Exome sequencing and network analysis identifies shared mechanisms underlying spinocerebellar ataxia

Nibbeling, Esther A R; Duarri, Anna; Verschuuren-Bemelmans, Corien C; Fokkens, Michiel R; Karjalainen, Juha M; Smeets, Cleo J L M; de Boer-Bergsma, Jelkje J; van der Vries, Gerben; Dooijes, Dennis; Bampi, Giovana B

Published in:
Brain

DOI:
10.1093/brain/awx251

Publication date:
2017

doi:10.1093/brain/awx251 BRAIN 2017: 140; 2860–2878 | 2860

BRAIN
A JOURNAL OF NEUROLOGY

Exome sequencing and network analysis identifies shared mechanisms underlying spinocerebellar ataxia

Esther A. R. Nibbeling,[1] Anna Duarri,[1] Corien C. Verschuuren-Bemelmans,[1] Michiel R. Fokkens,[1] Juha M. Karjalainen,[1] Cleo J. L. M. Smeets,[1] Jelkje J. de Boer-Bergsma,[1] Gerben van der Vries,[1] Dennis Dooijes,[2] Giovana B. Bampi,[1] Cleo van Diemen,[1] Ewout Brunt,[3] Elly Ippel,[2] Berry Kremer,[3] Monique Vlak,[4] Noam Adir,[5] Cisca Wijmenga,[1] Bart P. C. van de Warrenburg,[6] Lude Franke,[1] Richard J. Sinke[1] and Dineke S. Verbeek[1]

The autosomal dominant cerebellar ataxias, referred to as spinocerebellar ataxias in genetic nomenclature, are a rare group of progressive neurodegenerative disorders characterized by loss of balance and coordination. Despite the identification of numerous disease genes, a substantial number of cases still remain without a genetic diagnosis. Here, we report five novel spinocerebellar ataxia genes, *FAT2*, *PLD3*, *KIF26B*, *EP300*, and *FAT1*, identified through a combination of exome sequencing in genetically undiagnosed families and targeted resequencing of exome candidates in a cohort of singletons. We validated almost all genes genetically, assessed damaging effects of the gene variants in cell models and further consolidated a role for several of these genes in the aetiology of spinocerebellar ataxia through network analysis. Our work links spinocerebellar ataxia to alterations in synaptic transmission and transcription regulation, and identifies these as the main shared mechanisms underlying the genetically diverse spinocerebellar ataxia types.

1 Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
2 Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands
3 Department of Neurology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
4 Department of Neurology, Medical Center Haaglanden and Bronovo-Nebo, Den Hague, The Netherlands
5 Schulich Faculty of Chemistry, Technion–Israel Institute of Technology, Technion City, Israel
6 Department of Neurology, University Medical Center Nijmegen, Nijmegen, The Netherlands

Correspondence to: Dineke Verbeek,
University Medical Center Groningen, Antonius Deusinglaan 1, 9700RB, Groningen, The Netherlands
E-mail: D.S.Verbeek@umcg.nl

Keywords: spinocerebellar ataxia; whole exome sequencing; synaptic transmission; neurodegeneration; genetic network

Abbreviations: SCA = spinocerebellar ataxia; WES = whole exome sequencing

Introduction

Autosomal dominant cerebellar ataxia (ADCA) patients suffer from coordination problems, loss of balance, gait abnormalities, and slurred speech. These symptoms are often linked to other neurological features, such as pyramidal or extrapyramidal signs, and sometimes to mild intellectual disability (Durr, 2010; Klockgether, 2011). The cerebellar symptoms are caused by cerebellar atrophy and loss of Purkinje cells in the cerebellum and subsequent

Received March 20, 2017. Revised August 1, 2017. Accepted August 5, 2017. Advance Access publication October 6, 2017

neuronal loss in spinocerebellar tracts (Shakkottai *et al.*, 2011). Disease onset is usually during midlife but can also occur in childhood or at older ages. Next to the clinical heterogeneity, the disorder is also genetically highly heterogeneous with over 40 spinocerebellar ataxia (SCA) types currently recognized, for which mutations have been identified in 32 genes (Bakalkin *et al.*, 2010; Durr, 2010; Wang *et al.*, 2010; Kobayashi *et al.*, 2011; Duarri *et al.*, 2012; Hekman *et al.*, 2012; Lee *et al.*, 2012; Cadieux-Dion *et al.*, 2014; Delplanque *et al.*, 2014; Di Gregorio *et al.*, 2014; Tsoi *et al.*, 2014; Coutelier *et al.*, 2015; Fogel *et al.*, 2015; Depondt *et al.*, 2016; Seixas *et al.*, 2017). The majority of patients (~65%) are characterized by coding and/or non-coding repeat expansions, with only a minority of SCAs caused by truncating, missense, and nonsense mutations (Durr, 2010). Despite the identification of a large number of SCA genes, routine diagnostics yield a genetic diagnosis in only 60–70% of clinically clear ADCA cases. Furthermore, knowledge on the expected shared pathophysiology underlying cerebellar neurodegeneration is very limited. We therefore aimed to identify the genetic cause of ADCA in 20 families who remained undiagnosed after regular DNA diagnostics. To achieve this, we use a combination of whole exome sequencing (WES), targeted resequencing, gene network analysis, and functional validation (Supplementary Fig. 1).

Materials and methods

Patients

Twenty families with autosomal dominant cerebellar ataxia were selected for WES analysis. In these families repeat expansion mutations causing SCA1, 2, 3, 6, 7, and 17 were excluded by routine diagnostics using direct amplification and detection of the corresponding repeat as WES is not very suitable to detect these. Additionally, we already excluded the presence of mutations in the genes causing SCA13, 19/22, and 23 by Sanger sequencing in previous research projects. From these 20 families, 40 individuals were analysed including two or three most-distantly-related family members of multiplex families ($n = 8$), parent-child trios ($n = 2$), affected parent-child pairs ($n = 5$) and single cases ($n = 5$) in simplex pedigrees (Supplementary Table 1). All individuals gave informed consent as approved by the Medical Ethical Committee of the University Medical Center Groningen. Sanger sequencing and co-segregation analysis were used to validate the WES findings in available affected family members. The 96 singletons (all independent referrals) selected for gene panel analysis were obtained from the genetic diagnostic center of Groningen upon exclusion of mutations in routinely tested SCA genes. The extended DNA analyses were performed in a diagnostic setting (accredited diagnostic DNA laboratory) as follow-up testing of patients in line with the original diagnostic request. Patients were excluded if they had indicated that they did not agree with the use of their DNA for future (anonymous) studies to help develop or improve diagnostic techniques on the original 'Request for DNA Test' form.

Whole exome sequencing

Exome capturing was performed using the SureSelect Human all Exon V4 kit (Agilent technologies) according to the manufacturer's protocol. One-hundred base pair paired-end reads were generated on a HiSeq2000 platform (Illumina Inc.). Sequences were aligned to hg19, and variants were identified through an in-house bioinformatics pipeline (https://github.com/molgenis/NGS_DNA). Variants were interpreted using Ingenuity Variant Analysis (Qiagen, Hilden, Germany) and NGS Bench Lab software Cartagenia (Agilent Technologies). Variants with a minor allele frequency >0.1% in genetic databases including the Exome Aggregation Consortium (ExAC) (assessed October 2015, processed variants recently checked); 1000 Genomes; dbSNP; Exome Variant Server; NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (October 2015 assessed, processed variants recently checked); and GoNL (1000 Genomes Project Consortium *et al.*, 2010; Tennessen *et al.*, 2012; Genome of the Netherlands Consortium, 2014) were excluded from further analysis. Variants shared by affected family members were assessed for their computationally predicted pathogenicity by SIFT, PolyPhen-2, and MutationTaster (Kumar *et al.*, 2009; Adzhubei *et al.*, 2010; Schwarz *et al.*, 2010). Variants were included in further analyses when at least two of the three programs predicted them to be damaging or disease-causing. For overall summary statistics of the exome sequencing data see Supplementary Table 2.

Targeted resequencing

The capturing kit, designed using Agilent SureDesign software containing (Agilent Technologies), contained 28 candidate genes selected upon whole exome analysis in 10 SCA-negative (no mutations in known SCA genes including SCA1–3, 6, 7, 17, and 23) families in which multiple candidate variants were identified by WES supplemented with 14 known SCA genes that were not routinely screened for by diagnostics (for complete gene list see Supplementary Tables 3 and 4). The resulting array was designed to screen the entire coding region of the corresponding genes including 25 base pairs flanking both sides of the exons to cover intron–exon boundaries. The density of probes was set to five probes per target nucleotide and boosting of probes high in GC content was set to maximal performance. Sequence reads 150 bp in length were generated on a MiSeq platform (Illumina Inc.). Variants were annotated and interpreted as described above. For summary statistics of the targeted sequencing data see Supplementary Table 2.

Plasmids

pUC119-SR-*Fat2* (rat) (gift from Prof. Nakayama, Kazusa DNA research institute, Chiba, Japan) was used as a backbone to generate the plasmids used in this work. In short, pUC119-SR-*Fat2* was cut with MluI and NotI and the cDNA of *Fat2* was subcloned into a MluI/NotI digested pEGFP-N1 plasmid in which the PstI site was replaced for a MluI site (for primers see Supplementary Table 5). After this subcloning, we generated a plasmid containing the full-length cDNA of *Fat2* lacking

the EGFP-called pCMV-Fat2. To introduce the mutations, a PCR fragment of *Fat2* containing an AvrII site on the N-terminus and a SalI site on the C-terminus was subcloned in pJET1.2/blunt followed by site-directed mutagenesis (for primers see Supplementary Table 5). This facilitated the subcloning of the mutated cDNA fragments back into the pCMV-Fat2 plasmid. To fuse the *Fat2* cDNA to an EGFP or haemagglutinin (HA) tag, the C-terminus of *Fat2* (cDNA fragment between SalI and NotI) was subcloned into pJET1.2/blunt. The HA tag was introduced by PCR, whereas the cDNA of EGFP (fragment SacII-EGFP-NotI from the EGFP-N1 plasmid) was subcloned. Both C-terminally tagged fragments were subcloned in the pCMV-Fat2 plasmid between the cutting sites SalI and Not1 generating pCMV-Fat2-EGFP and pCMV-Fat2-HA plasmids (for primers see Supplementary Table 5).

The cDNAs in the plasmids pcDNA3.1-V5-HisB-FAT1-Trunc-FLAG (gift from Dr Chan, Memorial Sloan-Kettering Cancer Center, NY, NY, USA), pcDNA3.1/Myc-His(−) B-TGM6 (gift from Prof. Li, State Key Laboratory of Medical Genetics, Changsha City, Hunan Prov, China), pCMV6-AC-turboGFP-PLD3 (gift from Dr Cruchaga, Washington University, St Louis, MO, USA), and pgC1A-SGFP2 (gift from Dr Goedhart, University of Amsterdam, Amsterdam, The Netherlands) were used as backbones to introduce the mutations using site-directed mutagenesis PCR following the protocol of the manufacturer (for primers see Supplementary Table 5). The plasmid sequences were verified by Sanger sequencing (for sequencing primers see Supplementary Table 6).

Cell culture and transfection

HEK293T and COS-7 cells were grown in Dulbecco's modified Eagle medium supplemented with 10% foetal bovine serum and 1% penicillin-streptomycin in a 37°C incubator with 5% CO_2. Transfections were performed using polyethylenimine (Polysciences) according to the manufacturer's instructions. Cells were grown on glass cover slips in 24-well plates for immunocytochemistry or in 6-well plates for western blotting, quantitative RT-PCR and wound healing assays. Cells were cultured 48 h post-transfection.

Immunocytochemistry

Cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 15 min at room temperature and washed three times with PBS. With the exception of cells transfected with GFP/EGFP/turboGFP tagged proteins, cells were permeabilized and blocked in 0.1% Triton™ X-100, 5% foetal bovine serum in PBS for 30 min at room temperature, followed by incubation overnight at 4°C with primary antibodies including mouse anti-His (Santa Cruz Biotechnology; 1:100), mouse anti-Calnexin (Santa Cruz Biotechnology; 1:50), rabbit anti-turboGFP (Evrogen; 1:1000), and mouse anti-Myc (Cell Signaling Technology; 1:300) antibodies in phosphate buffer. The next day, the cells were washed with PBS and incubated with secondary antibodies including mouse and rabbit anti-Alexa Fluor® 488 and mouse and rabbit anti-Cy3 antibodies (Santa Cruz Biotechnology; both 1:500) for 1 h at room temperature in phosphate buffer. All slides were mounted in Vectashield medium with 4', 6-diamidino-2-phenylindole (DAPI; Vector Laboratories). Stack images were obtained using a Leica DMI 6000 Inverted microscope and processed using ImageJ software (http://fiji.sc/, National Institutes of Health, Bethesda, MA, USA).

