozygous also in the affected sister but not in two unaffected sisters (Figure 1B). Analysis for variants causing the disease in the two siblings due to parental mosaicism did not yield any candidate variants (data not shown).

In addition to the above-mentioned cohort of 30 affected individuals sequenced at the Wellcome Trust Sanger Institute, we have exome sequenced and analyzed in-house six additional families (five of which are Finnish) with severe epileptic syndromes. One of these (family B; Figure 1B), ascertained by clinicians (H.R., M.P., R.K.-F., and R.H.) at the Oulu University Hospital Finland, has four affected individuals with clinical features similar to the affected individuals in family A. After informed consent was given by parents of family B, exome capture (Nimblegen SeqCap EZ Human Exome Library v.2.0), sequencing (Illumina HiSeq 1500; performed at the Institute for Molecular Medicine Finland), sequence read alignment to hg19 (GRCh37), and variant calling (samtools)[21] was done as described previously with minor modifications.[22] Exome sequencing of two affected siblings (B-3 and B-8), their parents, and one unaffected sibling (B-4) revealed compound heterozygous *UBA5* variants c.855C>A (p.Tyr285Ter) and c.1111G>A (p.Ala371Thr) in the affected siblings (Figures 1B–1E and S1B, Tables S1–S3). No other variants passed filtering under the recessive model (Table S2). No candidate variants were identified when assessing the possibility of parental mosaicism (data not shown). Sanger sequencing of one additional affected individual and two additional unaffected siblings confirmed autosomal-recessive segregation of the *UBA5* variants with the disease in the family (Figure 1B).

To attempt to identify additional individuals with biallelic *UBA5* variants, we used GeneMatcher website[23] and accessed data on 178 exomes or whole genomes of epileptic encephalopathy cases generated in EuroEPINOMICS Rare Epilepsy Syndromes consortium, 455 exomes from the Northern Finland Intellectual Disability cohort, as well as the exomes from the UK-based Deciphering Developmental Disorders (DDD) study.[24,25] In the DDD study with more than 4,000 exome-sequenced families with developmental disorders, of which 3,072 were without genetic diagnosis after initial

receptor coactivator, ASC1 (official symbol TRIP4) via UFL1 (E3-ligating enzyme).[5,6] The conjugates are cleaved by UFSP2,[3] implying the reversibility of the UFM1 conjugating system.
(B) Pedigrees of five families with biallelic variants in *UBA5*. Variants present in each family are shown above the pedigrees. Exome-sequenced individuals are marked with asterisks. Plus sign (+) indicates wild-type.
(C) A schematic of the exon structure of *UBA5* showing the locations of the variants.
(D) A schematic of the domain structure of UBA5 protein.
(E) ClustalX alignment of the Arg55 and Ala371 residues of UBA5 in metazoa and plants. Asterisks (*) and colons (:) indicate fully conserved and highly conserved residues, respectively.

analysis,[24] we identified two unrelated individuals (C-4 of British and D-3 of Northern Irish and Romanian ancestry; Figure 1B) who are similarly affected to siblings in families A and B and are compound heterozygous for rare *UBA5* variants. Exome data of C-4 and D-3 did not reveal other plausible candidate variants (data not shown). Families C and D were included in the DDD study after informed consent, and the study has an UK Research Ethics Committee approval (10/H0305/83 granted by the Cambridge South REC and GEN/284/12 granted by the Republic of Ireland REC). Both C-4 and D-3 are compound heterozygous for the c.1111G>A (p.Ala371Thr) variant and a nonsense variant (C-4: c.562C>T [p.Arg188Ter]; D-3: c.181C>T [p.Arg61Ter]) (Figures 1B and S2, Table S3). Finally, within the Northern Finland Intellectual Disability cohort, we identified one affected individual (E-3) compound heterozygous for c.1111G>A (p.Ala371Thr) and c.562C>T (p.Arg188Ter), the same variant combination as in C-4 (Figures 1B and S2, Table S3).

The p.Ala371Thr substitution, present in heterozygous state in all five families, has an allele frequency of 0.19% in the total of approximately 60,000 individuals in the ExAC database, 0.28% in non-Finnish Europeans, and 0.46% in Finns, with no homozygotes identified. There is one heterozygous carrier for p.Arg55His in ExAC and two heterozygous carriers for p.Arg188Ter in EVS, whereas the p.Arg61Ter and p.Tyr285Ter changes are novel. The missense substitutions occur at residues conserved down to *C. elegans* (p.Ala371Thr) or *A. thaliana* (p.Arg55His; Figure 1E) and are predicted to be deleterious by all four in silico methods used (Table S3).

Because p.Ala371Thr is present at low frequency in the general population, we assessed the probability for observing rare (<1%) biallelic *UBA5* variants in five unrelated families in the study populations. We employed a recently established method, recessiveStats, developed as part of the DDD study,[24] in which for a given gene, the number of unrelated cases with one of three possible biallelic genotype classes is tallied after filtering exome variant data for rare (<1%) variants. The three biallelic genotype classes are loss-of-function/loss-of-function, loss-of-function/functional, and functional/functional, with functional variants defined as those altering amino acid sequence but not likely to cause complete loss of function, e.g., missense variants. Then the probability to observe n number of unrelated individuals with such biallelic genotypes in the study cohort is determined after calculating the expected frequency of the genotypes based on the summed allele frequencies of rare (<1%), functional, or loss-of-function variants in the ExAC database (see original publication[24] for methodological details and Table S4 for analysis details in this study). The statistical analysis was performed separately for the three genotype classes. For observing four independent families (B–E) with compound heterozygous functional (p.Ala371Thr) and loss-of-function variants, we obtained a p value of 3.30×10^{-10}, strongly indicating that our datasets are enriched for causal, compound heterozygous *UBA5* variants (Table S4). The p value remains statistically significant after conservative correction for multiple testing for the three genotype classes and for the total number of annotatable protein coding genes based on GENCODE release 19 ($0.05/3/17{,}370 = 9.60 \times 10^{-7}$). Notably, even though affected siblings in family A were not tallied in the functional/loss-of-function genotype class, they have in effect a combination of a missense and a loss-of-function variant, because the second *UBA5* variant in this family, c.164G>A (p.Arg55His), encodes UBA5 mutant with severely reduced enzymatic activity and also affects splicing (see below). Finally, to exclude the presence of p.Ala371Thr/loss-of-function variant genotype in the general population, we have queried genotype information in >75,000 exomes of control individuals with no severe pediatric diseases (60,706 exomes in ExAC, 10,490 Finnish exomes of which ~7,000 are not in ExAC, and ~8,000 parents from the DDD study). In these datasets, none of the few carriers of loss-of-function variants in *UBA5* also have p.Ala371Thr.

All affected individuals in non-consanguineous families A–E presented in early infancy with irritability, and most had pronounced dystonic movements (e.g., axial hyperextension, head version, tonic upward gaze deviation, and pronation of arms and legs) and truncal hypotonia and they later developed spasticity (Table S5). In families A, B, C, and E, epileptic seizures, including myoclonic jerks (family A and B) and infantile spasms (A–C), started in infancy. Individual D-3, who is currently 5 years and 7 months of age, has not had seizures and EEG has been normal. The affected individuals did not reach any motor milestones and had severe intellectual disability, besides D-3 who is considered to have moderate intellectual deficit. The siblings in family B and individual E-3 also showed progressive growth failure. All individuals developed progressive microcephaly, besides E-3 who had occipito-frontal circumference of −2 SD at 9 years, which does not reach the applied diagnostic threshold for microcephaly of −3 SD. The affected individuals did not share any distinctive dysmorphic features, but individuals A-4 and C-4 presented with expressionless face, tented upper lip, long and deep philtrum, micrognathia, and puffy hands and feet with tapering fingers (Figure S3). Brain MRI (performed in seven individuals) findings were subtle and included mildly delayed myelination, slight T2-hyperintensity in thalami (Figure S4), thalamic volume reduction, and mild cerebral or cerebellar atrophy. Four out of nine affected individuals have died at 5, 12, 16, and 21 years of age, and five, aged 3–42 years, are alive.

All affected individuals had normal karyotype. Microarray-based analyses on individuals A-4, A-6, C-4, and D-3 had yielded inherited microdeletions or microduplications interpreted as benign. Sequencing of the *SLC16A2* (MIM: 300095) and *ARX* (MIM: 300382) in case A-4 had not revealed pathogenic variants.

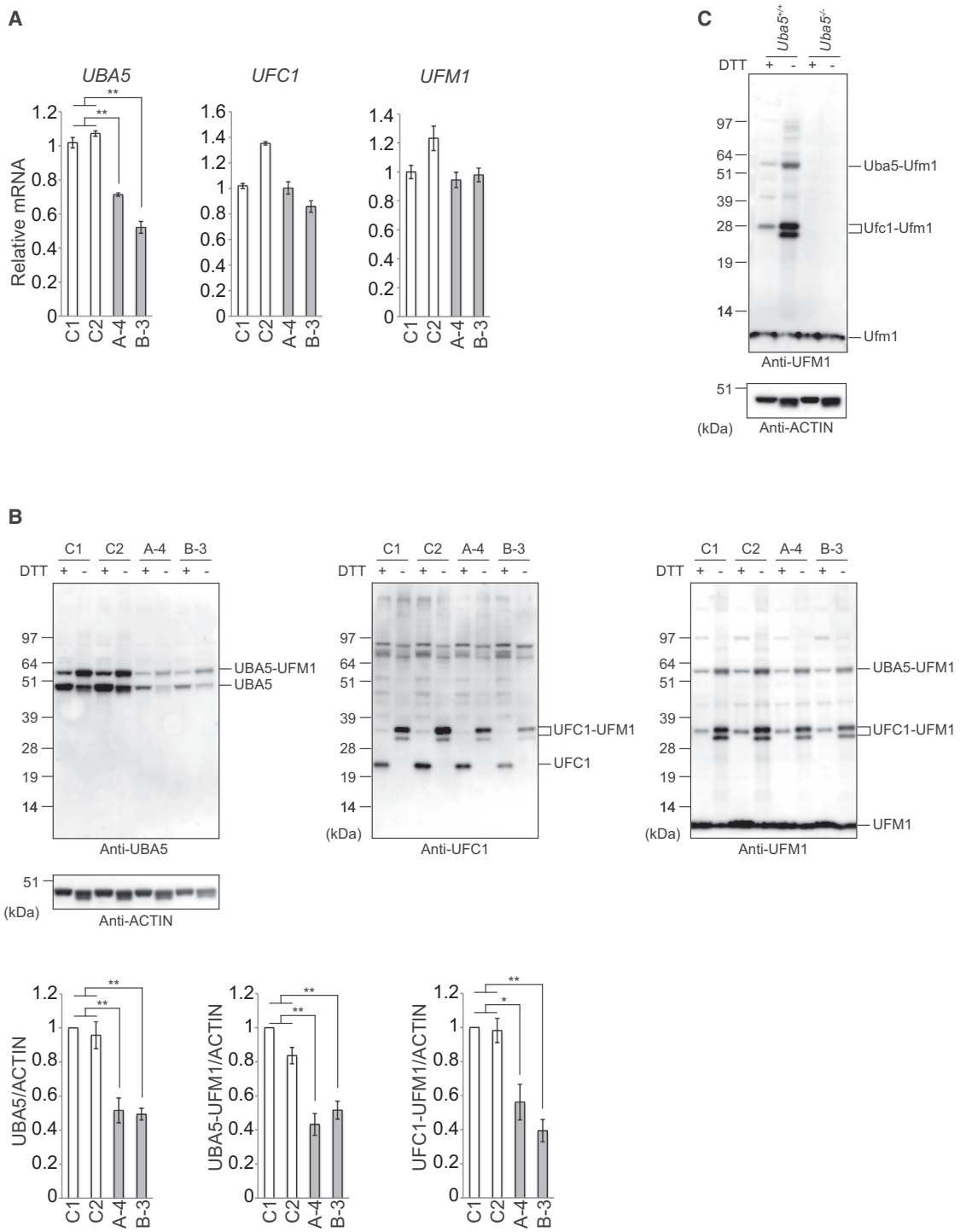

Figure 2. Defective E1 Activity of UBA5 in Fibroblasts Derived from Subjects with Pathogenic Biallelic *UBA5* Variants

(A) Quantitative real-time PCR analyses of *UBA5*, *UFC1* (MIM: 610554) and *UFM1* (MIM: 610553) in case (A-4 and B-3) and control (C1: female, age 26; C2: female, age 43) primary skin fibroblasts. Using a Transcriptor First Strand cDNA Synthesis Kit (Roche Applied Science), cDNA was synthesized from 1 μg of total RNA extracted from indicated fibroblasts. Quantitative PCR was performed using LightCycler 480 Probes Master (Roche Applied Science) in a LightCycler 480 (Roche Applied Science). Signals were assessed relative to that of *GAPDH* (MIM: 138400). Values were normalized to the amount of mRNA in control C1. The experiments were performed three times. The sequences of the primers are shown in Table S7. Statistical analysis was performed using the unpaired t test (Welch test). Data are means ± SE. **$p < 0.01$.

(B and C) Immunoblot analysis of UBA5, UFC1, and UFM1 with reducing and nonreducing samples that were prepared from fibroblasts of affected individuals and human controls (B) and mouse embryonic fibroblasts (C). Indicated fibroblasts were lysed with ice-cold TNE buffer (10 mM Tris-HCl [pH 7.5], 1% Nonidet P-40, 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid [EDTA], and protease inhibitors). Samples were prepared with NuPAGE-loading buffer in presence or absence of DTT, separated using a NuPAGE system (Life Technologies) on 4%–12% Bis-Tris gels in MOPS-SDS buffer, and then transferred to a polyvinylidene difluoride (PVDF) membrane. Mouse

(legend continued on next page)

The neuropathological findings in A-4, B-5, B-8, and B-10 were mild and nonspecific (data not shown). White matter volume seemed slightly reduced and corpus callosum was somewhat thin, but no signs of an ongoing leukodystrophy could be found. Slight diffuse atrophy could be seen in thalami, brain stem, and cerebellum. The density of especially mediodorsal and reticular neurons was diminished in thalami, and some postnecrotic calcified neurons were seen. In the mesencephalon there was slight neuronal loss and vacuolization in the periaqueductal gray matter, pontine basis was slightly thinned, and cerebello-olivary fibers were thinned in medulla oblongata. Cerebellar cortex showed slightly narrowed molecular and granular cell layers and marginal drop-out of Purkinje cells. Dentate nucleus harbored several pyknotic neurons. Both the clinical features and neuropathological findings were non-specific, resembling the pattern seen in various progressive encephalopathies.

We studied the consequence of the *UBA5* variants on RNA level in primary skin fibroblasts obtained from affected individuals A-4 and B-3, who are both heterozygotes for the missense substitution c.1111G>A (p.Ala371Thr). The other variant in B-3, the nonsense variant c.855C>A (p.Tyr285Ter) in the exon 9 (Figure 1C), is predicted to result in nonsense-mediated decay (NMD). As expected, we noted an approximately 50% reduction in *UBA5* mRNA levels in B-3 fibroblasts (Figure 2A). Sequencing of *UBA5* cDNA in B-3 fibroblasts revealed that the only allele present on RNA level is c.1111G>A (Figure S5A). We observed an approximately 30% reduction in mRNA levels also in fibroblasts of A-4 (Figure 2A), whose other heterozygous variant is c.164G>A. This change, which is predicted to cause the p.Arg55His substitution, occurs in the third nucleotide in the 5′ end of exon 2 (Figure S5B) and thus possibly affects splicing. Capillary sequencing of cDNA amplified using primers in exons 1 and 3 implied that the minority of sequences are from the c.164G>A mutant allele (Figure S5B). Suggesting that the variant facilitates exon 2 skipping, RT-PCR showed a weak band whose size and sequence corresponds to a *UBA5* transcript where exon 2 is skipped (Figures S5C and S5D). This would lead to a frameshift and consequent NMD. Confirming further that the *UBA5* allele with c.164G>A is expressed at lower levels, cDNA sequencing of exon 11 revealed that the majority of *UBA5* mRNA has c.1111G>A (Figure S5B).

We observed that expression of not only the *UBA5* mRNA but also the UBA5 protein was lower in fibroblasts of affected individuals than of control subjects (Figure 2B). By immunoblot analysis with non-reducing samples, we detected the intermediates of UFM1-UBA5 and of UFM1-UFC1. As expected, the level of UFM1-UBA5 intermediates was significantly lower in fibroblasts of affected individuals compared to control subjects (Figure 2B). Likewise, the formation of the UFM1-UFC1 intermediate declined (Figure 2B). Surprisingly, UFM1 conjugates were hardly detected in human fibroblasts regardless of the genotypes (Figure 2B). We confirmed a similar pattern in mouse embryonic fibroblasts (Figure 2C).

To test whether the missense variants affect the E1-like activity of UBA5 in cells, we used a UBA5 mutant in which the active site, cysteine (Cys250), was substituted with serine (termed UBA5^{Cys250Ser}). When the cysteine residue at the active site of E1 and E2 enzymes is replaced with serine, an *O*-ester bond instead of a thioester bond is formed with its respective modifier proteins, and the intermediates become stable even under reducing conditions.[4] In addition, to exclude the effect of endogenous UBA5, *UBA5* in HEK293T cells was deleted by CRISPR/Cas9 technology (Figures 3A and S6). We expressed a FLAG-tagged UBA5^{Cys250Ser} (FLAG-UBA5^{Cys250Ser}) together with MYC-tagged UFM1ΔC2, a mature form of UFM1 with the glycine residue at the C terminus (Figure 1A), in *UBA5*-deficient HEK293T cells, and analyzed the cell lysates by immunoblot assay. An intermediate between FLAG-UBA5^{Cys250Ser} and MYC-UFM1ΔC2 was clearly recognized (Figure 3A). We next analyzed UBA5 constructs carrying both p.Cys250Ser and one of the missense variants observed in affected individuals (FLAG-UBA5$^{Arg55His/Cys250Ser}$ or FLAG-UBA5$^{Ala371Thr/Cys250Ser}$). FLAG-UBA5$^{Arg55His/Cys250Ser}$ had lower ability to form the intermediate with MYC-UFM1ΔC2 than UBA5^{Cys250Ser} (Figure 3A). With FLAG-UBA5$^{Ala371Thr/Cys250Ser}$, we did not observe a statistically significant decrease in the intermediate formation (Figure 3A). Next, we examined the effect of UBA5 mutants on transferring the activated UFM1 to the E2 enzyme UFC1 (see Figure 1A) because the Ala371 residue is located within the C-terminal transthiolation domain critical for the transfer of UFM1 to UFC1 (Figure 1D).[26] We detected the intermediate between FLAG-UFC1^{Cys116Ser} and MYC-UFM1ΔC2 when wild-type UBA5 was expressed in *UBA5*$^{-/-}$ HEK293T cells (Figure 3B). By contrast, in the case of expressing either UBA5^{Arg55His} or UBA5^{Ala371Thr} mutant, formation of such intermediates was suppressed (Figure 3B).

In good agreement with these results, in vitro thioester formation assay with recombinant proteins revealed the decreased E1 activity of UBA5^{Arg55His} and UBA5^{Ala371Thr}

monoclonal anti-actin antibody (Chemicon International cat# MAB1501R), rabbit monoclonal anti-UFM1 antibody (Abcam cat# ab109305, RRID: AB_10864675), anti-UBA5 antibody,[4] and anti-UFC1 antibody[4] were used for immunodetection. The immunoreactive bands were detected by LAS-4000 (GE Healthcare UK). In the cases of samples prepared without DTT, the intermediates corresponding to UBA5-UFM1 and UFC1-UFM1 were clearly detected. Bar graphs indicate the quantitative densitometric analyses using Multi Gauge Version 3.2 Image software (Fuji Film) of UBA5, UBA5-UFM1, and UFC1-UFM1 intermediates relative to ACTIN.
Statistical analysis was performed using the unpaired t test (Welch test). The data represent the means ± SE of five separate experiments. *$p < 0.05$ and **$p < 0.01$.

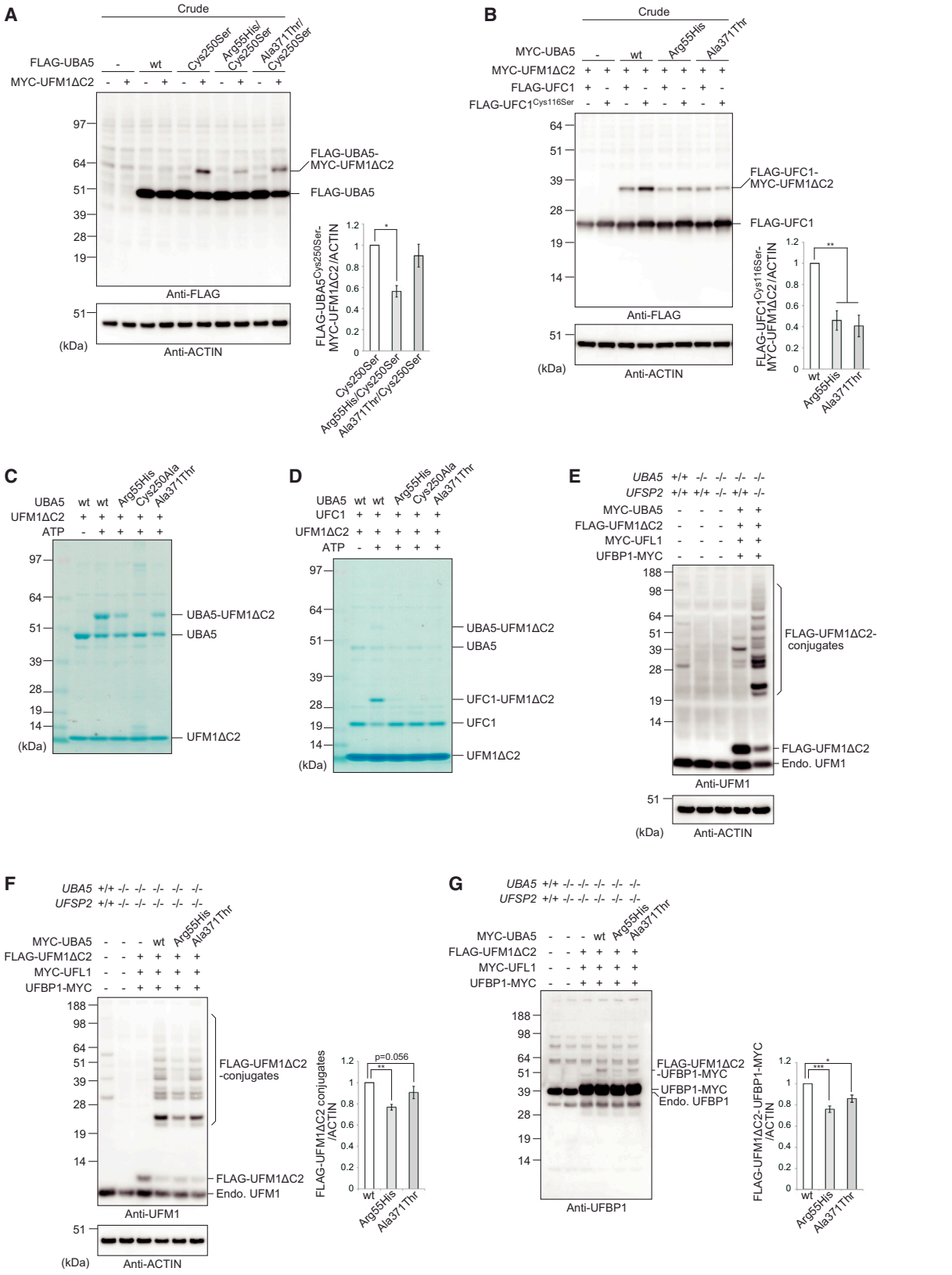

Figure 3. Impaired Function of UBA5 Mutants

(A and B) Immunoblot assay of UBA5 mutant p.Cys250Ser and double mutants p.Arg55His/p.Cys250Ser and p.Ala371Thr/p.Cys250Ser in *UBA5*$^{-/-}$ HEK293T cells. Indicated constructs (0.1 μg for UBA5, 0.5 μg for UFC1, and 2 μg for UFM1ΔC2) were expressed in *UBA5*-deficient HEK293T cells. 24 hr after transfection, the cell lysates were subjected to immunoblot analysis with indicated antibodies as described in Figure 2B. Bar graphs indicate the quantitative densitometric analyses of UBA5-UFM1 and UFC1-UFM1 intermediates

(legend continued on next page)

mutants. As shown in Figure 3C, the UBA5-UFM1ΔC2 intermediate was formed in an ATP-dependent manner in vitro. Although both mutants still had ability to form the intermediate, their activity was weaker compared to that of wild-type UBA5 (Figure 3C). In addition, the UBA5^{Arg55His} and UBA5^{Ala371Thr} mutants failed to transfer the activated UFM1 to UFC1 in the initial stage of reaction (Figure 3D). In vitro kinetics analysis of multiple time points up to 60 min of reaction revealed that the UFC1-UFM1ΔC2 intermediate was formed in the case of both mutants, but that, compared to wild-type UBA5, their reaction rates were slower (Figure S7). Whereas the activity of UBA5^{Arg55His} was approximately half of that with wild-type UBA5, UBA5^{Ala371Thr} had approximately 70%–80% activity (Figure S7).