Cell fractionation

Subcellular fractionation was performed to separate nuclear, membrane, and cytoplasmic fractions using a centrifugation method. Cells were lysed with 250 µl of LB buffer [250 mM sucrose, 20 mM HEPES (pH 7.4), 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM EDTA plus protease inhibitor cocktail (Roche Diagnostics)]. Lysates were passed through a 25-gauge needle 10 times using a 1 ml syringe and incubated on ice for 20 min. To obtain the nuclear pellet, lysates were centrifuged at 720*g* for 5 min at 4°C. Supernatants were kept and the nuclear pellet was washed once by adding 500 µl of LB buffer and passed through a 25-gauge needle 10 times, then centrifuged at 720*g* for 10 min. The pellet was resuspended with 1× Laemmli buffer containing 10% β-mercaptoethanol and sonicated briefly. To obtain the cytosolic and membrane fractions, supernatants were centrifuged at 13 500*g* for 15 min at 4°C, and the pellet was resuspended in 1× Laemmli buffer containing 10% β-mercaptoethanol and sonicated. Fractions were boiled for 5 min at 95°C and analysed by western blot.

Western blot

After 48 h of transfection, cell extracts were homogenized using a 2% sodium dodecyl sulphate (SDS)/PBS buffer containing protease inhibitor cocktail (Roche Diagnostics). Proteins were quantified using the BCA kit (Thermo Fisher Scientific) and the supernatant was mixed with loading sample buffer containing 10% β-mercaptoethanol and boiled for 5 min at 95°C.

For protein analysis, 50 µg or 100 µg of proteins was loaded on SDS-polyacrylamide electrophoresis gel followed by immunoblot analysis. Nitrocellulose membranes were incubated with rabbit anti-turboGFP (Evrogen, 1:15000), mouse anti-Flag (Sigma; 1:1000), mouse anti-actin (MP Biochemicals; 1:5000), mouse anti-HA (Roche Diagnostics, 1:5000), mouse anti-Myc (Cell Signaling Technology; 1:2500), rabbit anti-GRP78/BIP (Abcam; 1:5000), and mouse anti-Histone 3 (Abcam, 1:5000) antibodies. Protein densitometry was performed using QuantityOne® (Bio-Rad).

Reverse transcription and quantitative real-time PCR

Total RNA was extracted from snap-frozen mouse tissues and HEK293T cells using TRIzol® (Life Technologies, Invitrogen). Complementary DNA was generated using oligo-d(T) primers and the RevertAid™ cDNA kit (Thermo-Scientific, Fermentas) according to the manufacturer's protocol. Reverse transcription-PCR was performed using AmpliTAQ Gold 360 Master mix (Thermo Fisher Scientific) and PCR products were analysed using gel electrophoresis. Quantitative real time expression analysis was performed as described before (Smeets *et al.*, 2015). The list of primers used is given in Supplementary Table 7.

Wound healing assay

The wound-healing assay was performed in COS-7 cells grown to confluence in 6-well plates. A line was scratched with a 200 µl plastic pipette tip in the cell monolayer and cells were washed by replacement of the medium. Photomicrographs were taken at 6 h and 24 h post scratching using an EVOS FL fluorescence microscope (Thermo Fisher Scientific). Wound edge distance was measured at three different points with ImageJ software.

Aggregation assay

The aggregation assay was performed as described (Matsui *et al.*, 2008) with some modifications. Transfected HEK293T cells were washed with PBS, trypsinated with 0.01% trypsin containing 1 mM $CaCl_2$ at 37°C for 15 min. Cells were collected, washed with Ca^{2+} and Mg^{2+} free HEPES-buffered (pH 7.4) Hanks' balanced salt solution (HBSS) and passed through a 100 µm cell strainer to obtain a single cell suspension. Cells in HBSS with or without 1 mM $CaCl_2$ were incubated in albumin-coated 48-well plates shaking at 130 rpm for 1 h. Photomicrographs were taken using an EVOS FL fluorescence microscope (Thermo Fisher Scientific). Aggregation was quantified using the ImageJ software.

Phospholipase D activity

COS-7 cells were transfected with an empty GFP vector or GFP-PLD3 wild-type and GFP-PLD3-L308P expressing plasmids. PLD3 activity was measured using a colorimetric assay [Phospholipase D (PLD) Activity Colorimetric Assay Kit; BioVision, San Francisco] following manufacturer's instructions.

Gene network analysis

Previously we described a method to predict the likely function of genes using 2206 principal components identified using gene co-expression based on a compendium of 77 840 samples (Fehrmann *et al.*, 2015). The input is a gene set (here the established set of SCA genes at the time of analysis; $n = 24$) and it is determined for each of the components whether the principal component is informative for this gene set (T-test). Subsequently, the evidence from each of the 2206 components is combined, and for each gene a *P*-value is calculated that indicates the likelihood that the gene is involved in the gene set. Please note that given the known SCA genes are used as seeds to generate the network, they do not appear in the gene-network analysis gene list.

We first determined whether the set of known SCA genes form a coherent set of genes, using a leave-one-out cross-validation. By leaving one known SCA gene out and then ascertaining the rank of this left-out gene we could calculate an area under the curve and a *P*-value (Wilcoxon Mann-Whitney U-test) that allowed us to predict the genes that cause SCA are significantly co-regulated. (This is visualized on genenetwork.nl, click on 'method'). We subsequently ascertained whether adding the five newly identified genes to the gene set increased the performance (again using the leave-one-out strategy) and compared this to the effect of adding five randomly chosen genes. We repeated this 100 times, permitting us to empirically determine whether the five genes are significantly better co-regulated to the known SCA genes than five randomly chosen genes.

We subsequently ascertained all protein-coding genes. Per gene we determined whether that gene was significantly stronger co-regulated with the known SCA genes as compared to co-regulation with all other genes (Wilcoxon Mann-Whitney U-test). If we selected those genes most strongly co-regulated with the known SCA-genes, without requiring that they show limited co-regulation with all other genes, we would end up with hub-genes that show strong global co-regulation. Accordingly, this method enables us to select and focus on those genes that are specifically co-regulated with known SCA genes.

Protein modelling

The full-length amino acid sequences of CACNA1 isoform 2 and AGF3L2 proteins were obtained from the RCSB Protein Data bank (http://www.rcsb.org/pdb/home/home.do) (PDB entries 2VAY and 2LNA2, respectively). Predicted 3D structures were obtained by homology-based or threading structural modelling engines [Phyre2 (Mezulis *et al.*, 2016), I-TASSER (Yang *et al.*, 2015), Raptor-X (Källberg *et al.*, 2012) and SWISS-Model (Biasini *et al.*, 2014)], using different templates from the PDB. For CACNA1–2, the most significant model that included the p.Asp1341Tyr mutation site was obtained from the electron microscopy (EM) structure of CACNA1S (Wu *et al.*, 2016) and by merging structural information from models built by both Phyre2 and Raptor-X using the 5GJV and 5X0M cryo-EM structures, respectively. In the case of the AGF3L2 protein, all algorithms provided similar modelling results based on the 2DHR crystal structure, and we have used the I-TASSER model here. Visualization, graphics and virtual mutagenesis were performed using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.)

Results

Whole exome sequencing identified multiple genes linked to SCA

We used WES to identify the mutations and corresponding genes underlying the dominant cerebellar ataxia in 20 Dutch families who tested negative for mutations in the SCA1–3, 6, 7, 17, 19/22, and 23 genes. We performed WES on 40 individuals; two or three affected cases selected from multiplex families ($n = 8$), trios ($n = 2$), one parent plus one affected case ($n = 5$), and singletons from simplex families ($n = 5$) (Supplementary Table 1). WES showed an average coverage of $>20\times$ for 81.7% of the target regions (Supplementary Table 2) and ~82 000 high-quality variants were identified, per individual, of which ~35% were coding variants. Upon exclusion of variants with a minor allele frequency $>0.01\%$ in known databases and our in-house database of over 200 exomes and 500 genomes, variants were prioritized by their predicted deleterious effect and amino acid conservation. If possible, we tested segregation of each candidate variant meeting these criteria.

Our analysis resulted in the identification of mutations in four known SCA genes that were not, at the moment of executing this work, part of routine diagnostic screening [20% (4/20) of the families in our cohort] (Supplementary Tables 1 and 8 and Supplementary Fig. 2A–D). The genes involved were: *CACNA1A*, which has been linked to SCA6, episodic ataxia type 2, and familial hemiplegic migraine [OMIM 183086, OMIM1 08500, and OMIM 141500 (Ophoff *et al.*, 1996; Zhuchenko *et al.*, 1997)]; *PRKCG*, linked to SCA14 [OMIM 605361 (Chen *et al.*, 2003)]; *TMEM240*, linked to SCA21 [OMIM 607454 (Delplanque *et al.*, 2014)]; and *AFG3L2*, linked to SCA28 [OMIM 610246 (Di Bella *et al.*, 2010)]. The variants c.4645C>T; p.Arg1549* in *CACNA1A*, c.239C>T; p.Thr80Met in *TMEM240*, and c.1996A>G;p.M666V in *AFG3L2* have all been previously reported to be pathogenic (Jen *et al.*, 1999; Cagnoli *et al.*, 2010; Delplanque *et al.*, 2014). The predicted damaging variant c.187G>C; p.Gly63Arg in the highly conserved C1A subdomain of *PRKCG* has not been described before (Supplementary Table 8). Notably, another mutation of glycine 63, a substitution to valine, has been reported to cause SCA14, suggesting that the p.Gly63Arg variant is very likely disease-causing (Nolte *et al.*, 2007). Additionally, all three missense mutations affected conserved amino acids and co-segregated with the disease (Supplementary Fig. 2E–G).

We identified three novel candidate genes in multiplex families (Fig. 1A–C) in which only a single deleterious variant segregated (Supplementary Table 3). These included *FAT2* (FAT atypical cadherin 2), *PLD3* (phospholipase D family member 3), and *KIF26B* (Kinesin family member 26B) (Table 1). All three were predicted to be highly deleterious and to affect conserved amino acids (Fig. 1D–F). The variations (c.10758G>C; p.Lys3586Asn in *FAT2*, c.923T>C; p.Leu308Pro in *PLD3*, were located in the 34th cadherin repeat domain and the PLD phosphodiesterase domain 2, respectively) (Fig. 1I and J). Variant c.5710G>A; p.Asp1904Asn in *KIF26B* was located in C-terminal tail region that binds to Nedd4, an E3-ubiquitin ligase (Terabayashi *et al.*, 2012) (Fig. 1K).

The three families all suffered from slowly progressive cerebellar ataxia with a spectrum of additional neurological symptoms (Table 2). In short, the proband of Family RF14 carrying the c.10758G>C; p.Lys3586Asn mutation in *FAT2* exhibited a relatively pure cerebellar syndrome including limb and gait ataxia, downbeat nystagmus, and dysarthria. The disease presented after 40 years of age. Unfortunately, no detailed clinical information was available for the remaining affected family members. A subset of Family RF28 carrying the c.923T>C; p.Leu308Pro variant in *PLD3* had been clinically described by van Dijk *et al.* (1995). A recent re-evaluation of the extended family confirmed a variable combination of sensory neuropathy and cerebellar ataxia in affected family members. In a small minority of them, the cerebellar phenotype was more prominent than the sensory neuropathy. Cerebellar dysarthria was absent in some, but present in most cases and oculomotor examination was abnormal in all but one case. The average age of onset was relatively late, 53.5 years, and ranged from 35 to almost 70 years. The affected cases of Family 2002-0206 carrying the c.5710G>A; p.Asp1904Asn mutation in *KIF26B* had late onset (average 51 years) of symptoms that included spasticity and gait/limb ataxia. The disease progression was very slow and the oldest patient is currently 92 years old. No ocular movement disorders or other features were seen in any of these patients. The MRIs of two patients showed very mild cerebellar atrophy. Given that these candidate genes were not previously linked to SCA, additional evidence is required to prove pathogenicity.

No mutations were detected in known SCA genes or in the novel genes *FAT2*, *PLD3*, and *KIF26B* in 13 families. In two of these families, no candidate variant fulfilled our selection criteria. However, in the remaining 11 families, multiple candidate variants met our selection criteria (Supplementary Table 1), and the majority co-segregated within the families (Supplementary Table 3). Given that the variant list was very long from the families for whom we only exome-sequenced a single case ($n = 4$; Supplementary Table 1), we excluded these variants from further analysis for practical reasons. Thus, WES very likely identified three novel SCA genes (*FAT2*, *PLD3*, and *KIF26B*) and additional putative candidate genes in these families that needed further investigation.

Validation of results in an independent SCA cohort

To validate *FAT2*, *PLD3*, and *KIF26B* as novel SCA genes and obtain further evidence for involvement of the other putative candidate genes ($n = 28$) in the pathogenesis of SCA (MacArthur *et al.*, 2014), we screened an additional cohort of 96 unrelated cerebellar ataxia patients [familial and sporadic; no mutations in SCA1–3, SCA6 (polyQ), SCA 7, and SCA 17] with a dedicated gene panel for mutations in these genes. For completeness, the candidate gene panel was complemented with 14 known but rarer SCA genes that are not routinely screened in current genetic diagnostics (Supplementary Table 4).