Finally, we examined whether UBA5^{Arg55His} and UBA5^{Ala371Thr} mutants suppress the UFM1 conjugate formation in cells. HEK293T cells deficient for both *UBA5* and *UFSP2*, of which latter is the unique de-conjugating enzyme for UFM1 conjugates (see Figure 1A), were generated by CRISPR/Cas9 technology (*UBA5*$^{-/-}$;*UFSP2*$^{-/-}$; Figure S8). As shown in Figure 3E, only a few UFM1 conjugates were detected when we expressed E3-ligating enzyme UFL1, its adaptor protein UFBP1[5] (see Figure 1A), and UFM1ΔC2 together with wild-type UBA5 in single knockout *UBA5*$^{-/-}$ HEK293T cells. Remarkably, concomitant loss of the de-conjugating enzyme UFSP2 dramatically increased both number and amount of UFM1 conjugates in the *UBA5*-deficient HEK293T cells, indicating that most UFM1 conjugates are de-conjugated by UFSP2. We next evaluated the effect of UBA5 mutants (p.Arg55His or p.Ala371Thr) in this experimental setting. The UFM1 conjugate formation in the *UBA5*$^{-/-}$;*UFSP2*$^{-/-}$ double knock-out (DKO) cells expressing UBA5^{Arg55His} was markedly suppressed (Figure 3F) ($p = 0.002$; ratio of actin-normalized quantity of UFM1 conjugates with UBA5^{Arg55His} to that with wild-type: 0.77, with 95% CI 0.72–0.82), whereas UBA5^{Ala371Thr} only showed suggestive decrease ($p = 0.056$; ratio of actin-normalized quantity of UFM1 conjugates with UBA5^{Ala371Thr} to that with wild-type: 0.91, with 95% CI 0.79–1.02) (Figure 3F). Because ufmylation of UFBP1 is required for its tight binding to UFL1 and in turn for promotion of the E3 ligase activity,[5] we further examined the level of the UFM1-UFBP1 conjugate in the DKO cells. We found that the conjugate formation was impaired in both DKO cells expressing either UBA5^{Arg55His} ($p = 0.0005$; ratio of actin-normalized quantity of UFM1-UFBP1 conjugate with UBA5^{Arg55His} to that with wild-type: 0.76, with 95% CI 0.70–0.82) or UBA5^{Ala371Thr} ($p = 0.013$; ratio of actin-normalized quantity of UFM1-UFBP1 conjugate with UBA5^{Ala371Thr} to that with wild-type: 0.86, with 95% CI 0.79–0.93) (Figure 3G). Taken together, we conclude that the UBA5 mutants exhibit decreased E1 activity with attenuated ability to transfer the activated UFM1 to UFC1, which might cause impaired UFM1 conjugate formation, with p.Arg55His exhibiting more pronounced defects.

Germline *Uba5*[7] or *Ufm1* (M.K., unpublished data) knockout is embryonic lethal in mice. We generated conditional knockout mice for *Ufm1* (*Ufm1*$^{f/f}$) and crossed them with transgenic mice expressing Cre recombinase under the control of the nestin promoter[27] (nestin-*Cre*) to create CNS-specific *Ufm1* knockout (*Ufm1*$^{f/f}$;nestin-*Cre*) mice (Figure S9). The *Ufm1*$^{f/f}$;nestin-*Cre* mice allowed us to examine the neuronal pathology associated with deficiency of the UFM1 system in vivo. At embryonic day (E) 14.5, UFM1 protein was almost absent in brain of *Ufm1*$^{f/f}$;nestin-*Cre* but not of control (*Ufm1*$^{f/+}$;nestin-*Cre*)

relative to ACTIN. Statistical analysis was performed using the unpaired t test (Welch test). The data represent the means ± SE of four separate experiments. *$p < 0.05$ and **$p < 0.01$.

(C and D) In vitro thioester formation assay of UFM1 by UBA5 (C) and of UFM1 by UFC1 (D). Recombinant GST-UFM1ΔC2, GST-UFC1, and GST-UBA5, as well as UBA5 mutants p.Arg55His (GST-UBA5^{Arg55His}), p.Ala371Thr (GST-UBA5^{Ala371Thr}), and p.Cys250Ala (negative control; GST-UBA5^{Cys250Ala}) were produced in *E. coli* and the recombinant proteins were purified by chromatography on Glutathione Sepharose 4B (GE Healthcare UK). After digestion of GST by PreScission Protease (GE Healthcare UK), the recombinant proteins were dialyzed against 50 mM BisTris (pH 6.5), 100 mM NaCl, 10 mM $MgCl_2$, and 0.1 mM DTT (reaction buffer). Most thioester formation reactions contained reaction buffer with 0.8 μg UFM1ΔC2 and some of the following: 5 mM ATP, 0.08 (for UFC1-UFM1 thioester formation assay) or 0.8 (for UBA5-UFM1 thioester formation assay) μg UBA5 or UBA5 mutants, and 0.8 μg UFC1. Reactions were incubated for 5 min at 25°C and stopped by the addition of NuPAGE-loading buffer lacking reducing agent, followed by 10 min incubation at 37°C, NuPAGE (4%–12% acrylamide gradient), and Coomassie brilliant blue staining. Data shown are representative of three separate experiments.

(E) Immunoblot assay to detect UFM1 conjugates. MYC-UBA5 (0.1 μg) was expressed in combination with indicated constructs (each 1 μg) in *UBA5*-deficient or *UBA5-UFSP2* double-deficient HEK293T cells. Cells were lysed by 200 μL of TNE, and the lysate was then centrifuged at 10,000 × *g* for 10 min at 4°C to remove debris. The supernatant was subjected to immunoblot analyses with indicated antibodies.

(F) Immunoblot assay to study the effect of UBA5 mutants on UFM1 conjugate formation. MYC-UBA5 or MYC-UBA5 mutants (0.1 μg) were expressed in combination with indicated constructs (each 1 μg) in *UBA5-UFSP2* double-deficient HEK293T cells. Cells were lysed by 200 μL of TNE, and the lysate was then centrifuged at 10,000 × *g* for 10 min at 4°C to remove debris. The supernatant was subjected to immunoblot analyses with indicated antibodies. Bar graph indicates the quantitative densitometric analyses of FLAG-UFM1 conjugates relative to ACTIN. Statistical analysis was performed using the unpaired t test (Welch test). The data represent the means ± SE of six separate experiments. *$p < 0.05$.

(G) Immunoblot assay to study the effect of UBA5 mutants on UFM1-UFBP1 conjugate formation. Transfection and subsequent immunoblot analysis were conducted as shown in (F). Bar graph indicates the quantitative densitometric analyses of FLAG-UFM1-UFBP1-MYC relative to ACTIN. Rabbit polyclonal anti-UFBP1 antibody[6] was used for immunodetection. Statistical analysis was performed using the unpaired t test (Welch test). The data represent the means ± SE of six separate experiments. *$p < 0.05$ and ***$p < 0.001$.

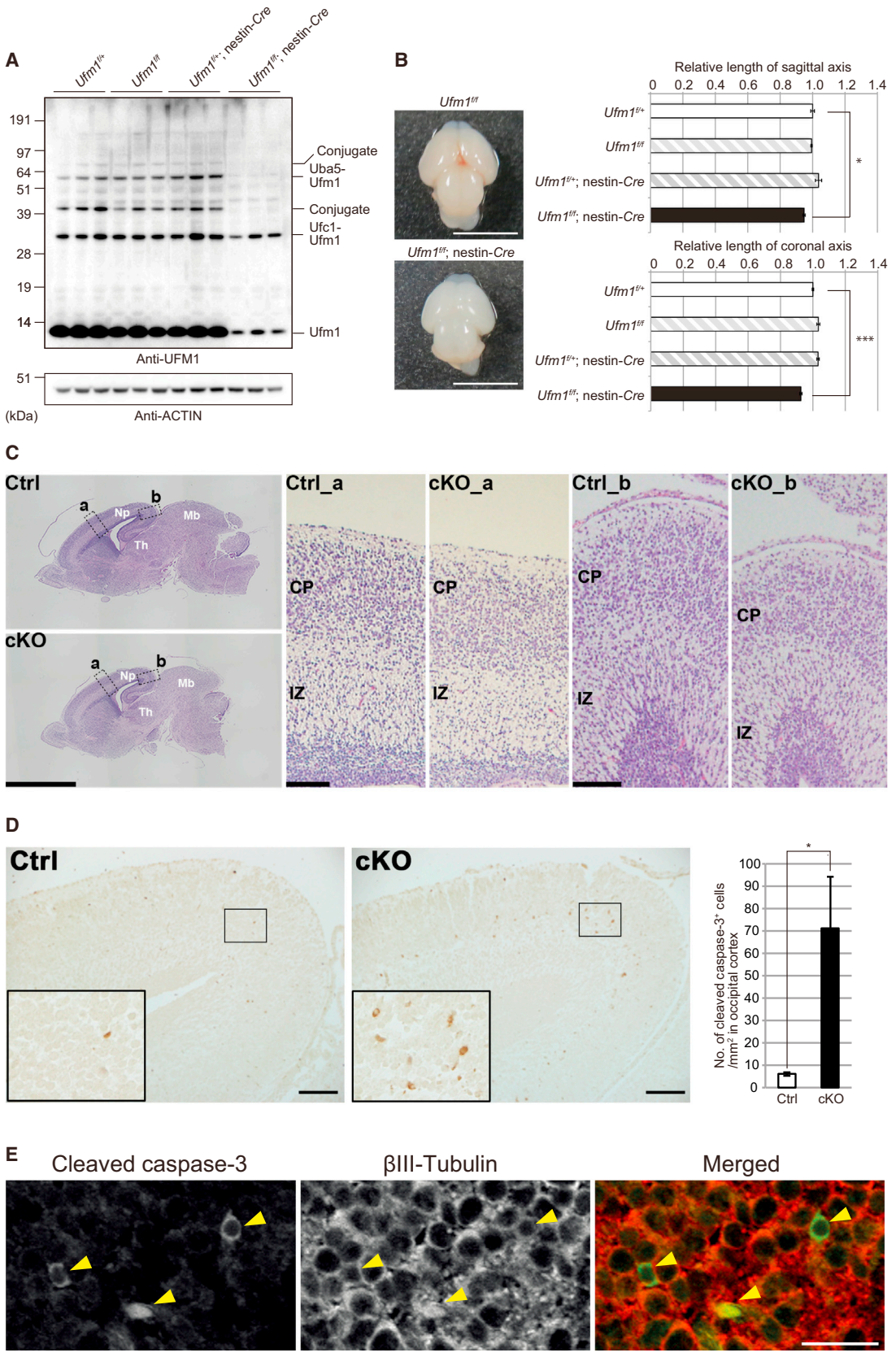

Figure 4. Loss of UFM1 in Central Nervous System Causes Microcephaly

(A) Immunoblot analysis of UFM1 in mice with indicated genotypes. Mice were delivered by caesarean section at E18.5, and then mouse brains were homogenized in 0.25 M sucrose, 10 mM 2-[4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid (HEPES) (pH 7.4), and

(legend continued on next page)

mice (Figure S10). Although a few UFM1 conjugates were detected in brains of control mice by immunoblotting with an anti-UFM1 antibody at E18.5, they were weak or undetectable in brains of mutant mice (Figure 4A). These results indicate impairment of the UFM1 conjugation system in the CNS of *Ufm1*$^{f/f}$;nestin-*Cre* mice. *Ufm1*$^{f/f}$;nestin-*Cre* mice were viable at birth and indistinguishable in appearance from their littermates. However, all *Ufm1*$^{f/f}$;nestin-*Cre* mice died within 1 day of birth (Table S6). Macroscopic anatomical analysis of the brains of mice that were delivered by Caesarean section at E18.5 revealed that *Ufm1*$^{f/f}$;nestin-*Cre* mice had microcephaly (Figure 4B). Both transverse and longitudinal distances in the mutant brains were significantly shorter than those of control brains (Figure 4B). Histological analysis using Karachi's hematoxylin and eosin staining showed that some parts of the brain, such as the occipital region of neopallium, midbrain, and thalamus, of *Ufm1*$^{f/f}$;nestin-*Cre* mice were consistently smaller than those of control mice (Figure 4C). However, no apparent abnormality in cellular organization was found in the mutant brain. To examine whether loss of UFM1 causes cell death, we carried out immunohistochemical analysis using an antibody against cleaved Caspase-3, a hallmark of apoptosis. A marked increase in the number of cleaved Caspase-3-positive cells was noted in the occipital region of neopallium in the *Ufm1*$^{f/f}$;nestin-*Cre* mice at E18.5 (Figure 4D), compared with control mice. Double immunofluorescence analysis with antibodies against βIII Tubulin, a neuronal marker, and cleaved Caspase-3 revealed that neurons underwent apoptosis (Figure 4E). We hardly detected such apoptotic cells in other brain regions of *Ufm1*$^{f/f}$;nestin-*Cre* mice at E18.5 (data not shown). Taken together, findings in the *Ufm1*$^{f/f}$;nestin-*Cre* mice suggest that an intact UFM1 system is pivotal for neuronal development and survival.