We identified an additional variation, c.10946G>A; p.Arg3649Gln, affecting a highly conserved amino acid located in the linker between the last cadherin repeat and the laminin A-G motif in *FAT2* in an apparently sporadic case with unknown family history (Fig. 1G–I and Table 1), supporting a role for *FAT2* in the pathogenesis of SCA. The patient suffered from late onset (~50 years of age) slowly progressive gait and limb ataxia and from dysarthria. MRI showed atrophy of the cerebellar vermis and hemosiderine depositions in the mesencephalon (Table 2). No additional mutations were identified in *PLD3* and *KIF26B* in this independent validation cohort, making further functional evidence necessary to understand the potential role of these genes in SCA.

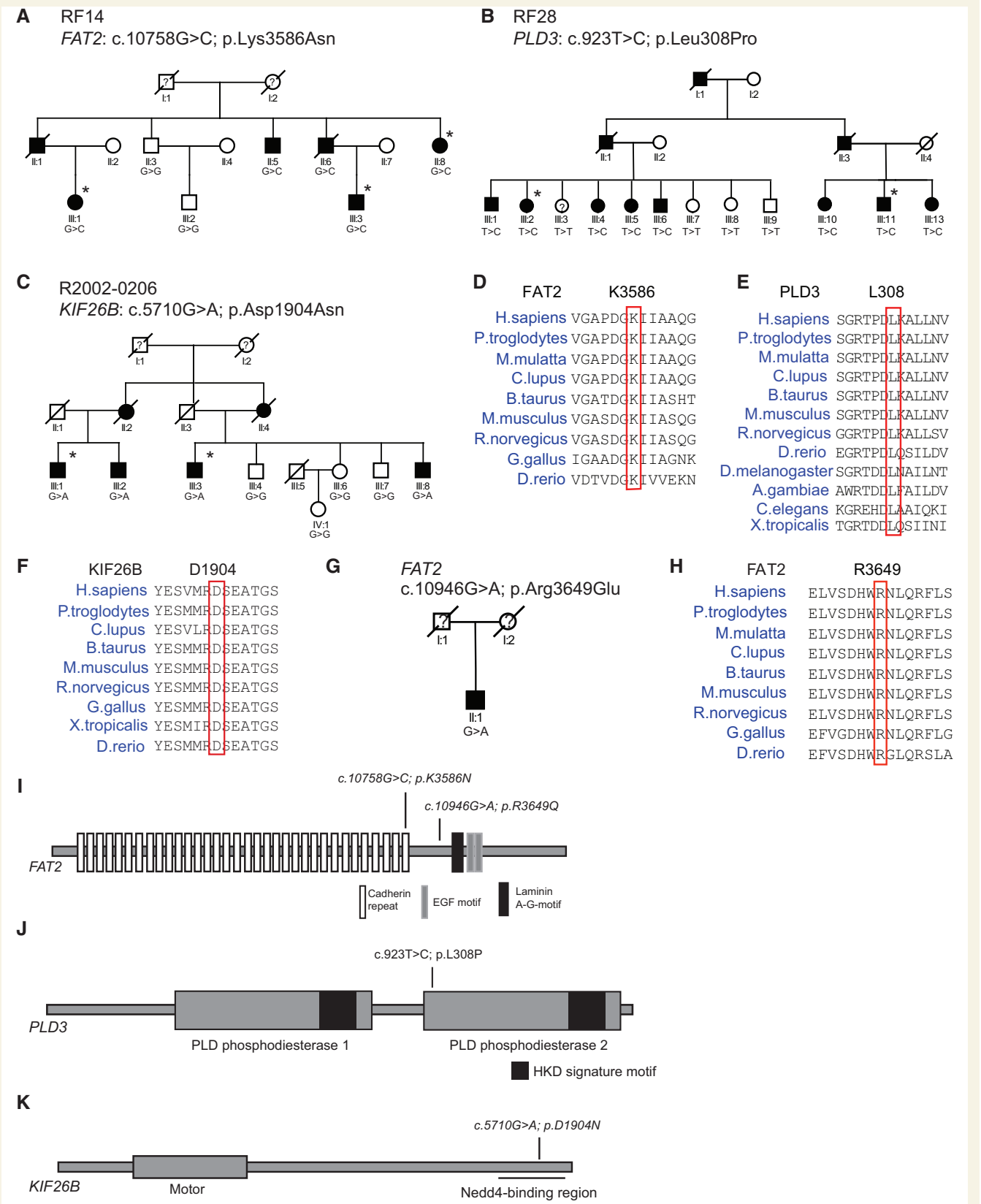

Figure 1 Mutations in *FAT2*, *PLD3*, and *KIF26B* causing SCA. (**A**) Pedigree of Family RF14 carrying the c.10758G>C; p.Lys3586Asn mutation in *FAT2*. The mutation was identified in five affected individuals but not in two unaffected family members. (**B**) Pedigree of Family RF28 carrying the c.923T>C; p.Leu308Pro in *PLD3*. The mutations were identified in eight affected individuals and were absent in three unaffected family members. One family member with mild complaints did not carry the variant. (**C**) Pedigree of the family carrying the c.5710G>A; p.Asp1904Asn in *KIF26B*. The variant was detected in the four affected family members but was absent in the four unaffected relatives. Closed

(continued)

Two of the putative candidate genes *FAT1* and *EP300* (E1A binding protein p300), extracted from the WES data of RF25 and RF19, respectively, were both mutated in multiple unrelated cases. In addition to the affected cases of Family RF25, three individuals in the replication cohort carried *in silico* predicted damaging variants in *FAT1* (Fig. 2A–D and Table 2) that change conserved amino acids (Fig. 2E). The *FAT1* mutations were located in cadherin repeats (numbers 5, 10, and 19) and in the intracellular C-terminal tail (Fig. 2F). Moreover, both affected family members in RF19 and one case from the replication cohort carried different mutations in *EP300* (Fig. 2G, H). The truncating mutation c.6693-6694insC; p.Gln2232fs70X found in the two members of Family RF19 led to a loss of the C-terminal part of the transactivation domain of *EP300* (Fig. 2J). The missense mutation c.6020A>G; p.Gln2007Arg identified in one male (Case DNA008784) in the replication cohort changed a conserved amino acid located in the linker preceding the transactivation domain (Fig. 2I and J).

The cases in whom the mutations in *FAT1* and *EP300* were identified displayed variable phenotypes (Table 2). These ranged from late-onset progressive limb ataxia, dysarthria, and downbeat nystagmus in one female (Case DNA004952) carrying the c.13265C>T; p.Thr4422Met mutation in *FAT1*, to childhood-onset episodic ataxia of the limbs and dysarthria in one female (Case DNA-057446) carrying the c.5788G>C; p.Asp1930His mutation in *FAT1*, to late-onset slowly progressive cerebellar ataxia in one male (Case DNA008784) carrying the c.6020A>G; p.Gln2007Arg mutation in *EP300*. For all the cases in the replication cohort, no additional family members were available for genetic testing.

In eight cases from this cohort we also identified mutations in known, but rarer, SCA genes including *CACNA1A*, *PRKCG*, *KCND3* (SCA19/22; OMIM 607346), *AFGL3L2* (SCA28; OMIM 604581), and *TGM6* (SCA35; OMIM 613908) (Supplementary Table 9). For the two novel mutations in *CACNA1A* (c.4021G>T; p.Asp1341Tyr) and *AFG3L2* (c.1861C>G; p.Leu621Val), segregation with the disease was observed and the mutations affected highly conserved amino acids (Supplementary Fig. 3A–D). Additionally, structural modelling showed that novel c.4021G>T; p.Asp1341Tyr variant in CACNA1A, homologous residue in CACNA1S is 889, is located at a critical position of the channel subunit that may serve as a major entrance site for calcium ions (Wu *et al.*, 2016). Additionally, changing the negatively charged aspartate to a polar/aromatic tyrosine residue is predicted to have a vibrant effect on the functionality of the channel (Supplementary Fig. 3E). The modification of the leucine to valine at position 621 in AGF3L2 could be considered of less structural importance as valine is only one methylene group smaller than leucine. Yet, the position of the amino acid change is located in a loose hydrophobic group of α-helixes and, therefore, might lead to a folding problem or less strong stabilization of the bundle of helixes (Supplementary Fig. 3F). No additional family members were available to test for segregation of the novel c.1277C>A; p.Thr426Asn mutation in *TGM6* (Supplementary Fig. 3G), but the mutation also affected a highly conserved amino acid (Supplementary Fig. 3H). Thus, mutations in the known but rarer SCA genes account for 8.3% (8/96) of the cases in our independent SCA cohort.

Functional validation of novel mutations in known SCA genes using cell models

First we validated the putative pathogenic effect of the novel mutation in *TGM6* (found in Case DNA009145) for which we were unable to perform co-segregation analysis. To this end, we investigated the characteristics of the mutant protein in HEK293T cells, which also has been used to study the consequences of two other reported *TGM6* mutations (Guan *et al.*, 2013). We determined the protein localization of wild-type (WT) TGM6 and TGM6-T426N in transiently transfected HEK293T cells. Immunocytochemistry showed that TGM6-T426N exhibited increased endoplasmic reticulum (ER) localization compared to TGM6-WT (Supplementary Fig. 4A–C). This coincided with significantly reduced protein stability of mutant TGM6 compared to TGM6-WT upon cycloheximide (CHX) treatment for 6 and 14 h as shown by western blotting [wild-type: 98.7 ± 13.5% (6 h) and 37.3 ± 12.3% (14 h) versus T426N: 65.3 ± 2.5% (6 h)

Figure 1 Continued

symbols = affected; open symbols = unaffected; question mark = disease status unknown; forward slash = deceased; asterisk indicates individuals were used in the WES analysis. (**D–F**) Multiple sequence alignments showing the amino acid sequence homology of the affected amino acids. (**G**) Pedigree of the patient with the second *FAT2* mutation: c.10946G>A; p.Arg3649Gln. (**H**) Multiple sequence alignment showing the amino acid sequence homology of the affected amino acid. (**I**) Schematic representation of *FAT2* encoding FAT atypical cadherin 2, a very large protein (480 kDa) containing two EGF-like motifs, one laminin A-G motif, and 34 cadherin repeat domains. DNA sequence analysis identified two missense mutations in SCA subjects: c.10758G>C; p.Lys3586Asn and c.10946G>A; p.Arg3649Gln in the last cadherin repeat and the linker between the last cadherin repeat and the first laminin A-G motif. (**J**) Schematic representation of the structure of *PLD3*, a single-pass type II membrane protein containing two phosphodiesterase domains with HKD signature motifs. DNA sequence analysis identified a missense mutation, c.923T>C; p.Leu308Pro, in the second phosphodiesterase domain of *PLD3*. (**K**) Schematic representation of the structure of *KIF26B*, a dynein motor protein containing a motor domain. DNA sequencing identified a missense mutation, c.5710G>A; p.Asp1904Asn in an interaction motif for Nedd4, an E3-ubiquitin ligase located in the C-terminus of *KIF26B*.

Table 1 Novel candidate SCA genes

Gene symbol	Gene name	Putative biological function	Family #	Position (hg19)	Nucleotide change	Deduced protein change	Segregation within family	SIFT	Poly Phen-2	Mutation Taster	1000G, %	ExAC, %
FAT2	FAT atypical cadherin 2	Cell adhesion molecule	RF14	chr5:150901396	c.10758G>C	p.Lys3586Asn	Yes	D	PD	D	Absent	8.279×10^{-6}
			DNA056251	chr5:150901208	c.10946G>A	p.Arg3649Gln	N.A.	D	PD	D	Absent	1.676×10^{-5}
KIF26B	Kinesin family member 26B	Microtubule binding protein	2002-0206	chr1:245851995	c.5710G>A	p.Asp1904Asn	Yes	D	PD	D	Absent	4.14×10^{-5}
PLD3	Phospholipase D family, member 3	Phospholipase	RF28	chr19:40880431	c.923T>C	p.Leu308Pro	Yes	D	PD	D	Absent	Absent
FAT1	FAT atypical cadherin 1	Cell adhesion molecule	RF25	chr4:187584592	c.3441A>C	p.Glu1147Asp	Yes	T	Pos D	D	Absent	Absent
			DNA004952	chr4:187510248	c.13265C>T	p.Thr4422Met	N.A.	T	PD	D	Absent	Absent
			DNA057446	chr4:187541952	c.5788G>C	p.Asp1930His	N.A.	T	PD	D	Absent	8.283×10^{-6}
			DNA003292	chr4:187629345	c.1637C>T	p.Pro546Leu	N.A.	D	PD	D	Absent	Absent
EP300	E1A binding protein p300	Histone acetyl-transferase	RF19	chr22:41574408	c.6693_6694insC	p.Gln2232fs70X	Yes	N.A	N.A.	N.A.	Absent	Absent
			DNA008784	chr22:41573735	c.6020A>G	p.Gln2007Arg	N.A.	T	Pos D	D	Absent	8.245×10^{-6}

D = damaging; N.A. = not analysed; PD = probably damaging; Pos D = possibly damaging; T = tolerated.