Here, we combined genetic, statistical, phenotypic, and functional data to show that the hypomorphic p.Ala371Thr variant in *trans* with a loss-of-function change in *UBA5* causes a severe, early-onset encephalopathy. The likelihood for observing biallelic missense/loss-of-function *UBA5* variants only by chance in the assessed disease cohorts was determined to be negligible. The assessment of UBA5 in fibroblasts of the affected individuals and the biochemical analyses with mutant UBA5 proteins corresponding to the two missense variants indicated that the compound heterozygous status in the affected individuals is not accompanied by a complete loss of UBA5 function. The p.Ala371Thr variant, which is present in all five families with altogether nine affected individuals, encodes UBA5 with only mildly reduced enzymatic activity. In contrast, the impact of the c.164G>A (p.Arg55His) variant on UBA5 function in family A is more pronounced. This variant is causing aberrant splicing, but this effect does not appear to be complete in cells of the affected individuals. However, given the low enzymatic activity of p.Arg55His UBA5 mutant protein, the function of UBA5 can be predicted to be severely compromised from this allele also when the gene is normally spliced. The nonsense variant p.Tyr285Ter in family B seems to undergo NMD based on analysis of *UBA5* mRNA expression and is thus likely to result in severely compromised enzymatic activity. It is likely that the gene product is degraded also in individuals C-4, D-3, and E-3 with nonsense variants. Thus, the affected individuals seem to have one variant with a major and one with a milder, hypomorphic effect on UBA5 function. It is plausible that biallelic, complete loss-of-function variants in *UBA5* would not be compatible with life as observed in the *Uba5*$^{-/-}$ mice.[7] Interestingly, the p.Ala371Thr variant with only mildly compromised UBA5 activity is present in considerable frequencies in Finns (0.46% allele frequency in ExAC) and

1 mM dithiothreitol (DTT). The homogenates were subjected to immunoblot analysis with indicated antibodies. Samples prepared from three mice with indicated genotype were loaded.
(B) A dorsal view of brains of *Ufm1*$^{f/f}$ and *Ufm1*$^{f/f}$;nestin-*Cre* mice delivered by Caesarean section at E18.5. Graphs show axial distance (from the anterior edge of cerebrum to posterior edge of mid brain) and maximal lateral distance of brains of indicated genotype mice. Data presented as mean ± SE of *Ufm1*$^{f/+}$ (n = 4), *Ufm1*$^{f/f}$ (n = 4), *Ufm1*$^{f/+}$;nestin-*Cre* (n = 5), and *Ufm1*$^{f/f}$;nestin-*Cre* (n = 6). Statistical analysis was performed using the unpaired t test. *$p < 0.05$ and ***$p < 0.001$.
(C) Histological analyses of brains of *Ufm1*$^{f/+}$;nestin-*Cre* (Ctrl) and *Ufm1*$^{f/f}$;nestin-*Cre* (cKO) mice. Embryos at E18.5 were delivered by Caesarean section, and their heads were fixed by immersion in 0.1 M phosphate buffer (pH 7.4) containing 4% paraformaldehyde and 4% sucrose. Each brain was carefully dissected and processed for paraffin embedding, and then 3 μm sagittal sections were prepared for haematoxylin and eosin staining. Images were captured with BZ-9000 (Keyence) and BX51 microscopes (Olympus). Boxed regions a and b in the neopallium are magnified and shown on the right as indicated. Note that the occipital region (b) of neopallium in the mutant brain is thinner than that in control, while the difference in the parietal region (a) is less apparent. Scale bars are 2 mm and 0.1 mm. Abbreviations are as follows: Np, neopallium; Mb, midbrain; Th, thalamus; IZ, intermediate zone; CP, cortical plate.
(D) Apoptotic cells in the occipital region of neopallium of *Ufm1*$^{f/+}$;nestin-*Cre* (Ctrl) and *Ufm1*$^{f/f}$; nestin-*Cre* (cKO) mice at E18.5. Sections prepared as described in (C) were immunostained by rabbit polyclonal anti-cleaved caspase-3 antibody (Cell Signaling Technology [CST] cat# 9661, RRID: AB_2314091; 1:500) as described previously.[28] Images were captured with BX53 microscope (Olympus). Each inset is a magnified image. Scale bars represent 100 μm. For quantification, the number of cleaved caspase-3-positive cells per unit area was calculated in each occipital cortex, which was defined as the cerebral cortex located posterior to the hippocampus. Statistical analysis was performed using the unpaired t test (n = 3 animals for each group). Data represent the means ± SE. *$p < 0.05$. The area was measured by NIH Image/ImageJ.
(E) Double-immunofluorescence analysis. Section of cKO brain (occipital region of neopallium) prepared as described in (C) was double-immunostained with anti-cleaved-caspase-3, mouse monoclonal anti-βIII Tubulin antibody (clone 5G8, Promega, 1:1,000), goat anti-mouse Alexa Fluor 594, and goat anti-rabbit Alexa Fluor 488 (Molecular Probes, 1:1,000). Images were captured with confocal FV1200 microscope (Olympus). Scale bar represents 20 μm.

other Europeans (0.28%). This implies that the syndrome associated with deficient UBA5 function should be encountered in the European population. Indeed, in addition to the five families identified in this study, individuals from two families of European ancestry with biallelic compound heterozygous variants in *UBA5* and a disease manifestation comparable to the affected individuals in our study have the p.Ala371Thr variant in combination with a loss-of-function *UBA5* variant (D. Bonneau, personal communication).

Notably, given its relatively mild effect on UBA5 function, the phenotypic consequences of p.Ala371Thr when occurring in homozygosity could be predicted to be mild. To search for individuals homozygous for p.Ala371Thr (no homozygotes were present in the databases used in exome data filtering, see above), we queried ~10,490 exomes from various Finnish population-based and disease cohorts (see SISu in Web Resources). In the exome data from the FINRISK cohort,[29] we identified one homozygous individual (V. Salomaa, personal communication) who is in his fifties and, based on data gathered from National Health registers, does not have any hospitalizations or chronic medication for any neurological or neuropsychiatric diseases. This observation suggests that, as described in this paper, p.Ala371Thr indeed contributes to a severe neurological phenotype only when it is in *trans* with a severe loss-of-function variant. This observation also has relevance for filtering strategies of exome variant data, because it exemplifies that homozygosity for a relatively rare missense variant in "control" individuals does not necessarily mean non-pathogenicity. During the review of this paper, Duan and colleagues reported a family with two siblings with compound heterozygous variants in *UBA5* (p.Lys310Glu and p.Arg246Ter) and a childhood-onset neurological disease with ataxia as the primary symptom.[30] Altogether, the existing data on different combinations of biallelic *UBA5* variants both in humans and in mouse models suggest that the phenotypic consequences range from embryonic lethality (biallelic loss-of-function in knockout mice), severe infantile-onset encephalopathy (this study), childhood-onset neurological disease (Duan et al.[30]), to possibly no or very mild phenotype.

The symptoms in individuals with biallelic *UBA5* variants are predominantly related to CNS. Previously, however, we and other groups have reported the function of the ubiquitously expressed UFM1 system outside the CNS.[5,7–11] For example, mice with a germline deletion of *Uba5*,[7] *Ufm1* (M.K., unpublished data), or *Ddrgk1* (*Ufbp1*)[8] are embryonic lethal and show a severe defect of erythroid differentiation. The affected individuals in our study, with partial loss of UBA5 function, do not present with anemia, suggesting that the remaining UBA5 activity is sufficient for normal hematopoiesis. Finally, to further support the role of UFM1 system beyond CNS, a suggestive association of a *UFSP2* (MIM: 611482; UFM1-specific peptidase) variant with Beukes hip dysplasia was recently established.[31]

Consistent with imaging and neuropathological observations in affected individuals, the analysis of CNS-specific *Ufm1* knockout mice revealed that dysfunction of the UFM1 system causes atrophy in several regions of the brain and results in neonatal death. Our data imply compromised UFM1 conjugate formation in the brains of the CNS-specific *Ufm1* knockout mice with neuronal apoptosis in restricted regions (i.e., the occipital region of neopallium). Therefore it is possible that spatiotemporally regulated and cell-specific ufmylation is necessary not only for prevention of neuronal cell death but also for neuronal development. Further studies assessing the specific function of UFM1 conjugation to its target proteins in CNS and elsewhere are warranted.

Accession Numbers

The raw aligned sequence reads of family A were submitted to the European Genome-phenome Archive by Wellcome Trust Sanger Institute under study accession IDs EGAS00001000190 and EGAS00001000386. Exome sequencing data from the DDD study are downloadable under study accession ID EGAS00001000775.

Supplemental Data

Supplemental Data include ten figures, seven tables, and Supplemental Acknowledgments and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.06.020.

Received: January 26, 2016
Accepted: June 22, 2016
Published: August 18, 2016

Web Resources

1000 Genomes, http://www.1000genomes.org
Clustal: Multiple Sequence Alignment, http://www.clustal.org/
European Genome-phenome Archive (EGA), https://www.ebi.ac.uk/ega
ExAC Browser, http://exac.broadinstitute.org/
GATK, https://www.broadinstitute.org/gatk/
GenBank, http://www.ncbi.nlm.nih.gov/genbank/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
OMIM, http://www.omim.org/
RRID, https://scicrunch.org/resources
SISu Project, www.sisuproject.fi
Variant Effect Predictor, http://useast.ensembl.org/Homo_sapiens/Tools/VEP

References

1. van der Veen, A.G., and Ploegh, H.L. (2012). Ubiquitin-like proteins. Annu. Rev. Biochem. *81*, 323–357.
2. Schulman, B.A., and Harper, J.W. (2009). Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. Nat. Rev. Mol. Cell Biol. *10*, 319–331.
3. Kang, S.H., Kim, G.R., Seong, M., Baek, S.H., Seol, J.H., Bang, O.S., Ovaa, H., Tatsumi, K., Komatsu, M., Tanaka, K., and Chung, C.H. (2007). Two novel ubiquitin-fold modifier 1

(Ufm1)-specific proteases, UfSP1 and UfSP2. J. Biol. Chem. *282*, 5256–5262.
4. Komatsu, M., Chiba, T., Tatsumi, K., Iemura, S., Tanida, I., Okazaki, N., Ueno, T., Kominami, E., Natsume, T., and Tanaka, K. (2004). A novel protein-conjugating system for Ufm1, a ubiquitin-fold modifier. EMBO J. *23*, 1977–1986.
5. Yoo, H.M., Kang, S.H., Kim, J.Y., Lee, J.E., Seong, M.W., Lee, S.W., Ka, S.H., Sou, Y.S., Komatsu, M., Tanaka, K., et al. (2014). Modification of ASC1 by UFM1 is crucial for ERα transactivation and breast cancer development. Mol. Cell *56*, 261–274.
6. Tatsumi, K., Sou, Y.S., Tada, N., Nakamura, E., Iemura, S., Natsume, T., Kang, S.H., Chung, C.H., Kasahara, M., Kominami, E., et al. (2010). A novel type of E3 ligase for the Ufm1 conjugation system. J. Biol. Chem. *285*, 5417–5427.
7. Tatsumi, K., Yamamoto-Mukai, H., Shimizu, R., Waguri, S., Sou, Y.-S., Sakamoto, A., Taya, C., Shitara, H., Hara, T., Chung, C.H., et al. (2011). The Ufm1-activating enzyme Uba5 is indispensable for erythroid differentiation in mice. Nat. Commun. *2*, 181.
8. Cai, Y., Pi, W., Sivaprakasam, S., Zhu, X., Zhang, M., Chen, J., Makala, L., Lu, C., Wu, J., Teng, Y., et al. (2015). UFBP1, a key component of the Ufm1 conjugation system, is essential for ufmylation-mediated regulation of erythroid development. PLoS Genet. *11*, e1005643.
9. Hertel, P., Daniel, J., Stegehake, D., Vaupel, H., Kailayangiri, S., Gruel, C., Woltersdorf, C., and Liebau, E. (2013). The ubiquitin-fold modifier 1 (Ufm1) cascade of *Caenorhabditis elegans*. J. Biol. Chem. *288*, 10661–10671.
10. Lemaire, K., Moura, R.F., Granvik, M., Igoillo-Esteve, M., Hohmeier, H.E., Hendrickx, N., Newgard, C.B., Waelkens, E., Cnop, M., and Schuit, F. (2011). Ubiquitin fold modifier 1 (UFM1) and its target UFBP1 protect pancreatic beta cells from ER stress-induced apoptosis. PLoS ONE *6*, e18517.
11. Zhang, Y., Zhang, M., Wu, J., Lei, G., and Li, H. (2012). Transcriptional regulation of the Ufm1 conjugation system in response to disturbance of the endoplasmic reticulum homeostasis and inhibition of vesicle trafficking. PLoS ONE *7*, e48587.
12. Daniel, J., and Liebau, E. (2014). The ufm1 cascade. Cells *3*, 627–638.
13. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
14. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 1–33.
15. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.
16. Muona, M., Berkovic, S.F., Dibbens, L.M., Oliver, K.L., Maljevic, S., Bayly, M.A., Joensuu, T., Canafoglia, L., Franceschetti, S., Michelucci, R., et al. (2015). A recurrent de novo mutation in KCNC1 causes progressive myoclonus epilepsy. Nat. Genet. *47*, 39–46.
17. McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., and Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics *26*, 2069–2070.
18. Ramu, A., Noordam, M.J., Schwartz, R.S., Wuster, A., Hurles, M.E., Cartwright, R.A., and Conrad, D.F. (2013). DeNovoGear: de novo indel and point mutation discovery and phasing. Nat. Methods *10*, 985–987.
19. Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., O'Donnell-Luria, A., Ware, J., Hill, A., Cummings, B., et al. (2015). Analysis of protein-coding genetic variation in 60,706 humans. bioRxiv http://dx.doi.org/10.1101/030338.
20. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
21. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
22. Sulonen, A.-M., Ellonen, P., Almusa, H., Lepistö, M., Eldfors, S., Hannula, S., Miettinen, T., Tyynismaa, H., Salo, P., Heckman, C., et al. (2011). Comparison of solution-based exome capture methods for next generation sequencing. Genome Biol. *12*, R94.
23. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum. Mutat. *36*, 928–930.
24. Akawi, N., McRae, J., Ansari, M., Balasubramanian, M., Blyth, M., Brady, A.F., Clayton, S., Cole, T., Deshpande, C., Fitzgerald, T.W., et al.; DDD study (2015). Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. Nat. Genet. *47*, 1363–1369.
25. Deciphering Developmental Disorders Study (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature *519*, 223–228.
26. Xie, S. (2014). Characterization, crystallization and preliminary X-ray crystallographic analysis of the human Uba5 C-terminus-Ufc1 complex. Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. *70*, 1093–1097.
27. Tronche, F., Kellendonk, C., Kretz, O., Gass, P., Anlag, K., Orban, P.C., Bock, R., Klein, R., and Schütz, G. (1999). Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. Nat. Genet. *23*, 99–103.
28. Horie, M., Watanabe, K., Bepari, A.K., Nashimoto, J., Araki, K., Sano, H., Chiken, S., Nambu, A., Ono, K., Ikenaka, K., et al. (2014). Disruption of actin-binding domain-containing Dystonin protein causes dystonia musculorum in mice. Eur. J. Neurosci. *40*, 3458–3471.
29. Borodulin, K., Vartiainen, E., Peltonen, M., Jousilahti, P., Juolevi, A., Laatikainen, T., Männistö, S., Salomaa, V., Sundvall, J., and Puska, P. (2015). Forty-year trends in cardiovascular risk factors in Finland. Eur. J. Public Health *25*, 539–546.
30. Duan, R., Shi, Y., Yu, L., Zhang, G., Li, J., Lin, Y., Guo, J., Wang, J., Shen, L., Jiang, H., et al. (2016). UBA5 mutations cause a new form of autosomal recessive cerebellar ataxia. PLoS ONE *11*, e0149039.
31. Watson, C.M., Crinnion, L.A., Gleghorn, L., Newman, W.G., Ramesar, R., Beighton, P., and Wallis, G.A. (2015). Identification of a mutation in the ubiquitin-fold modifier 1-specific peptidase 2 gene, UFSP2, in an extended South African family with Beukes hip dysplasia. S. Afr. Med. J. *105*, 558–563.