Table 2 Clinical features of affected individuals from families with mutations in candidate SCA genes

Family number #	Gene symbol	Mutation	Average AoO, years	Progression	Ataxia type	Dysarthria	Nystagmus	Familial/ sporadic	MRI	Other features
RF14	*FAT2*	p.Lys3586Asn	<60	Slow	Gait/limb	Yes	Downbeat	Familial	Cerebellar atrophy	
DNA056251	*FAT2*	p.Arg3649Gln	50	Slow	Gait/limb	Yes	No	Sporadic	Haemosiderine depositions in mesencephalon, vermal cerebellar atrophy	Cerebral aneurism
RF25	*FAT1*	p.Glu1147Asp	60	Slow	Gait/limb	Yes, but mild	Downbeat	Familial	Pronounced sulci cerebellar hemispheres	Macular degeneration, intention tremor extremities
DNA004952	*FAT1*	p.Thr4422Met	65	Progressive	Limb	Yes	Downbeat	Familial	Cerebellar and cerebrum atrophy	Pyramidal features, sensory dysfunction
DNA057446	*FAT1*	p.Asp1930His	<10	U	Limb	Yes	U	Familial	NA	Episodic
DNA053310	*FAT1*	p.Ile1478Thr	70	Progressive	Gait/limb	Yes	No	U	Mild cerebellar atrophy	Diabetic polyneuropathy, essential thrombocytosis
DNA003292	*FAT1*	p.Pro546Leu	<55	Slow	Cerebellar ataxia	U	U	Sporadic	NA	
2002-0206	*KIF26B*	p.Asp1904Asn	Range 74–90	Slow	Spastic ataxia	Yes	U	Familial	NA	
RF19	*EP300*	p.Gln2232fs70X		Slow	Gait	U	U	Familial	Loss parenchym vermis	
DNA008784	*EP300*	p.Gln2232Arg	20	Slow	Cerebellar ataxia	U	U	Sporadic	Cerebellar atrophy	Spasticity
RF28	*PLD3*	p.Leu308Pro	53.5, range 35–69	Slow	Mixed sensory and cerebellar; Gait>>limb	Variable: absent to severe	Variable: jerky pursuit, SWJ, GEN, DBN, slow saccades, saccadic dysmetria	Familial	Absent or mild cerebellar atrophy	Mild to predominant sensory axonal neuropathy

AD = autosomal dominant; AoO = age of onset; DBN = downbeat nystagmus; GEN = gaze-evoked nystagmus; N.A. = not available; SWJ = square-wave jerk; U = unknown.

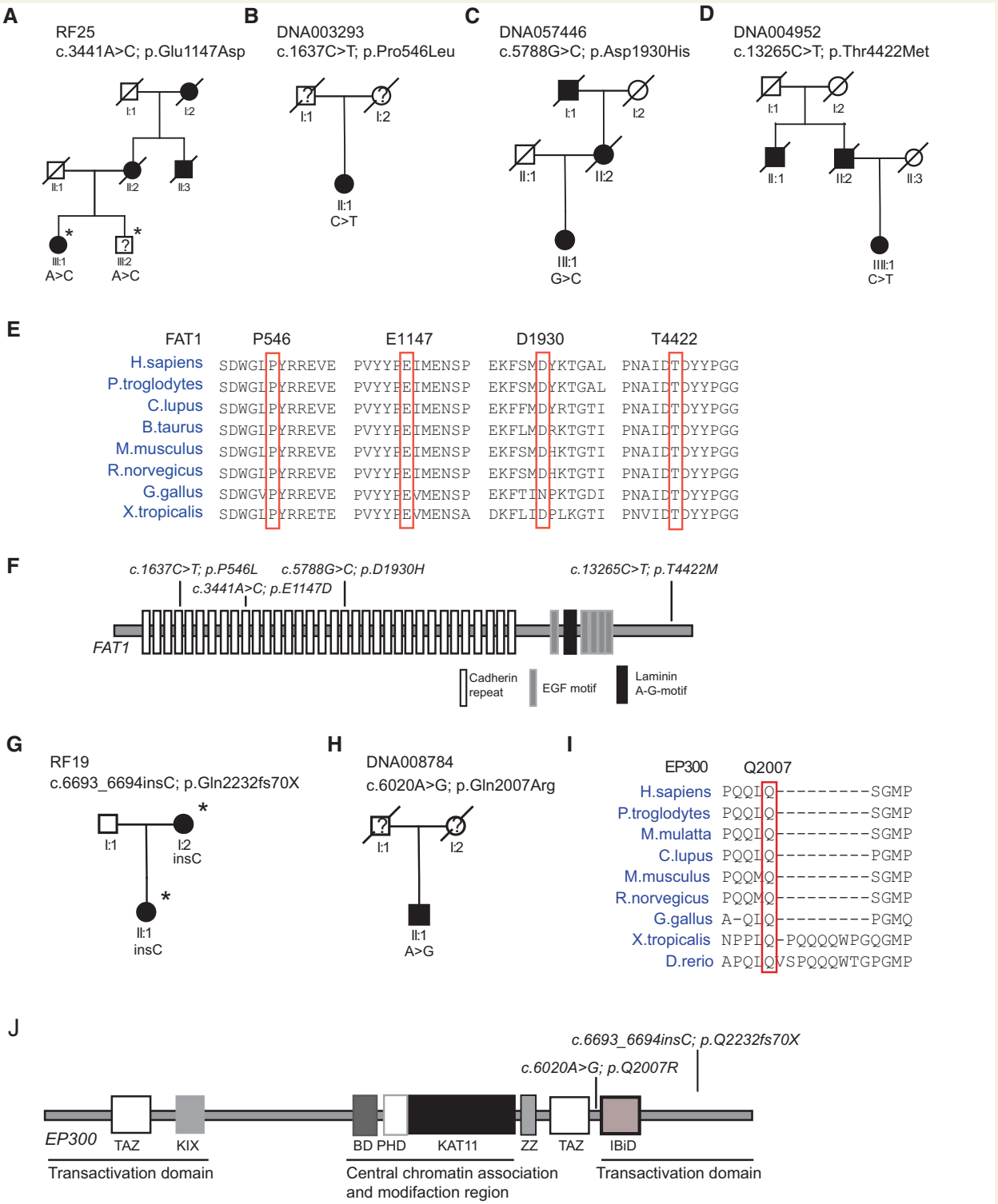

Figure 2 Mutations in *FAT1* and *EP300* causing SCA. Pedigrees of the cases carrying *FAT1* mutations: (**A**) c.3441A>C, p.Glu1147Asp; (**B**) c.1637C>T, p.Pro546Leu; (**C**) c.5788G>C, p.Asp1930His; and (**D**) c.13265C>T, p.Thr4422Met. No additional family members were available for genetic testing for Cases DNA003293, DNA057446 and DNA004952. Closed symbols = affected; open symbols = unaffected; question mark = disease status unknown; forward slash = deceased; asterisk indicates individuals were used in the WES analysis. (**E**) Multiple sequence alignments showing the amino acid sequence homology of the affected amino acids. (**F**) Schematic representation of *FAT1* encoding FAT atypical cadherin 1, containing 5 EGF-like motifs, 1 laminin A-G motif, and 34 cadherin repeat domains. Three missense mutations: c.3441A>C, p.Glu1147Asp; c.1637C>T, p.Pro546Leu; and c.5788G>C, p.Asp1930His, are located in cadherin repeat domains whereas mutation c.13265C>T, p.Thr4422Met is located in the C-terminal tail of *FAT1*. Pedigrees of the cases carrying *EP300* mutations. (**G**) c.6693_6694insC, p.Gln2232fs70X. (**H**) c.6020A>G, p.Gln2007Arg. No additional family members were available for genetic testing for Case DNA008784.

(continued)

and 10.8 ± 3.6% (14 h); Supplementary Fig. 4D and E]. To confirm that mutant TGM6 sensitizes cells to apoptosis, as previously reported (Guan *et al.*, 2013), we treated transiently transfected HEK293T cells expressing either an empty vector, TGM6-WT, or TGM6-T426N with calcium ionophore (A23187) for 30 h. Both TGM6-WT and TGM6-T426N led to significantly increased cell death upon A23187 treatment compared to the empty vector transfected cells (Supplementary Fig. 4F and G). TGM6-T426N apparently induced more cell death than TGM6-WT, but this observation was borderline significant. Notably, no increased cell death was observed in untreated conditions (data not shown). Based on the marked altered cellular localization and significantly reduced protein stability of TGM6-T426N, we suggest that this variant has damaging effects.

Second, we determined whether the mutations p.Gly63Val and p.Gly63Arg in *PRKCG* located in the phorbol-ester responsive C1A domain also affect phorbol-ester-induced membrane translocation, as was reported for established SCA14 mutations located in the C1B subdomain (Verbeek *et al.*, 2008). Therefore, we transiently transfected HEK293T cells, which are successfully used to study the translocation pattern and thereby activation of protein kinase C's (PKCs) (Hui *et al.*, 2014), with wild-type C1A-SGFP2 and Gly63Val and Gly63Arg mutant C1A-SGFP2 (Supplementary Fig. 5A) following 10 min of phorbol 12-myristate 13-acetate (PMA) treatment. Whereas PMA was able to induce efficient membrane translocation of wild-type C1A-SGFP2, as shown by fluorescent microscopy, the mutant subdomains showed markedly reduced or no PMA-induced membrane translocation (Supplementary Fig. 5B and C). Additionally, the Gly63Arg mutant C1A-SGFP2 exhibited prominent perinuclear aggregates that were not observed for Gly63Val mutant or wild-type C1A-SGFP2 (Supplementary Fig. 5B). These data indicate that both p.Gly63Val and p.Gly63Arg mutations are functionally impairing phorbol-ether-induced translocation of the C1A subdomain and thus alter the activation of PKCγ, which very likely causes disease.

Functional validation of candidate genes with expression and cell models

To understand the potential role of the putative candidate genes *FAT2*, *FAT1*, *EP300*, *PLD3*, and *KIF26B* in SCA, we profiled their expression through different tissues using publicly available human gene expression data (Fehrmann *et al.*, 2015). Both these novel candidate genes and all known SCA genes had their highest average expression in cerebellum compared to other brain regions (Fig. 3). *FAT2* was almost exclusively expressed in the cerebellum, while the other candidate genes were also expressed in other brain regions (Supplementary Fig. 6A). This strongly supports a cerebellar role for *FAT2*, and further validates its role in the pathogenesis of SCA. However, *PLD3*, *KIF26B*, *FAT1*, and *EP300* showed broadly distributed expression patterns across multiple mouse tissues using RT-PCR, suggesting functions outside the brain as well (Supplementary Fig. 6B).

To reveal the consequences of the two mutations, p.Lys3586Asn and p.Arg3649Gln, on FAT2 localization, we transiently transfected C-terminally EGFP tagged *Fat2*-WT, *Fat2*-K3588N (corresponding to human K3586N), and *Fat2*-R3651Q (corresponding to human R3649Q) cDNAs in COS7 cells following protocol described for FAT1 (Moeller *et al.*, 2004). We assessed the subcellular localization of the different FAT2 proteins and noted significant increased co-localization with Golgin97, a Golgi-apparatus marker, by confocal microscopy (Fig. 4A and B). No differences in protein expression levels between the FAT2 proteins were detected via immunoblotting (Fig. 4C and D). We further determined the cell adhesion properties of the mutant FAT2 proteins in a similar fashion to what has been done for fat cadherin (Oda *et al.*, 1994; Matsui *et al.*, 2008). We performed an aggregation assay using HEK293T cells expressing the various FAT2 proteins and maintained the cells in the absence or presence of calcium for 1 h. The aggregation capacity of the cells was not increased by the expression of the various FAT2 proteins. In contrast, the addition of calcium elevated the aggregation properties of the cells in all conditions (Fig. 4E and F). In the cells expressing the mutant FAT2 proteins, the aggregation properties were further increased but only significantly higher in the cells expressing Fat2-K3588N compared to Fat2-WT and mock transfected cells (K3588N: 2.64 ± 0.06 versus WT: 1.72 ± 0.14 and mock: 1.86 ± 0.13).

Since Fat1 has been implicated in the strength of cell contacts at the wound margins and cell migration (Tanoue and Takeichi, 2004), we hypothesized that FAT2 may also be involved in cell migration and mutant FAT2 may alter this process. As in our hands HEK293T cells could not be used for a migration assay, we assessed the migration capacity conferred by various FAT2 proteins in

Figure 2 Continued

(**I**) Multiple sequence alignments showing the conservation of the affected amino acid. (**J**) Schematic representation of *EP300* encoding E1A binding protein p300. *EP300* is composed of two transactivation domains and a central chromatin association and modification region. The first transactivation domain contains a CREB and MYB interaction domain (KIX) and a cysteine/histidine region (TAZ). The second transactivation domain contains an interferon response binding domain (IBiD). The central region contains multiple domains including the bromodomain (BD), a plant homeodomain (PHD), and a catalytic domain (KAT11). The two mutations, c.6693_6694insC; p.Gln2232fs70X and c.6020A>G; p.Gln2007Arg, are located close to or within the second transactivation domain of *EP300*.