The American Journal of Human Genetics, Volume *99*

Supplemental Data

Biallelic Variants in *UBA5* Link Dysfunctional UFM1 Ubiquitin-like Modifier Pathway to Severe Infantile-Onset Encephalopathy

Mikko Muona, Ryosuke Ishimura, Anni Laari, Yoshinobu Ichimura, Tarja Linnankivi, Riikka Keski-Filppula, Riitta Herva, Heikki Rantala, Anders Paetau, Minna Pöyhönen, Miki Obata, Takefumi Uemura, Thomas Karhu, Norihisa Bizen, Hirohide Takebayashi, Shane McKee, Michael J. Parker, Nadia Akawi, Jeremy McRae, Matthew E. Hurles, the DDD Study, Outi Kuismin, Mitja I. Kurki, Anna-Kaisa Anttonen, Keiji Tanaka, Aarno Palotie, Satoshi Waguri, Anna-Elina Lehesjoki, and Masaaki Komatsu

Acknowledgements

This study was supported by the Folkhälsan Research Foundation (A-E.L.), Wellcome Trust grants [089062] and [098051] (A.P.), the European Commission FP7 project no. 242167 SynSys (A.P.), Health-2010 -project no. 261433 BioSHare (A.P.), the Academy of Finland grants no. 251704 and 263401 (A.P), Emil Aaltonen Foundation (M.M. and A.-K.A.), Epilepsy Research Foundation (M.M. and A.L.), University of Helsinki Funds (M.M.), Doctoral Programme in Biomedicine (M.M.), Arvo and Lea Ylppö Foundation (M.M. and A.-K.A.), Finnish Brain Foundation (M.M.), Paulo Foundation (M.M.), Biomedicum Helsinki Foundation (M.M.), Helsinki University Central Hospital Research Fund (A.-K.A.), a Grant-in-Aid for Scientific Research on Innovative Areas (25111006; M.K.), a Grant-in-Aid for Scientific Research (26253019; M.K.), a Grant-in-Aid for Exploratory Research (26650012; M.K.), a Human Frontier Scientific Program (RGP0055/2014; M.K.) and the Takeda Science Foundation (M.K.).

The authors thank the families for participation in the study, S. Lindh (Folkhälsan Institute of Genetics) for sample management, C. Scott and J. Durham (Wellcome Trust Sanger Institute) and Institute for Molecular Medicine Finland sequencing laboratory for exome sequence processing, CSC–IT Center for Science, Ltd. for the allocation of computational resources, A. Kyttälä (National Institute for Health and Welfare, Finland) for advice in cellular experiments, P. Häppölä (Institute for Molecular Medicine Finland) for checking the presence of p.Ala371Thr in Finnish population cohorts, V. Salomaa (National Institute for Health and Welfare, Finland) for sharing data from the FINRISK cohort, I. Helbig, S. Weckhuysen and other members of the EuroEPINOMICS RES consortium for data sharing, T. Kouno (Niigata University Graduate School of Medical and Dental Sciences) for assistant in mouse genetic study and K. Kanno (Fukushima Medical University School of Medicine) for assistance in histological study.

The DDD study presents independent research commissioned by the Health Innovation Challenge Fund (grant number HICF-1009-003), a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute (grant number WT098051). The views expressed in this publication are those of the author(s) and not necessarily those of the Wellcome Trust or the Department of Health. The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network.

M.M. is currently a part-time employee of Blueprint Genetics. Other authors do not report any conflicts of interest.

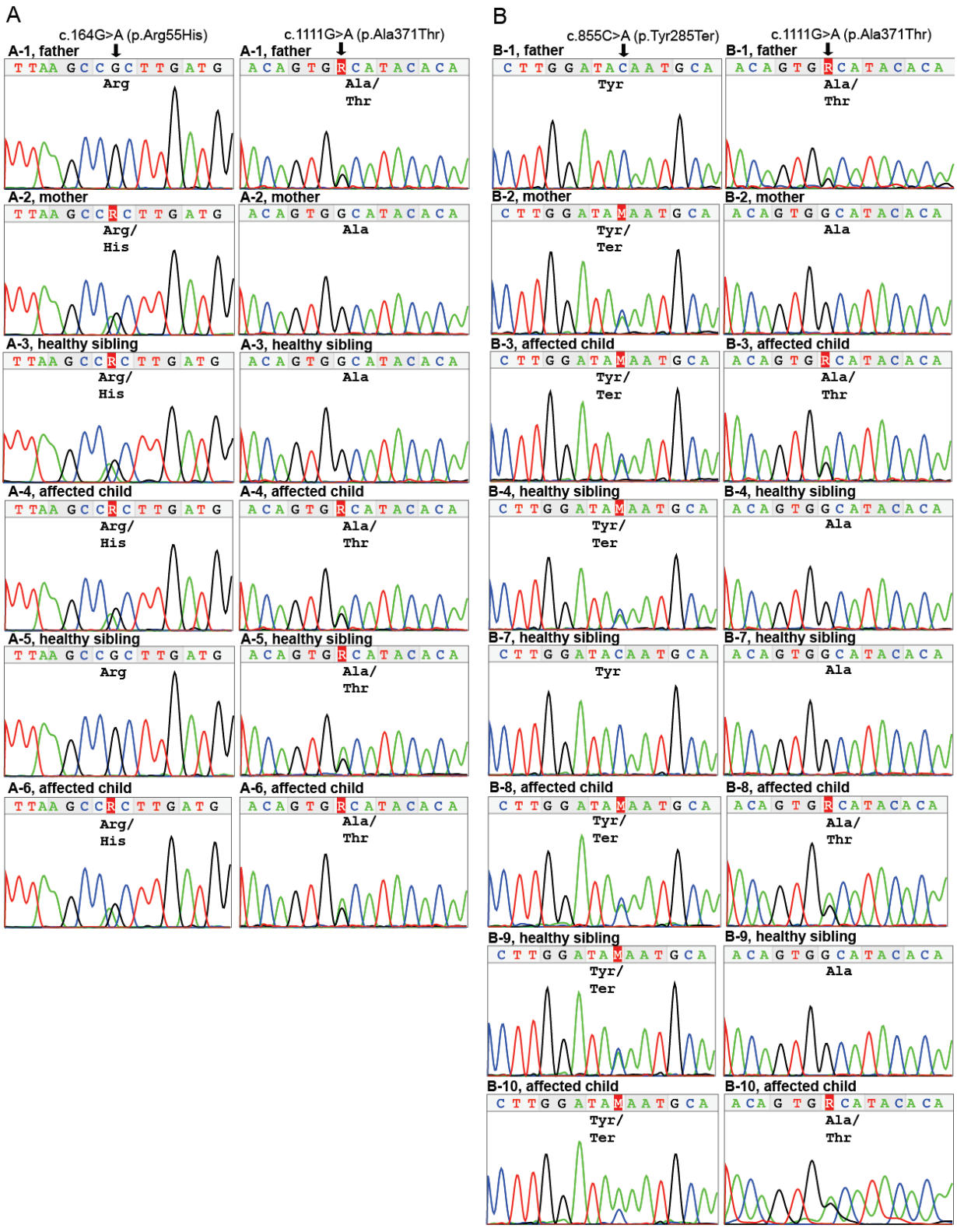

Figure S1. Validation of compound heterozygous *UBA5* variants. Capillary sequencing chromatograms of *UBA5* variants in families A (A) and B (B). Chromatograms were visualized by 4Peaks software (Nucleobytes B.V., Aalsmeer, The Netherlands).

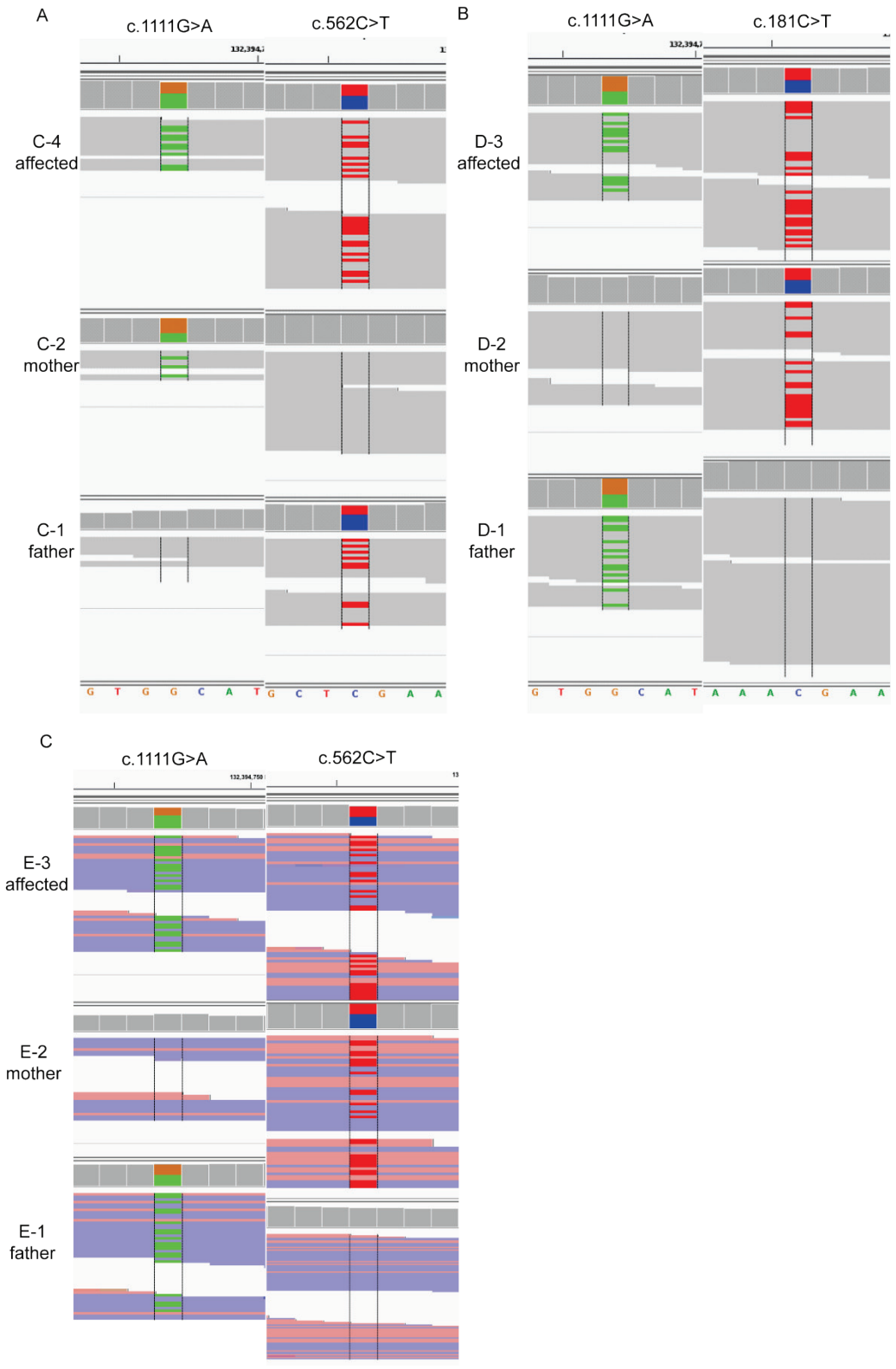

Figure S2. Visualization of *UBA5* variants in families C, D and E. Sequence reads at *UBA5* variant positions in family trios C (A), D (B) and E (C) were visualized using Integrative Genomics Viewer.[1]