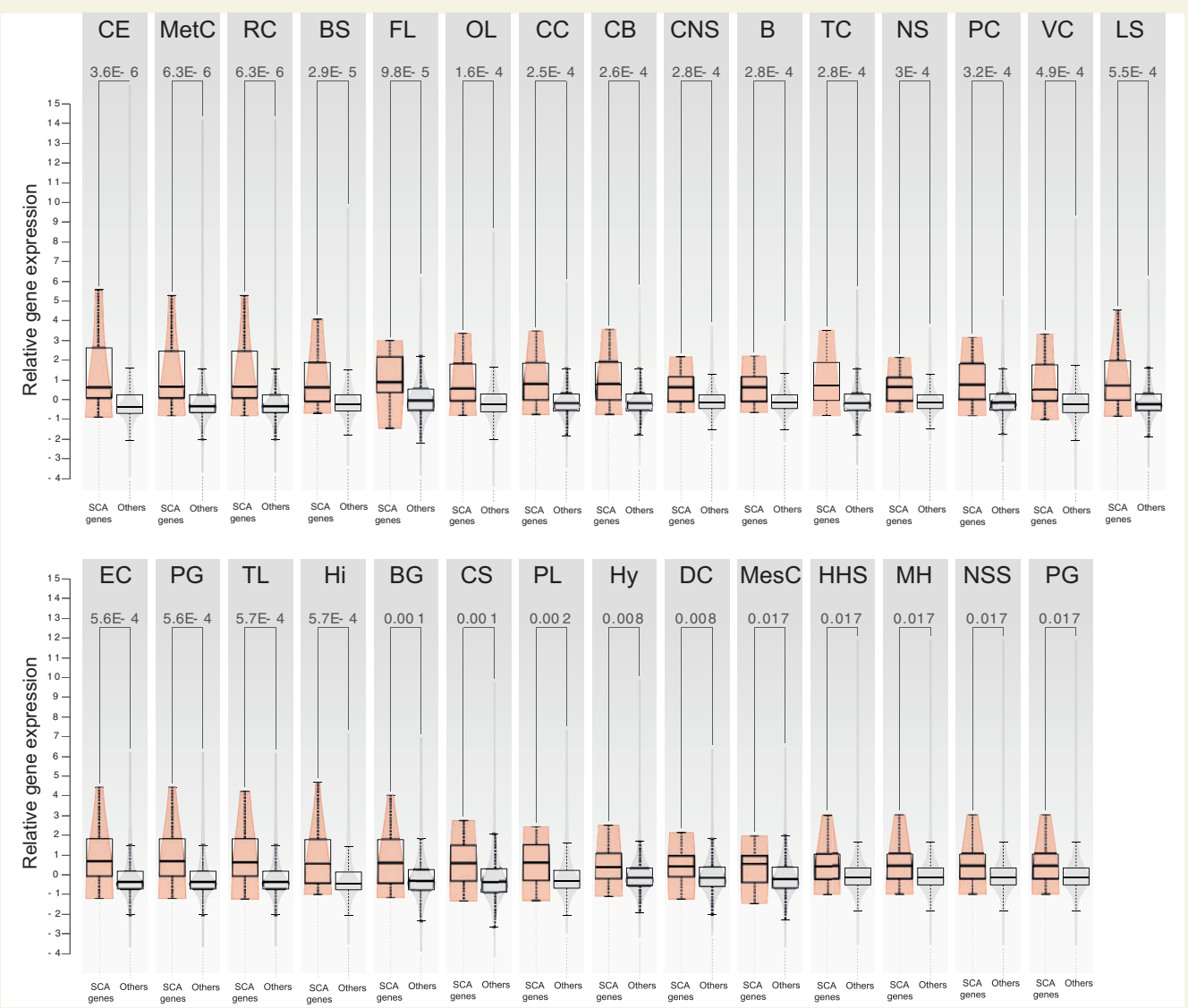

Figure 3 All known SCA genes and candidate genes are highly expressed in cerebellum. Box plot analysis of all known SCA genes and candidate SCA genes (SCA genes) mRNA expression levels compared to all other genes (others) over a range of human brain tissues extracted from publicly available RNA-seq data. The highest average SCA gene and candidate SCA gene expression levels were detected in cerebellum (CE), metencephalon (MetC), rhombencephalon (RC), and brainstem (BS). B = Brain; BG = basal ganglia; CB = cerebrum; CC = cerebral cortex; CS = corpus striatum; DC = diencephalon; EC = entorhinal cortex; FL = frontal lobe; HHS = hypothalamo-hypophyseal system; Hi = hippocampus; Hy = hypothalamus; LC = limbic system; MesC = mesencephalon; MH = middle hypothalamus; NS = nervous system; NSS = neurosecratory system; OL = occipital lobe; PC = prosencephalon; PG = parahippocampal gyrus; PG = pituitary gland; PL = parietal lobe; TC = telencephalon; TL = temporal lobe; VC = visual cortex.

transiently transfected COS7 cells by a wound-healing assay. Fat2-WT and mutant Fat2-overexpressing cells exhibited significantly reduced wound closure rates compared to the mock transfected cells at both 6 and 24 h post-scarring [Supplementary Fig. 7; GFP control: 69.5 ± 1.8% (6 h) and 100% (24 h); wild-type: 89.5 ± 1.6% (6 h) and 46.0 ± 4.8% (24 h); K3588N: 86.5 ± 2.4% (6 h) and 30.8 ± 4.7% (24 h); and R3651Q: 80.8 ± 1.8% (6 h) and 33.9 ± 4.2% (24 h)]. In contrast, no differences in wound closure rates were observed between Fat2-WT and mutant Fat2-expressing cells. Given that FAT2 seems to be a multifunctional protein, we speculate that the two variants do not affect the role of FAT2 in cell migration, but mainly change cell adhesion properties. Overall, the functional validation for Fat2-K3588N supports the genetic data, but additional functional work is necessary to reveal the role of Fat2 in cerebellar neurodegeneration. Additionally, p.Lys3586Asn could be a rare benign polymorphism and further investigation is warranted.

FAT1, among a range of functions, has been implicated in Hippo/Wnt signalling. It is thought to mediate the expression of several Hippo and Wnt targets, most likely via

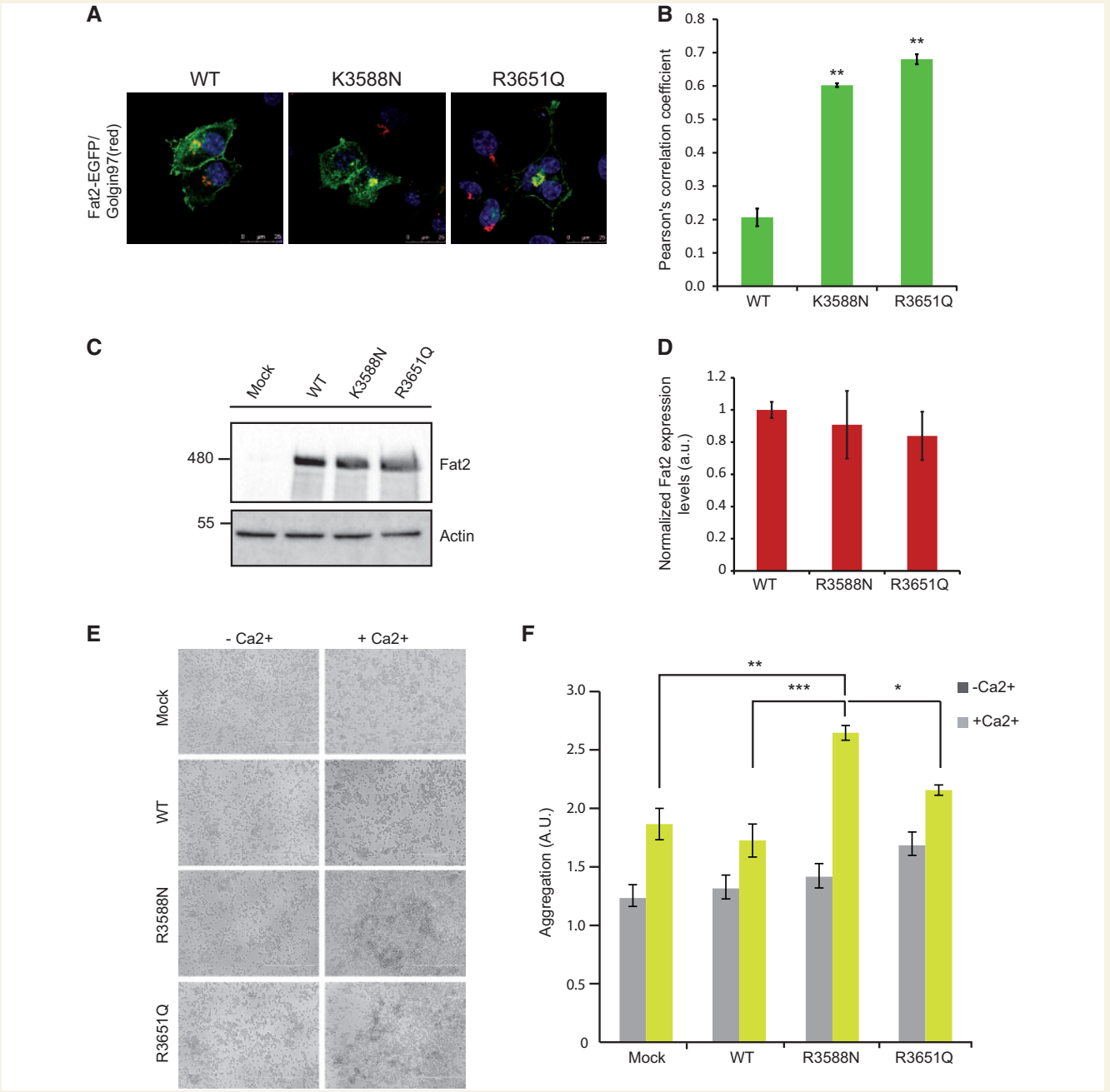

Figure 4 Fat2-K3588N exhibits increased cell adhesion characteristics. (**A**) Confocal images of transiently transfected COS-7 cells expressing Fat2-WT-EGFP, Fat2-K3588N-EGFP, and Fat2-R3651Q-EGFP stained with anti-Golgin97 (red), and DAPI (blue). Scale bar = 25 μm. (**B**) Quantification graph showing the Pearson's correlation coefficient indicative of the levels of co-localization. The missense mutations alter the cellular localization of FAT2 as more mutant FAT2 protein co-localized with the Golgi apparatus marker, Golgin97, compared to Fat2-WT (Fat2-WT: 0.21 ± 0.03; Fat2-K3588N: 0.60 ± 0.01; and Fat2-R3651Q: 0.68 ± 0.02, $P < 0.01$, ANOVA). (**C**) Immunoblot showing the expression levels of the various FAT2 proteins (Fat2-WT-HA, Fat2-K3588N-HA, and Fat2-R3651Q-HA) in HEK293T cells compared to mock transfected cells. Fat2-WT-HA has a predicted molecular weight of ~480 kDa. (**D**) Quantification graph of the immunoblot of lysates of HEK293T cells expressing EGFP (mock control), Fat2-WT-HA, Fat2-K3588N-HA, and Fat2-R3651Q-HA. No significant differences in expression levels were observed. Data are corrected for actin levels and normalized against Fat2-WT protein levels. (**E**) Representative pictures of an aggregation assay in dissociated HEK293T cells expressing EGFP (mock control), Fat2-WT-EGFP, Fat2-K3588N-EGFP, and Fat2-R3651Q-EGFP. The cells were treated with saline or calcium (1 mM) for 1 h before pictures were taken. Scale bar = 400 μm. (**F**) Quantification graph of the aggregation assay in HEK293T cells expressing EGFP (mock control), Fat2-WT-EGFP, Fat2-K3588N-EGFP, and Fat2-R3651Q-EGFP. The calcium treatment led to a significant increase in cellular aggregation in all cases. Mock: 1.23 ± 0.11 versus 1.86 ± 0.13 ($P < 0.01$). Fat2-WT: 1.31 ± 0.13 versus 1.72 ± 0.14 ($P < 0.05$). Fat2-K3588N: 1.41 ± 0.13 versus 2.64 ± 0.06 ($P < 0.0001$). Fat2-R3651Q 1.68 ± 0.13 versus 2.15 ± 0.04 ($P < 0.05$). Additionally, the expression of Fat2-K3588N led to a significant increase in cellular aggregation in the presence of calcium as compared to the Fat2-WT and Fat2-R3651Q expressing cells and the mock control (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, two-way ANOVA).

translocation of its N-terminal cytoplasmic domain to the nucleus by which it mediates neuronal differentiation (Magg *et al.*, 2005; Ahmed *et al.*, 2015). Since we were only able to obtain a truncated *FAT1* cDNA including the N terminus, the first two cadherin domains, all five epidermal growth factor (EGF) repeats, the transmembrane region, and the cytoplasmic domain [FAT1-Trunc-FLAG (Morris *et al.*, 2013)], we could only investigate the consequences of the C-terminal located p.Thr4422Met mutation (Fig. 2F). Similar gene expression levels of several Wnt targets detected by quantitative PCR were observed in transiently transfected HEK293T cells with FAT1-Trunc-WT (FAT1-WT) and FAT1-Trunc-T4422M (FAT1-T4422M) (Supplementary Fig. 8A). Given that in HEK293T cells we did not observe the inhibitory effect of FAT1-WT on the mRNA expression of these targets as previously reported in glioma cells (Morris *et al.*, 2013), checked the cellular distribution of FAT1-WT and FAT1-T4422M using cellular fractionation. Immunoblotting showed that both FAT1-WT and FAT1-T4422M translocated to the nucleus (Supplementary Fig. 8B). Based on these data we cannot exclude or confirm pathogenicity of the p.Thr4422Met variant. Future studies in a different cell line are required to assess the putative functional consequences of the p.Thr4422Met variant in *FAT1*.

To assess the functional consequence of the p.Leu308Pro mutation located in the second phosphodiesterase domain of *PLD3*, we determined the cellular localization of mutant PLD3 as shown by immunocytochemistry of COS7 cells expressing either PLD3-WT or PLD3-L308P (Fig. 5A). However, no differences in cellular localization between PLD3-WT and PLD3-L308P were detected (Supplementary Fig. 9). The p.Leu308Pro variant also did not alter PLD3's protein stability in COS7 cells upon CHX treatment up to 10 h as shown by western blot (Fig. 5B and C), nor did it affect the PLD3 protein maturation, as similar amounts of glycosylated PLD3 were detected in COS7 cell lysates transiently expressing PLD3-WT and PLD3-L308P (indicated by the asterisk in Fig. 5D). Additionally, tunicamycin treatment blocked glycosylation in a similar manner for both PLD3-WT and PLD3-L308P, and the cells expressing PLD3-L308P did not exhibit increased endoplasmic reticulum stress as compared to PLD3-WT-expressing cells as reflected by equal Bip levels (Fig. 5D). Finally, we determined the phospholipase D activity of PLD3 in transiently transfected COS7 cells. PLD3-L308P exhibited significantly reduced activity compared to PLD3-WT (PLD3-WT: 2.2 ± 0.2 versus L308P: 1.5 ± 0.06) (Fig. 5E), validating the damaging effect of the p.Leu308Pro variant.

Co-expression gene network analysis strengthens the role of candidate genes in SCA

To gain further insight into the functional role of the five candidate genes in SCA, we generated a gene network using all known SCA genes as seeds to predict functions of genes following previously described methodology (Fehrmann *et al.*, 2015), and assessed the contribution of the five candidate genes to the connectivity of this SCA gene network. The gene network based on known SCA genes at the time of analysis ($n = 24$) was more tightly co-expressed than a random set of genes (area under the curve = 0.59, Wilcoxon P-value = 0.138) indicating that the known SCA genes were more connected than expected for randomly selected genes (data not shown). The addition of the five candidate genes further increased the connectivity of the gene network genes (area under the curve = 0.64, Wilcoxon P-value = 0.007). The genes that were co-expressed with the known SCA genes are visualized in a gene network (Supplementary Fig. 10) and the top 100 co-expressed genes are listed in Supplementary Table 10, which was not observed for 100 random sets of five genes (empiric $P < 0.01$) (data not shown). This indicates that the candidate genes are more similar in biological function to known SCA genes than would be expected by chance.

Co-functionality analysis further suggests shared molecular pathways underlying SCA

Since all genetically distinct SCA types are characterized by pervasive Purkinje cell degeneration, we investigated whether the SCA genes (known and novel) converge into a limited number of specific biological pathways. Several SCA genes have been postulated to function in similar biological pathways including dysregulation of transcription, RNA-toxicity, protein misfolding, and synaptic transmission (Matilla-Dueñas *et al.*, 2014; Smeets and Verbeek, 2014). We assessed co-functionality of known and novel SCA genes as previously described (Pers *et al.*, 2015), and clustered these genes according to their predicted functionality (Fig. 6, Supplementary Fig. 11 and Supplementary Table 11). We identified a strong gene cluster comprised of many known SCA genes reflecting synaptic transmission (*TTBK2*, *ATXN1*, *ELOVL4*, *PDYN*, *FGF14*, *PRKCG*, *CACNA1A*, *SPTBN2*, *KCNC3*, *KCND3*, *TRPC3*, *PPP2R2B*, and *ITPR1)* that also contained the newly identified *FAT2* and *PLD3* genes, further validating their role in the pathogenesis of SCA. Two smaller clusters of genes (*NOP56*, *TBP*, *AFG3L2*, *ATXN10*, and *EEF2*; *FAT1*, *EP300*, *ATXN2*, and *ATXN7*) were both predicted to have a nuclear function. The fact that novel genes *FAT1* and *EP300*, both known to be involved in transcription regulation (Arany *et al.*, 1994; Morris *et al.*, 2013; Ahmed *et al.*, 2015), are part of these clusters validates the methodology. Nevertheless, some of the known and novel disease genes could not be mapped to one of these shared pathways, which suggests the involvement of other disease pathways. Taken together, our work reinforces that alterations in synaptic transmission and nuclear functioning, such as transcription regulation, cause SCA.

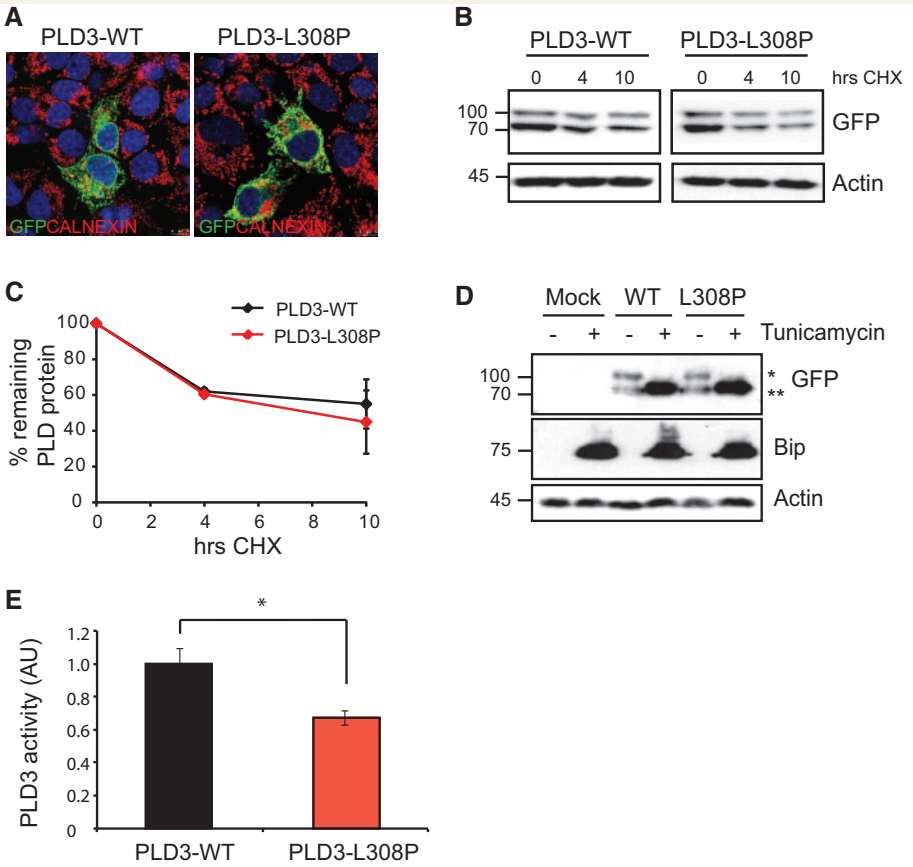

Figure 5 Missense mutation p.Leu308Pro leads to reduced phospholipase activity. (**A**) Confocal images of COS-7 cells transfected with turboGFP-PLD3-WT or turboGFP-PLD3-L308P (green) and stained with anti-Calnexin antibody (endoplasmatic reticulum marker; red) and DAPI (nucleus; blue). No differences in cellular localization were detected. Scale bar = 10 μm. (**B**) Representative western blot image showing the remaining PLD3 protein levels in COS-7 expressing turboGFP-PLD3-WT or turboGFP-PLD3-L308P following cycloheximide treatment (CHX; 25 μg/ml) for 0, 4, and 10 h. (**C**) Graph showing the quantification of the protein bands. No significant differences were observed [WT: 61.9 ± 1.6% (4 h) and 55.0 ± 13% (10 h) versus L308P: 60.4 ± 1.1% (4 h) and 44.8 ± 13% (10 h)]. Data are corrected for actin levels, normalized against time point 0 and shown as mean ± SD. (**D**) Representative western blot image showing protein levels in COS-7 cells transfected with turboGFP (mock control), turboGFP-PLD3-WT or turboGFP-PLD3-L308P untreated and treated with tunicamycin (1 μg/ml) for 24 h. No differences in PLD3 protein glycosylation were observed that was efficiently blocked by tunicamycin in both PLD3-WT and PLD3-L308P. Treatment with tunicamycin markedly increased BiP protein levels, but no differences were observed between cells expressing turboGFP, PLD3-WT, or PLD3-L308P. (**E**) Quantification of PLD3 activity in COS-7 cell lysates. Cells were transfected with turboGFP as a mock control, turboGFP-PLD3-WT, or turboGFP-PLD3-L308P and the activity was measured using a commercial available colorimetric assay. The relative PLD activity of cells transfected was: GFP 1.7 ± 0.1, PLD3-WT 2.2 ± 0.2, and PLD3-L308P 1.5 ± 0.06 (*$P < 0.05$). Data are normalized against the mock control and shown as mean ± SD.

Discussion

Using an approach combining WES, targeted resequencing, and gene network analysis in 20 families, we identified five novel disease genes for SCA (*FAT2*, *FAT1*, *PLD3*, *EP300*, and *KIF26B*) that explain 25% of this cohort. All mutations were predicted to be damaging to protein function. Furthermore, 20% of the families and nearly 10% of the replication cohort carried mutations in known but rarer SCA genes that are not part of routine screening in most genetic diagnostic laboratories. This illustrates the power of next-generation sequencing (NGS)-based diagnostic testing and the need to incorporate these rarer genes in standard diagnostics. In 60% of the families, we were unable to identify a single genetic defect using WES. This may be due to a number of factors: some parts of the exome are still not completely covered, mutations may be located in highly repetitive sequences, large insertions or deletions are missed by sequencing, mutations are non-coding, or insight into the functional consequences of variants is missing. This result demonstrates the need to improve current genetic diagnostics and the necessity of assessing the functional consequences of putative disease-causing variants. Our work clearly illustrates the complexity of studying the impact of missense variants on protein functioning in an autosomal dominant brain disease as this requires

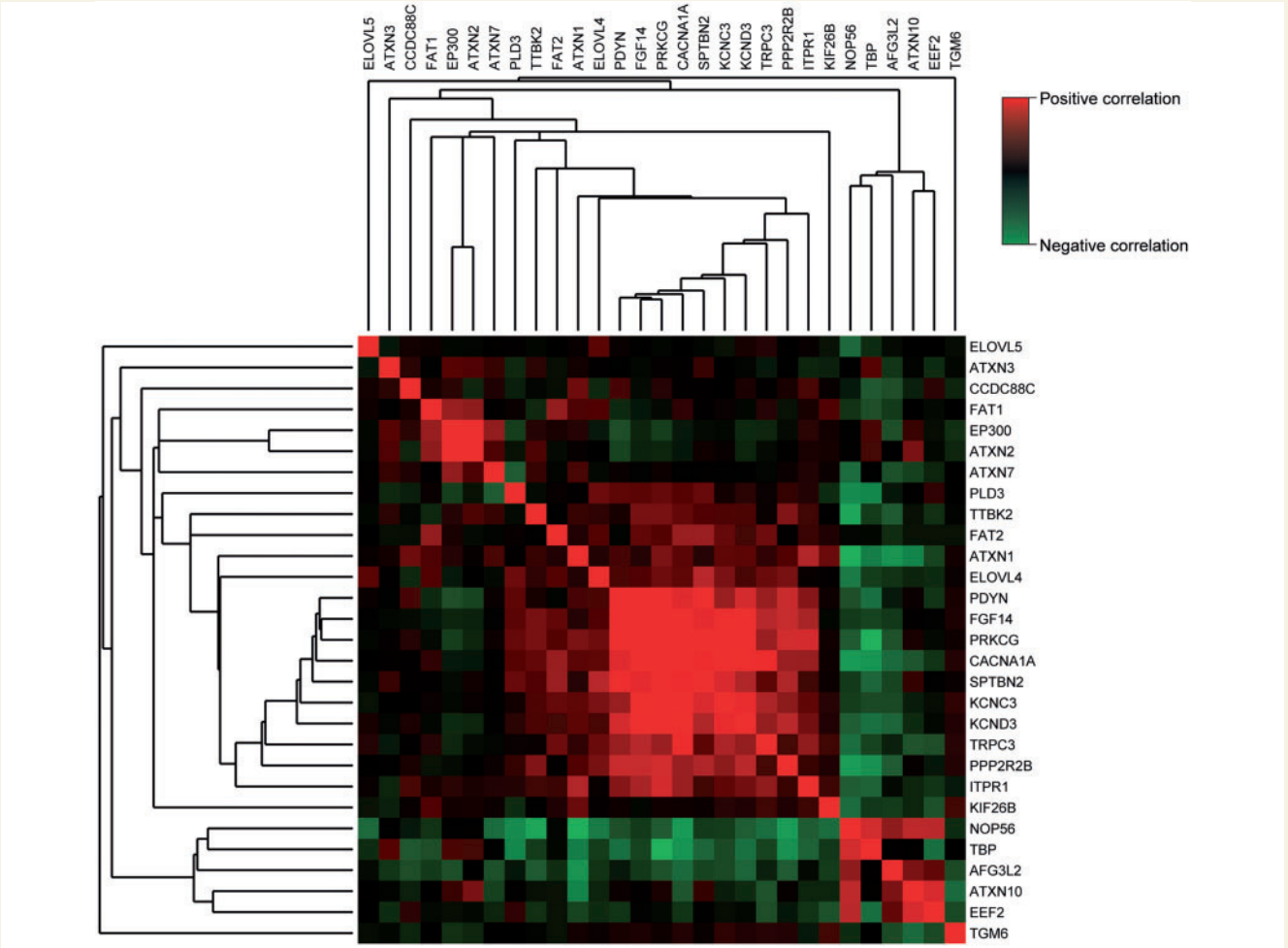

Figure 6 Known SCA genes and candidate genes highlight synaptic transmission and transcription regulation as shared mechanisms in the pathogenesis of SCA. Comparison of functionality between all known SCA genes and the five candidate genes that revealed the presence of two functional clusters that contain both the known SCA genes and candidate genes. The large cluster in the middle represents genes that co-function in synaptic transmission and includes the known SCA genes *PDYN*, *FGF14*, *PRKCG*, and *CACNA1A* and the candidate genes *FAT2*, *PLD3*, and *KIF26B*. The smaller *upper left* and *lower right* clusters are predicted to have a both nuclear function and role in transcription regulation. The *upper left* cluster contains known SCA genes such as *ATXN2* and *ATXN3* and candidate genes *FAT1* and *EP300*. The *lower right* cluster only contains known genes such as *NOP56*, *TBP*, and *EEF2* with functions in splicing, transcription, and translation.

functional studies that fit the profile of the candidate gene in overexpression cell models.

Since we are aware of the limitations of our functional studies, we used additional methods such as gene network analysis to support novel genetic findings. None of the novel candidate genes were previously linked to SCA, but our data showed that *FAT2*, *PLD3*, *FAT1*, and *EP300* are highly connected with known SCA genes based on their predicted functions. To date, *KIF26B* is linked to kidney agenesis in mice by controlling endothelial polarity and vascularization (Uchiyama *et al.*, 2010), but may play a yet-unrecognized role in sympathetic and autonomic CNS development and axon guidance as predicted by co-expression analysis. Moreover, our gene network analysis supports the idea that different SCA genes converge on a limited number of biological pathways and demonstrates the importance of synaptic transmission and transcription regulation in the pathophysiology of SCA. This is further strengthened by the notion that *fat1* and *fat2* are implicated in fatty acid synthesis in a zebrafish model (Pang *et al.*, 2014), just like *ELOVL4* and *ELOVL5*, the genes underlying SCA34 and SCA38, respectively (Cadieux-Dion *et al.*, 2014; Di Gregorio *et al.*, 2014). Loss of function mutations in protocadherin *FAT1* and truncating mutations in *EP300*, which encodes a histone acetyltransferase, have been implicated in human cancers (Gayther *et al.*, 2000; Morris *et al.*, 2013). Additionally, splicing mutations in *FAT1* cause a facioscapulohumeral dystrophy-like phenotype and dominant *de novo* mutations in *EP300* cause a rare form of Rubinstein-Taybi syndrome, a multiple congenital anomaly

syndrome characterized mainly by mental retardation and microcephaly (OMIM 613684) (Roelfsema *et al.*, 2005; Puppo *et al.*, 2015).

Despite the fact that both FAT1 and EP300 play a crucial role in transcription regulation, they seemingly also have a mitochondrial function. Notably, mitochondrial dysfunction has already been implicated in the pathogenesis of several known SCA types (Matilla-Dueñas *et al.*, 2014). FAT1 has been shown to control mitochondrial function by regulating complex I and II activity in smooth muscle cells (Cao *et al.*, 2016). In this study, FAT1 fragments were shown to accumulate in smooth muscle cell mitochondria and even directly interact with inner mitochondrial membrane proteins implicating a direct role of FAT1 in mitochondrial function. Additionally, EP300 was shown to modify key regulators of mitochondrial gene expression and to maintain mitochondrial integrity in the heart (Nakagawa *et al.*, 2009).

Interestingly, DNA replication stress due to altered transcription can be the cause of genome instability and subsequent double strand breaks in the DNA (Bertoli *et al.*, 2016). Notably, DNA damage plays a crucial role in the pathogenesis of several recessive ataxias including ataxia-telangiectasia (ATM), spinocerebellar ataxia with axonal neuropathy (SCAN1), ataxia-ocular motor apraxia (AOA1) and others (Jiang *et al.*, 2017). More recently, mutant α-synuclein was also shown to cause transcriptional deregulation and specifically affected the transcription of DNA repair genes underlying the neurotoxicity (Paiva *et al.*, 2017). These data may point towards a common role of transcriptional deregulation and accumulation of DNA damage in several neurodegenerative disorders.

Notably, three of five candidate genes (*FAT1*, *FAT2*, and *EP300*) are implicated in autophagy (Calamita and Fanto, 2011; Napoletano *et al.*, 2011; Mariño *et al.*, 2014). Autophagy plays an important role in protein turnover and recycling in the synapse and dendritic spine elimination (Shehata *et al.*, 2012; Khan *et al.*, 2014; Tang *et al.*, 2014), thereby mediating synaptic plasticity and neurotransmission. Alterations in autophagy have been shown to underlie neurodevelopmental disorders including autism (Dere *et al.*, 2014), but also may lead to neurodegeneration as pathological protein accumulations at synapse terminals or altered turnover of crucial receptors at the synapse may play a role in Parkinson's disease (Friedman *et al.*, 2012; Zhang *et al.*, 2015). Recently, the *ATG5* gene encoding autophagy-related 5 was implicated in a congenital recessive ataxia with developmental delay (Kim *et al.*, 2016), demonstrating that disrupted autophagy can lead to cerebellar ataxia. Additionally, accumulation of p62 caused defective ATM-mediated DNA repair and accumulation of protein-linked DNA breaks that might trigger C9orf72-related neurodegeneration (Walker *et al.*, 2017). Based on this work and the fact that defective DNA break repair caused by mutations in *TPB1* underlies SCAN1 (El-Khamisy *et al.*, 2005), we speculate that alterations in autophagy due to mutations in *FAT1*, *FAT2*, and *EP300* may affect DNA break repair leading to cerebellar neurodegeneration and ataxia. Notably, our gene network analysis based on co-functionality with known SCA genes did not identify autophagy or fatty acid synthesis as a shared mechanism for cerebellar ataxia, probably because the analysis does not use phenotype-specific hypotheses and considers multiple lines of complementary evidence to assess co-functionality.

EP300 was also shown to bind to ATXN3, the SCA3 protein, resulting in inhibition of histone acetylation and transcription (Li *et al.*, 2002), further validating its role in brain disorders including SCA. Finally, *PLD3* was recently implicated as a major risk gene for Alzheimer's disease (Cruchaga *et al.*, 2014) and, importantly, the p.Leu308Pro mutation was not observed in $\pm$2300 Alzheimer's disease cases and 2000 controls upon resequencing and was absent in any genetic database.

The gene networks (both on co-expression and co-functionality) implicate other candidate genes in SCA cases without a genetic diagnosis. For example, *CEND1*, was a top 25 gene from our co-expression network (for full gene list see Supplementary Table 10) and mice lacking CEND1 exhibit compromised cerebellar development and show deficits in motor coordination (Sergaki *et al.*, 2010), strengthening its putative role in the pathophysiology of SCA. Therefore, we encourage the screening of existing and new exome sequence datasets for damaging variants in these genes.

In conclusion, our work further validated an important role for synaptic transmission and transcription regulation as underlying shared biological pathways in SCA. Although more functional work needs to be performed to consolidate how mutations in these novel SCA genes cause cerebellar neurodegeneration, we provided more insights into the main shared molecular pathways leading to cerebellar neurodegeneration. Finally, we showed that this combined approach is highly successful in identifying mutations in genes in relatively small families and helps reveal the underlying biological pathways that lead to disease.

Web resources

1000 Genomes, http://browser.1000genomes.org
dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evesusgs.washington.edu/EVS/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
Exome Aggregation Consortium (ExAc), http://www.exac.broadinstitute.org

Acknowledgements

We would like to acknowledge all the patients and relatives that participated in this study, Kate Mc Intyre for editing of

the manuscript, and Mirjam Roffel for technical assistance. We would like to thank Prof. L.P ten Kate, Prof. K.P.J. Braun, Dr E. Brand for providing clinical information, and Prof. Nakayama, Dr Chan, Prof. Li, Dr. Cruchaga, and Dr Goedhart for sharing the DNA plasmids.

Funding

This work was funded by a Rosalind Franklin Fellowship awarded by the University of Groningen, the Prinses Beatrix Muscle Funds (W.OR10-38), and NutsOhra (1101-042). Part of the work was performed at the UMCG Imaging and Microscopy Center (UMIC), which is sponsored by NWO grants 40-00506-98-9021 and 175-010-2009-023.

Supplementary material

Supplementary material is available at *Brain* online.

References

1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature 2010; 467: 1061–73.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010; 7: 248–9.

Ahmed AF, de Bock CE, Lincz LF, Pundavela J, Zouikr I, Sontag E, et al. FAT1 cadherin acts upstream of Hippo signalling through TAZ to regulate neuronal differentiation. Cell Mol Life Sci 2015; 72: 4653–69.

Arany Z, Sellers WR, Livingston DM, Eckner R. E1A-associated p300 and CREB-associated CBP belong to a conserved family of coactivators. Cell 1994; 77: 799–800.

Bakalkin G, Watanabe H, Jezierska J, Depoorter C, Verschuuren-Bemelmans C, Bazov I, et al. Prodynorphin mutations cause the neurodegenerative disorder spinocerebellar ataxia type 23. Am J Hum Genet 2010; 87: 593–603.

Bertoli C, Herlihy AE, Pennycook BR, Kriston-Vizi J, de Bruin RAM. Sustained E2F-dependent transcription is a key mechanism to prevent replication-stress-induced DNA damage. Cell Rep 2016; 15: 1412–22.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 2014; 42: W252–8.

Cadieux-Dion M, Turcotte-Gauthier M, Noreau A, Martin C, Meloche C, Gravel M, et al. Expanding the clinical phenotype associated with ELOVL4 mutation: study of a large French-Canadian family with autosomal dominant spinocerebellar ataxia and erythrokeratodermia. JAMA Neurol 2014; 71: 470–5.

Cagnoli C, Stevanin G, Brussino A, Barberis M, Mancini C, Margolis RL, et al. Missense mutations in the AFG3L2 proteolytic domain account for ~1.5% of European autosomal dominant cerebellar ataxias. Hum Mutat 2010; 31: 1117–24.

Calamita P, Fanto M. Slimming down fat makes neuropathic hippo: the Fat/Hippo tumor suppressor pathway protects adult neurons through regulation of autophagy. Autophagy 2011; 7: 907–9.

Cao LL, Riascos-Bernal DF, Chinnasamy P, Dunaway CM, Hou R, Pujato MA, et al. Control of mitochondrial function and cell growth by the atypical cadherin Fat1. Nature 2016; 539: 575–8.

Chen D-H, Brkanac Z, Verlinde CLMJ, Tan X-J, Bylenok L, Nochlin D, et al. Missense mutations in the regulatory domain of PKC gamma: a new mechanism for dominant nonepisodic cerebellar ataxia. Am J Hum Genet 2003; 72: 839–49.

Coutelier M, Blesneac I, Monteil A, Monin M-L, Ando K, Mundwiller E, et al. A recurrent mutation in CACNA1G alters Cav3.1 T-type calcium-channel conduction and causes autosomal-dominant cerebellar ataxia. Am J Hum Genet 2015; 97: 726–37.

Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 2014; 505: 550–4.

Delplanque J, Devos D, Huin V, Genet A, Sand O, Moreau C, et al. TMEM240 mutations cause spinocerebellar ataxia 21 with mental retardation and severe cognitive impairment. Brain 2014; 137: 2657–63.

Depondt C, Donatello S, Rai M, Wang FC, Manto M, Simonis N, et al. MME mutation in dominant spinocerebellar ataxia with neuropathy (SCA43). Neurol Genet 2016; 2: e94.

Dere E, Dahm L, Lu D, Hammerschmidt K, Ju A, Tantra M, et al. Heterozygous ambra1 deficiency in mice: a genetic trait with autism-like behavior restricted to the female gender. Front Behav Neurosci 2014; 8: 181.

Di Bella D, Lazzaro F, Brusco A, Plumari M, Battaglia G, Pastore A, et al. Mutations in the mitochondrial protease gene AFG3L2 cause dominant hereditary ataxia SCA28. Nat Genet 2010; 42: 313–21.

Di Gregorio E, Borroni B, Giorgio E, Lacerenza D, Ferrero M, Buono Lo N, et al. ELOVL5 mutations cause spinocerebellar ataxia 38. Am J Hum Genet 2014; 95: 209–17.

Duarri A, Jezierska J, Fokkens M, Meijer M, Schelhaas HJ, Dunnen den WFA, et al. Mutations in potassium channel kcnd3 cause spinocerebellar ataxia type 19. Ann Neurol 2012; 72: 870–80.

Durr A. Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. Lancet Neurol 2010; 9: 885–94.

El-Khamisy SF, Saifi GM, Weinfeld M, Johansson F, Helleday T, Lupski JR, et al. Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1. Nature 2005; 434: 108–13.

Fehrmann RSN, Karjalainen JM, Krajewska M, Westra H-J, Maloney D, Simeonov A, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. Nat Genet 2015; 47: 115–25.

Fogel BL, Hanson SM, Becker EBE. Do mutations in the murine ataxia gene TRPC3 cause cerebellar ataxia in humans? Mov Disord 2015; 30: 284–6.

Friedman LG, Lachenmayer ML, Wang J, He L, Poulose SM, Komatsu M, et al. Disrupted autophagy leads to dopaminergic axon and dendrite degeneration and promotes presynaptic accumulation of α-synuclein and LRRK2 in the brain. J Neurosci 2012; 32: 7585–93.

Gayther SA, Batley SJ, Linger L, Bannister A, Thorpe K, Chin SF, et al. Mutations truncating the EP300 acetylase in human cancers. Nat Genet 2000; 24: 300–3.

Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet 2014; 46: 818–25.

Guan W-J, Wang J-L, Liu Y-T, Ma Y-T, Zhou Y, Jiang H, et al. Spinocerebellar ataxia type 35 (SCA35)-associated transglutaminase 6 mutants sensitize cells to apoptosis. Biochem Biophys Res Commun 2013; 430: 780–6.

Hekman KE, Yu G-Y, Brown CD, Zhu H, Du X, Gervin K, et al. A conserved eEF2 coding variant in SCA26 leads to loss of translational fidelity and increased susceptibility to proteostatic insult. Hum Mol Genet 2012; 21: 5472–83.

Hui X, Reither G, Kaestner L, Lipp P. Targeted activation of conventional and novel protein kinases C through differential translocation patterns. Mol Cell Biol 2014; 34: 2370–81.

Jen J, Yue Q, Nelson SF, Yu H, Litt M, Nutt J, et al. A novel nonsense mutation in CACNA1A causes episodic ataxia and hemiplegia. Neurology 1999; 53: 34–7.

Jiang B, Glover JNM, Weinfeld M. Neurological disorders associated with DNA strand-break processing enzymes. Mech Ageing Dev 2017; 161: 130–40.

Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX web server. Nat Protoc 2012; 7: 1511–22.

Khan MM, Strack S, Wild F, Hanashima A, Gasch A, Brohm K, et al. Role of autophagy, SQSTM1, SH3GLB1, and TRIM63 in the turnover of nicotinic acetylcholine receptors. Autophagy 2014; 10: 123–36.

Kim M, Sandford E, Gatica D, Qiu Y, Liu X, Zheng Y, et al. Mutation in ATG5 reduces autophagy and leads to ataxia with developmental delay. Elife 2016; 5: 213.

Klockgether T. Update on degenerative ataxias. Curr Opin Neurol 2011; 24: 339–45.

Kobayashi H, Abe K, Matsuura T, Ikeda Y, Hitomi T, Akechi Y, et al. Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. Am J Hum Genet 2011; 89: 121–30.

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009; 4: 1073–81.

Lee Y-C, Durr A, Majczenko K, Huang Y-H, Liu Y-C, Lien C-C, et al. Mutations in KCND3 cause spinocerebellar ataxia type 22. Ann Neurol 2012; 72: 859–69.

Li F, Macfarlan T, Pittman RN, Chakravarti D. Ataxin-3 is a histone-binding protein with two independent transcriptional corepressor activities. J Biol Chem 2002; 277: 45004–12.

MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. Nature 2014; 508: 469–76.

Magg T, Schreiner D, Solis GP, Bade EG, Hofer HW. Processing of the human protocadherin Fat1 and translocation of its cytoplasmic domain to the nucleus. Exp Cell Res 2005; 307: 100–8.

Mariño G, Pietrocola F, Eisenberg T, Kong Y, Malik SA, Andryushkova A, et al. Regulation of autophagy by cytosolic acetyl-coenzyme A. Mol Cell 2014; 53: 710–25.

Matilla-Dueñas A, Ashizawa T, Brice A, Magri S, McFarland KN, Pandolfo M, et al. Consensus paper: pathological mechanisms underlying neurodegeneration in spinocerebellar ataxias. Cerebellum 2014; 13: 269–302.

Matsui S, Utani A, Takahashi K, Mukoyama Y, Miyachi Y, Matsuyoshi N. Knockdown of Fat2 by siRNA inhibits the migration of human squamous carcinoma cells. J Dermatol Sci 2008; 51: 207–10.

Mezulis S, Sternberg MJE, Kelley LA. PhyreStorm: a web server for fast structural searches against the PDB. J Mol Biol 2016; 428: 702–8.

Moeller MJ, Soofi A, Braun GS, Li X, Watzl C, Kriz W, et al. Protocadherin FAT1 binds Ena/VASP proteins and is necessary for actin dynamics and cell polarization. EMBO J 2004; 23: 3769–79.

Morris LGT, Kaufman AM, Gong Y, Ramaswami D, Walsh LA, Turcan Ş, et al. Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. Nat Genet 2013; 45: 253–61.

Nakagawa Y, Kuwahara K, Takemura G, Akao M, Kato M, Arai Y, et al. p300 plays a critical role in maintaining cardiac mitochondrial function and cell survival in postnatal hearts. Circ Res 2009; 105: 746–54.

Napoletano F, Occhi S, Calamita P, Volpi V, Blanc E, Charroux B, et al. Polyglutamine Atrophin provokes neurodegeneration in Drosophila by repressing fat. EMBO J 2011; 30: 945–58.

Nolte D, Landendinger M, Schmitt E, Müller U. Spinocerebellar ataxia 14: novel mutation in exon 2 of PRKCG in a German family. Mov Disord 2007; 22: 265–7.

Oda H, Uemura T, Harada Y, Iwai Y, Takeichi M. A Drosophila homolog of cadherin associated with armadillo and essential for embryonic cell-cell adhesion. Dev Biol 1994; 165: 716–26.

Ophoff RA, Terwindt GM, Vergouwe MN, van Eijk R, Oefner PJ, Hoffman SM, et al. Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca2+ channel gene CACNL1A4. Cell 1996; 87: 543–52.

Paiva I, Pinho R, Pavlou MA, Hennion M, Wales P, Schütz A-L, et al. Sodium butyrate rescues dopaminergic cells from alpha-synuclein-induced transcriptional deregulation and DNA damage. Hum Mol Genet 2017; 26: 2231–46.

Pang S-C, Wang H-P, Li K-Y, Zhu Z-Y, Kang JX, Sun Y-H. Double transgenesis of humanized fat1 and fat2 genes promotes omega-3 polyunsaturated fatty acids synthesis in a zebrafish model. Mar Biotechnol 2014; 16: 580–93.

Pers TH, Karjalainen JM, Chan Y, Westra H-J, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun 2015; 6: 5890.

Puppo F, Dionnet E, Gaillard M-C, Gaildrat P, Castro C, Vovan C, et al. Identification of variants in the 4q35 gene FAT1 in patients with a facioscapulohumeral dystrophy-like phenotype. Hum Mutat 2015; 36: 443–53.

Roelfsema JH, White SJ, Ariyürek Y, Bartholdi D, Niedrist D, Papadia F, et al. Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. Am J Hum Genet 2005; 76: 572–80.

Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010; 7: 575–6.

Seixas AI, Loureiro JR, Costa C, Ordóñez-Ugalde A, Marcelino H, Oliveira CL, et al. A Pentanucleotide ATTTC repeat insertion in the non-coding region of DAB1, mapping to SCA37, causes spinocerebellar ataxia. Am J Hum Genet 2017; 101: 87–103.

Sergaki MC, Guillemot F, Matsas R. Impaired cerebellar development and deficits in motor coordination in mice lacking the neuronal protein BM88/Cend1. Mol Cell Neurosci 2010; 44: 15–29.

Shakkottai VG, do Carmo Costa M, Dell'Orco JM, Sankaranarayanan A, Wulff H, Paulson HL. Early changes in cerebellar physiology accompany motor dysfunction in the polyglutamine disease spinocerebellar ataxia type 3. J Neurosci 2011; 31: 13002–14.

Shehata M, Matsumura H, Okubo-Suzuki R, Ohkawa N, Inokuchi K. Neuronal stimulation induces autophagy in hippocampal neurons that is involved in AMPA receptor degradation after chemical long-term depression. J Neurosci 2012; 32: 10413–22.

Smeets CJLM, Jezierska J, Watanabe H, Duarri A, Fokkens MR, Meijer M, et al. Elevated mutant dynorphin A causes Purkinje cell loss and motor dysfunction in spinocerebellar ataxia type 23. Brain 2015; 138: 2537–52.

Smeets CJLM, Verbeek DS. Cerebellar ataxia and functional genomics: identifying the routes to cerebellar neurodegeneration. Biochim Biophys Acta 2014; 1842: 2030–8.

Tang G, Gudsnuk K, Kuo S-H, Cotrina ML, Rosoklija G, Sosunov A, et al. Loss of mTOR-dependent macroautophagy causes autistic-like synaptic pruning deficits. Neuron 2014; 83: 1131–43.

Tanoue T, Takeichi M. Mammalian Fat1 cadherin regulates actin dynamics and cell-cell contact. J Cell Biol 2004; 165: 517–28.

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 2012; 337: 64–9.

Terabayashi T, Sakaguchi M, Shinmyozu K, Ohshima T, Johjima A, Ogura T, et al. Phosphorylation of Kif26b promotes its polyubiquitination and subsequent proteasomal degradation during kidney development. PLoS One 2012; 7: e39714.

Tsoi H, Yu ACS, Chen ZS, Ng NKN, Chan AYY, Yuen LYP, et al. A novel missense mutation in CCDC88C activates the JNK pathway and causes a dominant form of spinocerebellar ataxia. J Med Genet 2014; 51: 590–5.

Uchiyama Y, Sakaguchi M, Terabayashi T, Inenaga T, Inoue S, Kobayashi C, et al. Kif26b, a kinesin family gene, regulates adhesion of the embryonic kidney mesenchyme. Proc Natl Acad Sci USA 2010; 107: 9240–5.

van Dijk GW, Wokke JH, Oey PL, Franssen H, Ippel PF, Veldman H. A new variant of sensory ataxic neuropathy with autosomal dominant inheritance. Brain 1995; 118 (Pt 6): 1557–63.

Verbeek DS, Goedhart J, Bruinsma L, Sinke RJ, Reits EA. PKC gamma mutations in spinocerebellar ataxia type 14 affect C1 domain accessibility and kinase activity leading to aberrant MAPK signaling. J Cell Sci 2008; 121: 2339–49.

Walker C, Herranz-Martin S, Karyka E, Liao C, Lewis K, Elsayed W, et al. C9orf72 expansion disrupts ATM-mediated chromosomal break repair. Nat Neurosci 2017; 45: 1159.

Wang J-L, Yang X, Xia K, Hu ZM, Weng L, Jin X, et al. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. Brain 2010; 133: 3510–18.

Wu J, Yan Z, Li Z, Qian X, Lu S, Dong M, et al. Structure of the voltage-gated calcium channel Ca(v)1.1 at 3.6 Å resolution. Nature 2016; 537: 191–6.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods 2015; 12: 7–8.

Zhang H, Duan C, Yang H. Defective autophagy in Parkinson's disease: lessons from genetics. Mol Neurobiol 2015; 51: 89–104.

Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, et al. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. Nat Genet 1997; 15: 62–9.