Figure S3. Dysmorphic features of affected individuals C-4 and A-4. (A) C-4 at 10 years and 11 months of age. The individual has generally myopathic and expressionless face, prominent supra-orbital ridge, strabismus, small and open mouth with tented upper lip, long and deep philtrum as well as windswept, puffy hands with tapering fingers. (B) Facial features of individual A-4 at 7 months. Note the uplifted ear lobes, small chin as well as long and deep philtrum. A-4 had also puffy hands and feet at 11 months (no figures available). In addition, at 2 years, tapered fingers, open mouth, high palate, and hypermobile small joints of hands were seen (no figures available).

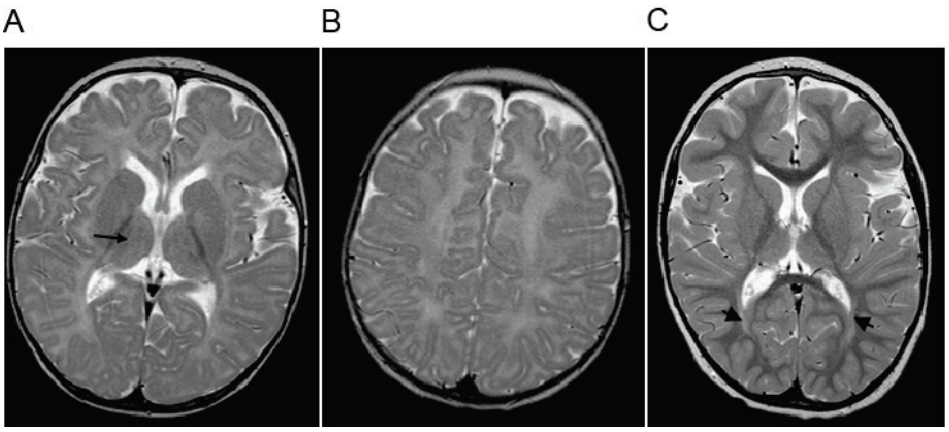

Figure S4. Brain MRI findings of two affected individuals with *UBA5* variants. (A and B) T2-weighted MR images of affected individual A-6 at the age of 6 months. Myelination is delayed and there is slight signal increase in thalami (arrow). (C) T2-weighted MR image of affected individual A-4 at the age of 5 years. There is signal increase periventricularly (arrows) and thalami are mildly atrophic. Myelination is appropriate for age.

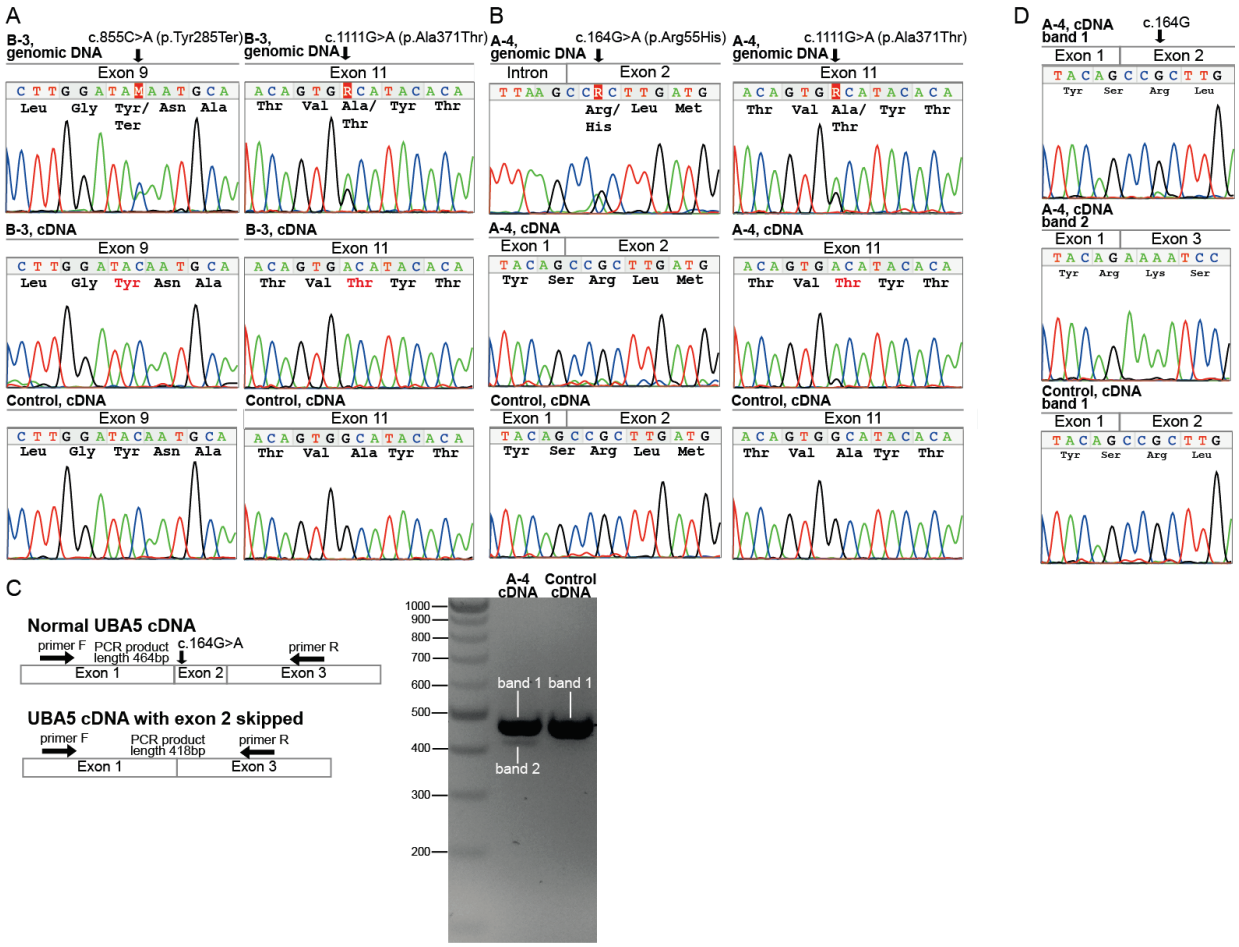

Figure S5. RNA level consequences of *UBA5* variants in the fibroblasts of affected individuals. (A and B**)** Capillary sequencing chromatograms of indicated *UBA5* variants in genomic DNA (top) or cDNA (middle) of affected individual B-3 (A) or A-4 **(**B**)** primary skin fibroblasts and in cDNA of control individual C2 primary skin fibroblasts (bottom). Total RNA was extracted from primary skin fibroblasts of cases A-4 and B-3 and an unrelated control individual C2 using RNeasy Plus Mini Kit (Qiagen, Hilden, Germany). Synthesis of cDNA from 400 ng of RNA was done using iScript cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA). *UBA5* (GenBank: NM_024818.3) variants (A-4: c.164G>A; A-4 and B-3: c.1111G>A; B-3: c.855C>A) were genotyped on the cDNA level by bidirectional Sanger sequencing. The primers used in PCR and sequencing reaction are available from authors upon request. As comparison variants were genotyped also in fibroblast genomic DNA, which was extracted using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). PCR products were subjected to bidirectional capillary sequencing. (C) Analysis of the effect of c.164G>A on *UBA5* exon 2 (exon numbering based on Ensembl: ENST00000356232, GenBank: NM_024818) splicing in primary skin fibroblasts of affected individual A-4. The left panel indicates the positions of primers (sequences available from authors) used in RT-PCR of *UBA5* cDNA and the expected size of the PCR products in the presence and absence of exon 2. The right panel shows an agarose gel (2%) electrophoresis of RT-PCR products from cDNA of either affected individual A-4 or control individual C2 using primers indicated in the left panel. (D) Capillary sequencing chromatograms of DNA extracted from the indicated bands in the gel shown in (C). DNA bands were extracted from the gel using NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany). Extracted DNA was subjected to bidirectional Sanger sequencing using the same primers as in the RT-PCR.

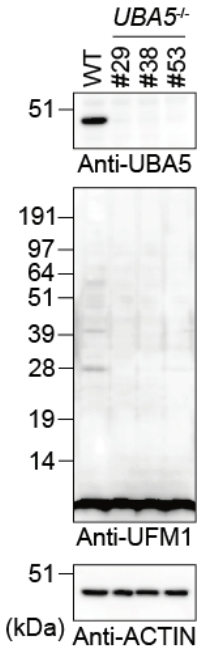

Figure S6. Immunoblot analysis of *UBA5* knockout HEK293T cells. For generation of *UBA5*-knockout cells, UBA5 guide RNA designed by CRISPR Design tool (http://crispr.mit.edu/) was subcloned into pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene #42230), a human codon-optimized SpCas9 and chimeric guide RNA expression plasmid. HEK293T cells were transfected with the vector together with pEGFP-C1 vector (#6084-1, Clontech Laboratories, Inc., Mountain View, CA), and cultured for 2 days. Thereafter, the GFP-positive cells were sorted and expanded. Loss of *UBA5* was confirmed by heteroduplex mobility assay followed by immunoblot analysis with anti-UBA5 antibody. Cell lysates of *UBA5*-deficient HEK293T cells (*UBA5*KO) and of intact HEK293T (WT) cells were subjected to NuPAGE, followed by immunoblot analysis with indicated antibodies. Data shown are representative of three separate experiments.

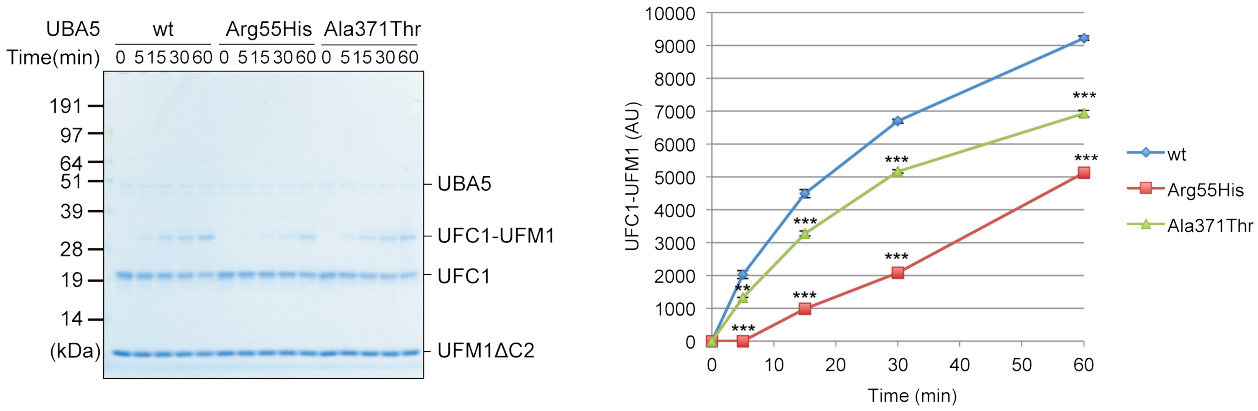

Figure S7. *In vitro* thioester formation assay of UFM1 by UFC1. Recombinant GST-UFM1ΔC2, GST-UBA5, GST-UBA5^{Arg55His}, GST-UBA5^{Ala371Thr} and GST-UFC1 were produced in *E. coli* and recombinant proteins were purified by chromatography on Glutathione sepharose 4B (GE Healthcare UK Ltd). After digestion of GST by PreScission Protease (GE Healthcare UK Ltd), the recombinant proteins were dialyzed against 50 mM BisTris (pH 6.5), 100 mM NaCl, 10 mM $MgCl_2$, and 0.1 mM DTT (reaction buffer). Thioester formation reactions contained reaction buffer with 5 mM ATP, 0.2 µg UFM1ΔC2, 0.5 µg UFC1 and either 0.05 µg UBA5 wildtype or 0.05 µg of one the UBA5 mutants. Reactions were incubated for indicated time at 25°C and stopped by the addition of NuPAGE-loading buffer lacking reducing agent, followed by 10 min incubation at 37°C, NuPAGE (12% Bis-Tris gel) and Coomassie Brilliant Blue staining. Data shown are representative of three separate experiments. Graph indicates the quantitative densitometric analyses of UFC1-UFM1 intermediates using NIH Image software. Statistical analysis was performed using the unpaired *t*-test (Welch test). The data represents the means ± SE of three separate experiments. **, $P<0.01$, and ***, $P<0.001$.

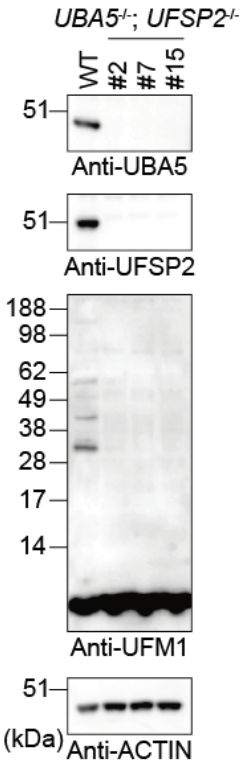

Figure S8. Immunoblot analysis of *UBA5 UFSP2* double knockout HEK293T cells. For generation of *UBA5*- and *UFSP2*-double knockout cells, each UBA5 and UFSP2 guide RNA designed by CRISPR Design tool (http://crispr.mit.edu/) was subcloned into pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene #42230), a human codon-optimized SpCas9 and chimeric guide RNA expression plasmid. HEK293T cells were transfected with both vectors together with pEGFP-C1 vector (#6084-1, Clontech Laboratories), and cultured for 2 days. Thereafter, the GFP-positive cells were sorted and expanded. Loss of *UBA5* and *UFSP2* was confirmed by heteroduplex mobility assay followed by immunoblot analysis with anti-UBA5 and UFSP2 antibodies. Cell lysates of *UBA5*- and *UFSP2*-double deficient HEK293T cells (*UBA5*$^{-/-}$; *UFSP2*$^{-/-}$) and that of intact HEK293T (WT) cells were subjected to NuPAGE, followed by immunoblot analysis with indicated antibodies. Rabbit monoclonal anti-UFSP2 antibody (ab185965; Abcam, Inc., Cambridge, MA) was used for immunodetection. Data shown are representative of three separate experiments.

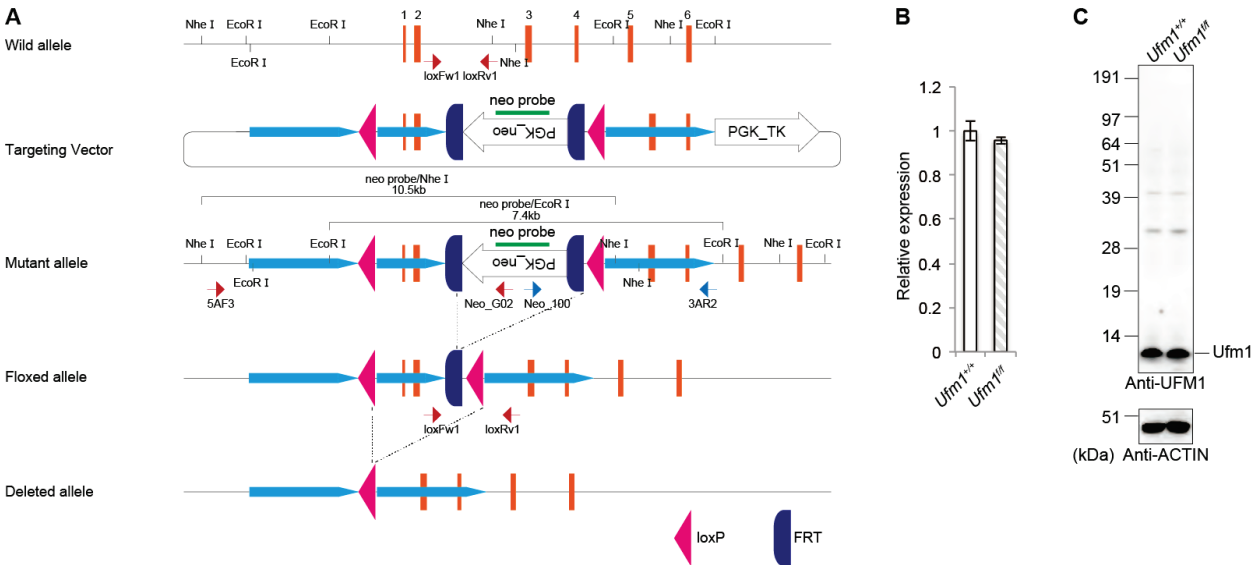

Figure S9. Generation of *Ufm1* conditional knockout mice. (A) Schematic representation of the targeting vector and the targeted allele of the *Ufm1* gene. The coding exons numbered in accordance with the initiation site as exon 1 are depicted by orange boxes. *LoxP* and *FRT* sequences are indicated. Arrows indicate the positions of PCR primers. EcoRI, *Eco*RI sites; NheI, *Nhe*I sites; neo, neomycin resistance gene cassette; TK, thymidine kinase. The targeting vector was electroporated into mouse RENKA ES cells, selected with G418 (250 μg/ml; Life Technologies, Carlsbad, CA), and then screened for homologous recombinants by PCR and Southern blot analyses. PCR primers were as follows: 5AF3 5'-AAGGCGCGCCATTAAATGAGGCATGACTGC-3', and Neo_G02 5'-ATCAGGACATAGCGTTGGCTAC-3' for 5' terminal side, and 3AR2 5'-AAGTCGACAATCCAGAATACACTAGGGG-3', and Neo_100 5'-AGGTGAGATGACAGGAGATC-3' for 3' terminal side. Southern blot analysis was performed by digestion of genomic DNA with *Nhe*I or *Eco*RI and hybridization to detect 10.5-kb and 7.4-kb bands that correspond to the wild-type and mutant allele. A probe for Southern blot analysis is shown as green bar. To delete the neomycin resistance gene, the mice with mutant allele were bred with *PGK-FLPo* transgenic mice[2] (MGI ID: 4415609). Resulting progeny containing the $Ufm1^{flox}$ allele were bred with *nestin*-Cre transgenic mice[3] (MGI ID: 2176173). Genotyping of the floxed mice by PCR was performed using the following two primers: loxFw1 5'-AGAGGGGGACAGGAGGAAGGGTGCGGCATT-3', and loxRv1 5'-ACTGAAGTTAAGAGTGTTCGGGTCGGACCT-3'. Mice were housed in specific-pathogen free facilities and the experimental protocol was approved by the Ethics Review Committee for Animal Experimentation of Niigata University School of Medicine. (B) Expression of *Ufm1* transcript. Total RNA was extracted from each genotyped liver. The *Ufm1* transcript level was detected by real time-PCR as shown in Figure 2A. Signals were assessed relative to that of glucuronidase (Gus). Values were normalized to the amount of mRNA in the wild-type mouse liver. The experiments were performed three times. The sequences of the primers are shown in Table S6. Data represented are means ± SE. (C) Immunoblot analysis of UFM1. The livers of indicated genotype mice were homogenized as shown in Figure 4A, and then homogenates were subjected to NuPAGE, followed by immunoblot analysis with indicated antibodies.

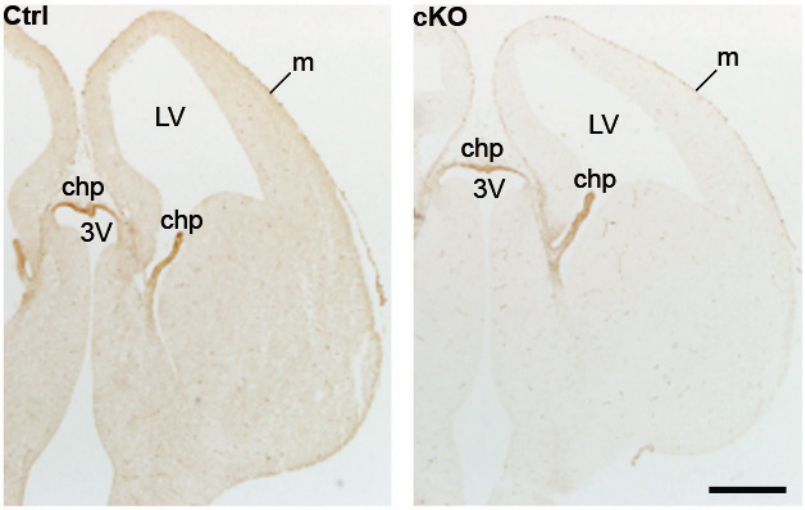

Figure S10. Loss of UFM1 immunoreactivity in the brain parenchyma of *Ufm1* conditional knockout mice. Cryosections of E14.5 brains were immunostained using anti-UFM1 antibody. Immunohistochemistry was performed as described in Figure 4D, using rabbit monoclonal anti-UFM1 antibody (ab109305; Abcam, 1:1000). Images were captured with BX53 microscopes (Olympus, Tokyo, Japan). Left panel control (*Ufm1*$^{f/+}$;nestin-*Cre*) brain and right panel knockout (*Ufm1*$^{f/f}$; nestin-*Cre*) brain. Meninges (m) and choroid plexus (chp) in the lateral ventricle (LV) and third ventricle (3V) are positive for UFM1 immunoreactivity in the knockout brain. It has previously been reported that Cre recombination does not affect meninges and choroid plexus in nestin-*Cre* transgenic mice.[4] Bar: 500 μm.

Table S1. Exome sequencing metrics.

	A-1 father	A-2 mother	A-4 affected	B-1 father	B-2 mother	B-3 affected	B-4 un-affected sibling	B-8 affected
Sequence (Gb)	2.7	2.6	4.7	1.7	1.7	1.8	1.9	1.5
Mean coverage	53.4	50.1	90.5	35.5	37.0	38.9	40.9	31.9
Percentage of reads at least 5×	93.3	93	97	89.8	90.5	90.6	91.6	90.6
Percentage of reads at least 10×	87.8	87.2	93.7	79.7	81.1	81.4	82.9	79.8

Exome sequence metrics within the exome bait regions of Agilent SureSelect Human All Exon 50 Mb V3 (family A) or Nimblegen SeqCap EZ Human Exome Library v2.0 (family B). Values were obtained after removal of duplicate reads.

Table S2. Variant filtering process under autosomal recessive inheritance model in families A and B.

	Variant counts		
Exome variant filtering step	**A-4**	**B-3**	**B-8**
All exome bait region single nucleotide variants and indels	39367	35524	33669
Missense, stop gained, stop lost, acceptor or donor splice site variants, frameshift indels, inframe indels in CCDS genes	8879	7687	7683
Rare (<1%) and not homozygous in variant databases (ExAC, Exome variant server, 1000 genomes)	169	203	197
Homozygous or compound heterozygous	2 (*UBA5*)	2 (*UBA5*)	2 (*UBA5*)
Shared with exome- or capillary-sequenced affected siblings and not present in unaffected siblings	2 (*UBA5*)	2 (*UBA5*)	2 (*UBA5*)

The number of variants remaining after each filtering step is presented for exome sequenced affected individuals A-4, B-3 and B-8 (see Figure 1B). Variant counts in the two last steps have been counted after removing low quality variants based on visualization of sequence reads using Integrative Genomics Viewer (https://www.broadinstitute.org/igv/). Different exome capture, sequencing and variant calling methods were used for families A and B explaining differences in variant counts (see Main text).

Table S3. *UBA5* variants in families A-E.

Family ID	Gene	Position (GRCh37)	dbSNP ID (build 138)	Ref/alt allele	ExAC all frequency	ExAC homozygotes	ExAC Europeans frequency	ExAC Finns frequency	Predicted consequence	Coding DNA change	Predicted protein change	CADD, SIFT, PolyPhen, Mutation Taster[a]
A	*UBA5*	3:132384669	-	G/A	8.29E-06	0	1.505E-05	0	missense[b]	c.164G>A	p.Arg55His[b]	D,D,D,D
D	*UBA5*	3:132384686	-	C/T	0	0	0	0	stop gained	c.181C>T	p.Arg61Ter	D,-,-,D
C,E	*UBA5*	3:132389876	rs374052333	C/T	0	0	0	0	stop gained	c.562C>T	p.Arg188Ter	D,-,-,D
B	*UBA5*	3:132394134	-	C/A	0	0	0	0	stop gained	c.855C>A	p.Tyr285Ter	D,-,-,D
A,B,C,D,E	*UBA5*	3:132394747	rs114925667	G/A	0.0019	0	0.002791	0.0046	missense	c.1111G>A	p.Ala371Thr	D,D,PD,D

Ref, reference; Alt, alternate; ExAC, Exome Aggregation Consortium; -, not available.
[a]In the CADD (v. 1.3) annotation, variants with a phred-scaled score of >15 were annotated as deleterious. PD, possibly deleterious (applies to PolyPhen only); D, deleterious; -, not available.
[b]The c.164G>A variant affects also splicing (see Main text).

Mutation nomenclature is based on the following Ensembl/GenBank transcripts: ENST00000356232/NM_024818.3

Table S4. Utilization of recessiveStats tool to determine probability to observe *n* number of affected individuals with rare, biallelic variants in *UBA5* in the study populations

	Ancestries of unrelated affected individuals in the study populations[1]					
	FIN	**NFE**	**AFR**	**EAS**	**SAS**	**Total**
Number of individuals	490	2,858	100	15	278	3,741
Cumulative allele frequencies of rare variants (<1%) in ExAC v. 0.3						
Functional[2] variants	0.00625	0.00630	0.00533	0.00522	0.00291	NA
LoF[3] variants	0.00015	0.00016	0.00010	0.00012	0.00006	NA
Expected number of individuals with biallelic *UBA5* variants in the study population[4]						
Biallelic functional/functional	0.03107	0.18376	0.00492	0.00072	0.00552	0.22670
Biallelic functional/LoF	0.00123	0.00774	0.00015	2.52E-05	0.00017	0.00933
Biallelic LoF/LoF	0.00030	0.00190	4.03E-05	7.03E-06	6.73E-05	0.00232
Observed number of individuals with biallelic *UBA5* variants						
Biallelic functional/functional[5]	1	2	0	0	0	3
Biallelic functional/LoF[6]	2	2	0	0	0	4
Biallelic LoF/LoF	0	0	0	0	0	0
recessiveStats *P* value for genotype enrichment[7]						
Functional/functional	NA	NA	NA	NA	NA	0.0018
Functional/LoF	NA	NA	NA	NA	NA	3.30×10^{-10}
LoF/LoF	NA	NA	NA	NA	NA	1

FIN, Finnish; NFE, non-Finnish Europeans; AFR, Africans; EAS, East Asians; SAS, South Asians; ExAC, Exome Aggregation Consortium; NA, not applicable; LoF, loss-of-function

[1] Study populations consist of 35 in-house exomes from Finnish individuals with severe epilepsy syndromes, one in-house exome of Iranian individual with severe epilepsy syndrome, 455 exomes from the Northern Finnish Intellectual disability cohort, 3,072 exomes of unsolved individuals from the Deciphering Developmental Disorders (DDD) study and 178 exomes of individuals with rare epilepsy syndromes in the EuroEPINOMICS study.

[2] Variants annotated as 'functional' have one of the following consequences: stop_lost, initiator_codon_variant, transcript_amplification, inframe_insertion, inframe_deletion, missense_variant, coding_sequence_variant.

[3] Variants annotated as 'LoF' have one of the following consequences: transcript_ablation, splice_donor_variant, splice_acceptor_variant, stop_gained, frameshift_variant. For AFR subpopulation, there are not any individuals with LoF variants in ExAC. The presented LoF allele frequency was determined assuming that the next individual recruited will have such a variant.

[4]The expected number of individuals with biallelic LoF variants was determined as follows:

$((rate_{lof}\text{^}2) * (1 - autozygosity\_rate) + rate_{lof} * autozygosity\_rate) * n$, where rate is the cumulative allele frequency of LoF variants in the population in the ExAC database and *n* the total number of individuals in the study populations.

Expected number of indidividuals with biallelic LoF/functional variants (also biallelic LoF variants were included in this category):

$((rate_{lof}\text{^}2) * (1 - autozygosity\_rate) + rate_{lof} * autozygosity\_rate + 2 * rate_{lof} * (1 - rate_{lof}) * rate_{functional}) * n$, where rate is the cumulative allele frequency of variants in the population in the ExAC database and *n* the total number of individuals in the study populations.

The expected number of individuals with biallelic functional variants:

$((rate_{functional}\text{^}2) * (1 - autozygosity\_rate) + rate_{functional} * autozygosity\_rate) * n$, where rate is the cumulative allele frequency of functional variants in the population in the ExAC database and *n* the total number of individuals in the study population.

Autozygosity rate of 0.003931 was used to account for increased rate of homozygous genotypes due to autozygosity in *UBA5* in the study populations. Autozygosity rate was determined using samples in the Finnish and DDD study populations using 'bcftools roh' tool.[5]

[5] The one Finnish individual with biallelic functional/functional genotype is individual A-4 who is heterozygous for p.Ala371Thr and p.Arg55His. Notably, the latter variant, 164G>A (p.Arg55His), is in effect a loss-of-function variant given that it affects splicing and the Arg55His UBA5 mutant has severely reduced enzymatic activity. The two individuals of European ancestry are from the DDD study. Their biallelic variants are not likely to be pathogenic as 1) the variant combination is not present in a similarly affected sibling or 2) the homozygous variant (p.Ala371Thr) in the affected individual is present as homozygous in an unaffected individual in his fifties in a Finnish population based cohort (see main text).

[6] These four individuals with biallelic functional/LoF variants represent families B-E in our study.

[7] Threshold for statistical significance after conservative multiple testing correction for three tests, one for each biallelic genotype class, and for the total number annotatable protein coding genes in GENCODE release 19: $0.05/3/17{,}370 = 9.60\times10^{-7}$.

Table S5. Clinical features of affected individuals with *UBA5* mutations.

Family and individual ID	A-4	A-6	B-3	B-5	B-8	B-10	C-4	D-3	E-3
Gender	Male	Female	Female	Male	Male	Male	Male	Female	Male
Ancestry	Finnish	Finnish	Finnish	Finnish	Finnish	Finnish	White British	Father Northern Irish, mother Romanian	Finnish
Year of birth	2008	2013	1974	1977	1982	1988	2000	2010	2006
Pregnancy and delivery	Normal	Normal	Normal	Normal	Caesarean section gw 37, Apgar 1/7	Normal	Normal	Caesarean section gw33 for pre-eclampsia, Apgar 8/9	Caesarean section gw 39 for breech presentation, Apgar 9/10/10
Birth weight/height/ OFC	3,335 g/48 cm/ 34 cm	3,285 g/47.5 cm/NA	3,720 g/49.5 cm/ 34.5 cm	2,780 g/46.5 cm/ 33.5 cm	3,450 g/49 cm/ 34.5 cm	3,700 g/48 cm/ 35 cm	3,490 g/NA/NA	1,580 g /43cm /29.5cm	3,030 g/47 cm/34 cm
Age at presentation	2 mo	1 wk	2-3 mo	1 wk	At birth	1 wk	2 wks	3wks	3-4 mo
Presenting symptoms	Irritability, back arching, poor eye contact	Irritability, back arching, poor eye contact	Jitteriness, back arching	Jitteriness, back arching	Jitteriness, bouts of arm and leg stiffening	Irritability, jitteriness, back arching	Excessive crying, irritability and jitteriness from birth	Irritability, GER, back arching, writhing movements, poor eye contact	Poor eye contact
Truncal hypotonia	Yes	Yes	Yes	Yes	Yes	Yes	Possible	Yes	Yes
Dystonic/athetoid movements	Yes	Yes	Yes	Yes	Yes	Yes	No record (little spontaneous movement)	Yes	Yes
Spasticity (age)	Yes (5 y)	No	Yes (NA)	Yes (NA)	Yes (NA)	Yes (NA)	Yes	Yes	Yes (5 mo)
Seizures (age at onset)	Infantile spasms (3 mo), myoclonus	Infantile spasms (4 mo), myoclonus	Myoclonus	Motor (9 mo), myoclonus	Infantile spasms (8-12 mo)	Motor, infantile spasms (4 mo)	Infantile spasms (6 mo) associated with hypsarrhythmia	No	Hypomotor seizures (3-5 mo)
Vision	Impaired: no fixation, dyskinetic squint	Impaired: fixation, not following	Follows objects, squint	Follows objects, squint	Follows objects, squint	Follows objects, squint	No record (subjectively limited eye movements?)	Intermittent squint; otherwise normal	Intermittent outward squint, partially follows objects
Optic fundi	Normal	NA	Normal	NA	Normal	Normal	No record	Normal	NA
Psychomotor development	Arrested. Vocalizing, no speech. Absent motor skills. Severe intellectual deficit.	Arrested. Vocalizing, no speech. Absent motor skills. Severe intellectual deficit.	Arrested. Vocalizing, no speech. Absent motor skills. Severe intellectual deficit.	Arrested. No speech. Absent motor skills. Severe intellectual deficit.	Arrested. No speech. Absent motor skills. Severe intellectual deficit.	Arrested. No speech. Absent motor skills. Severe intellectual deficit.	Arrested. Four-limb spasticity with associated scoliosis and contractures & excessive salivation (drooling); never spoke or sat independently.	Arrested. Vocalizing, no speech. Smiles and laughs appropriately. Good comprehension. Some ability to feed self. Motor milestones globally delayed. Moderate intellectual deficit.	Arrested. No speech. Absent motor skills. Severe intellectual deficit.
Acquired microcephaly	Yes	Yes	Yes	Yes	NA	Yes	Yes (OFC around 25th centile as baby, but approximately 50cm at 10y)	Yes	No (-1 SD at birth, -2 SD 9 yr)
Stature (age)	-0,1 SD (3 y)	1 SD (1.7 y)	-2.6 SD (1 y) -3.6 SD (3 y) -7 SD (15 y)	-4.8 SD (1.2 y) <-7 SD (16 y)	-2.7 SD (1 y) - 6 SD (21 y)	-2.7 SD (1.1y) -4.3 SD (12 y)	Weight at 15 years approx. 42kg, but difficult to measure length due to spasticity & scoliosis	Height 110cm (25th percentile) Weight 15.8kg (2nd percentile) (5y 7mo)	Height 111 cm (-5 SD) Weight 16.4 kg (5th percentile) (9 yr)
Brain MRI (age)	Delayed myelination (7 mo); Mild thalamic and cerebellar atrophy, SI periventricularly (5 y)	Normal (3 mo); Delayed myelination, mild WM volume loss, mild T2 SI in thalami (6 mo)	Mild cortical and central atrophy (19 y)	NA	NA	Mild cortical and central atrophy, SI periventricularly (4 y)	Non-specific small brain (8 mo)	Normal apart from slightly small cerebellum (3y 4mo)	Delayed myelination, frontal subcortical SI (left),
Other	Dysmorphic features (Figure S3), GER	GER					Dysmorphic (Figure S3, expressionless/myopathic facial features, GER		Acetabular dysplasia (right), GER, abnormal visual evoked potential (VEP) test
Outcome	Death at 5 y	Alive	Alive	Death at 16 y	Death at 21 y	Death at 12 y	Alive	Alive	Alive

Gw, gestational weeks; OFC, occipitofrontal circumference; GER, gastroesophageal reflux; WM, cerebral white matter; SI, signal increase; NA, data not available

Table S6. Genotype analysis of progeny intercross between *Ufm1*$^{f/+}$;nestin-Cre and *Ufm1*$^{f/f}$ mice.

	Ufm1$^{f/+}$	*Ufm1*$^{f/f}$	*Ufm1*$^{f/+}$; nestin-*Cre*	*Ufm1*$^{f/f}$; nestin-*Cre*	
				Alive	**Dead**
E14.5	2	3	5	2	0
E18.5	10	10	14	9	0
P1	11	8	11	0	9

Table S7. The sequences of the primers used in quantitative real-time PCR.

	Left	Right
UBA5 (OMIM: 610552)	CAGCACTGCCTAAACAAGAGG	TGAAACCTCAGATACCAGCTCA
UFC1 (OMIM: 610554)	TTTGGACTAGCTCATCTCATGG	GAATCAGATCAGGGATTTCCAC
UFM1 (OMIM: 610553)	GGCTGCCGTACAAAGTACTCA	TTCCATCATTGGTAATAATTGCAC
GAPDH (OMIM: 138400)	AGCCACATCGCTCAGACAC	GCCCAATACGACCAATCC
Ufm1	AGTACGCCGTTCACAGCAG	CATTAGTAATAATCGCACTTGTAGCTG
Gus	GATGTGGTCTGTGGCCAAT	TGTGGGTGATCAGCGTCTT

Supplemental References

1. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24-26.

2. Wu, Y., Wang, C., Sun, H., LeRoith, D., and Yakar, S. (2009). High-efficient FLPo deleter mice in C57BL/6J background. PLoS ONE *4*, e8054.

3. Tronche, F., Kellendonk, C., Kretz, O., Gass, P., Anlag, K., Orban, P.C., Bock, R., Klein, R., and Schutz, G. (1999). Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. Nat. Genet. *23*, 99-103.

4. Blaess, S., Graus-Porta, D., Belvindrah, R., Radakovits, R., Pons, S., Littlewood-Evans, A., Senften, M., Guo, H., Li, Y., Miner, J.H., et al. (2004). β1-Integrins are critical for cerebellar granule cell precursor proliferation. J. Neurosci. *24*, 3402-3412.

5. Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics.