

Title: Mutations in *ANXA11* cause familial and sporadic amyotrophic lateral sclerosis

Authors: Bradley N Smith^{1†}, Simon D Topp^{1†}, Claudia Fallini^{2†}, Hideki Shibata^{3†}, Han-Jou Chen^{1†}, Claire Troakes¹, Andrew King¹, Nicola Ticozzi⁴⁺⁵, Kevin P. Kenna², Athina Soragia-Gkazi¹, Jack W Miller¹, Akane Sato³, Diana Marques Dias¹, Maryangel Jeon², Caroline Vance¹, Chun Hao Wong¹, Martina de Majo¹, Wejdan Kattuah¹, Jacqueline C Mitchell¹, Emma L Scotter⁶, Nicholas W Parkin⁷, Peter C. Sapp², Matthew Nolan¹, Peter J Nestor⁸, Michael Simpson⁹, Michael Weale⁹, Monkel Lek¹⁰⁺¹¹, Frank Baas¹², Vianney de Jong JM¹², Anneloor LMA ten Asbroek¹², Alberto Garcia Redondo¹³, Jesús Esteban-Pérez¹³, Cinzia Tiloca⁴⁺⁵, Federico Verde⁴⁺⁵, Stefano Duga¹⁴⁺¹⁵, Nigel Leigh¹⁶, Pall H¹⁷, Karen E. Morrison¹⁸, Ammar Al-Chalabi¹, Pamela J. Shaw¹⁹, Janine Kirby¹⁹, Martin R. Turner²⁰, Kevin Talbot²⁰, Orla Hardiman²¹, Jonathan D. Glass²², Jacqueline de Belleruche²³, Masatoshi Maki³, Stephen E Moss²⁴, Christopher Miller¹, Cinzia Gellera²⁵, Antonia Ratti⁴⁺⁵, Safa Al-Sarraj¹, Robert H Brown Jr², Vincenzo Silani^{4+5*}, John E Landers^{2*}, Christopher E Shaw^{1*}

†* These authors made equal contributions to the manuscript

Affiliations:

1. Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 125 Coldharbour Lane, Camberwell, SE5 9NU, London, UK.
2. Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.
3. Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan.
4. Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, 20149 Milan, Italy
5. Department of Pathophysiology and Transplantation, 'Dino Ferrari' Center, University of Milan, 20122 Milan Italy
6. The Centre for Brain Research, Faculty of Medical and Health Sciences, University of Auckland, 85 Park Road, Grafton, Auckland, New Zealand.
7. Viapath Molecular Genetics, Guy's Genetics Centre, Guy's Hospital, Great Maze Pond, London, SE1 9RT.
8. German Center for Neurodegenerative Diseases, Leipziger Str. 44, 39120 Magdeburg, Germany.

9. Medical & Molecular Genetics, Division of Genetics and Molecular Medicine, Guys Tower, Kings College London, London Bridge, SE1 9RT.
10. Analytic and Translation Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
11. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.
12. Department of Genome analysis, University of Amsterdam, Academic Medical Centre, PO Box 22700, 1100DE, Amsterdam, The Netherlands.
13. Unidad de ELA, Instituto de Investigación Hospital 12 de Octubre de Madrid, SERMAS, and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER U-723), Madrid, Spain.
14. Department of Biomedical Sciences, Humanitas University, Rozzano - Milan Italy.
15. Humanitas Clinical and Research Center, Via Manzoni 56, 20089 Rozzano - Milan, Italy.
16. Trafford Centre for Medical Research, Brighton & Sussex Medical School, BN1 9RY
17. School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, The University of Birmingham, Birmingham, UK.
18. School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, UK, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust UK.
19. Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK.
20. Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Oxford, UK.
21. Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin, Republic of Ireland.
22. Department of Neurology, Center for Neurodegenerative Disease, Emory University School of Medicine, Atlanta, Georgia 30322, USA.
23. Neurogenetics Group, Division of Brain Sciences, Imperial College London, Hammersmith Hospital Campus, Burlington Danes Building, Du Cane Road, London, W12 0NN.
24. Institute of Ophthalmology, University College London, 11-43 Bath Street, EC1V 9EL, London, UK.
25. Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', 20133 Milan, Italy.

Corresponding author: Professor Christopher Shaw, Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 125 Coldharbour Lane, Camberwell, SE5 9NU, London, UK. (T) +44 207 848 0974, email: chris.shaw@kcl.ac.uk

One sentence Summary: A stringent filtering strategy applied to 751 ALS exomes identified p.D40G and other mutations in *ANXA11* that produced cytoplasmic aggregates in patient spinal cord and *in-vitro* and disrupted calyculin binding.

Abstract:

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disorder. We screened 751 familial ALS exomes and identified 6 novel, or extremely rare, *ANXA11* mutations in 13 individuals. A novel p.D40G mutation was absent from 70,000 control exomes; segregated with disease in two kindreds, was present in another two unrelated cases ($p=0.0102$), and all carriers shared a common founder haplotype. Annexin A11-positive aggregates were abundant in spinal motor neurons and hippocampal axons in a p.D40G patient. Transfected cells expressing p.D40G and other N-terminal mutations had altered calyculin binding while a p.R235Q mutant formed insoluble aggregates. *ANXA11* mutations accounted for ~1% of familial and ~1.7% of apparent sporadic cases in our cohort. Mutations in *ANXA11* highlight the role of defective cellular trafficking in the pathogenesis of ALS.

Introduction

Gene hunting in rare Mendelian disorders has been transformed by exome sequencing. This approach is particularly attractive for late-onset autosomal dominant syndromes with short disease durations, such as Amyotrophic Lateral Sclerosis (ALS) in which DNA from multiple affected individuals in the same kindred are rare. ALS has a life-time risk of 1 in 400 and is characterized by degeneration of brain and spinal cord motor neurons resulting in progressive paralysis and death within an average of 3 years [1]. 10% of cases are familial (FALS) and a causative gene mutation can be identified in ~60% of European kindreds [2]. Mutations in the same genes account for ~10% of sporadic ALS cases (SALS) reflecting incomplete penetrance. Non-synonymous mutations in *SOD1*, *TARDBP*, *FUS* and an intronic hexanucleotide repeat expansion in *C9orf72* together account for ~20% of ALS cases and other genes for ~1-3% [3, 4]. Whole Exome and/or Genome Sequencing (WES/WGS) has identified nine ALS genes by either shared variant segregation analysis in ALS kindreds (*VCP*, *PFN1*, *MATR3*, *CHCHD10*, *CCNF*) or rare variant burden analysis (*TUBA4A*, *TBK1*, *NEK1*, *C21orf2*) [5-15]. Here we analysed FALS exomes and identified a novel non-synonymous founder mutation in *ANXA11* that is present in all affected family members tested and is also found in multiple unrelated index cases. Annexin A11 is a widely expressed, calcium-dependent phospholipid-binding protein (505AA, 56kDa) that belongs to the larger human Annexin family of

12 members [16]. Each family member has four highly conserved annexin domains, many of which can complex a calcium ion facilitating binding to anionic cell membranes. Unique to the annexin family, Annexin A11 has the longest N-terminus (~196 amino acids) that is hydrophobic, disordered and binds several interacting partners: the most characterized being calcyclin (encoded by *S100A6*) [17]. In regard to involvement with disease, Annexin A11 is associated with autoimmune disorders such as systemic lupus erythematosus and case-control studies have found a genetic association of the R230C single nucleotide polymorphism (SNP) with sarcoidosis [18, 19]. Additionally, enhanced Annexin A11 expression is also found in breast cancer and other acquired malignancies [18]. Here we present a new role for Annexin A11 in a rare, neurodegenerative Mendelian disorder as evidenced by novel mutations that cause Amyotrophic Lateral Sclerosis (ALS).

Results

(i) Exome sequencing, filtering and gene identification

From our cohort of 751 European FALS exomes (negative for *C9orf72* GGGGCC expansions) we obtained exome data for two or more affected relatives from only 50 families (average=2.14 individuals per family), highlighting the difficulty in accessing DNA from extended kindreds for this late onset disorder. Families ranged from a simple pair of siblings (sharing an estimated 50% of their variants) to an index case, parent and a second cousin (sharing an estimated 1.5% of their variants). On average 84% (range=58-97%) of the protein coding bases contained in Refseq transcripts were sequenced in all family members to a depth of ≥ 10 reads, a cut-off threshold in line with that used by the Exome Aggregation Consortium (ExAC) for defining high quality variants [20]. Our filtering strategy was to detect novel, high quality, coding and splicing variants that are absent from 1000genomes, UK10K, EVS and ExAC ($n > 72,000$). This produced an average of ~10 candidates per family (range=0-27) (Supplementary Table 1). As proof of principle the analysis identified mutations shared in single kindreds from several known ALS genes, including *SOD1*, *TARDBP*, *FUS*, *DCTN1* and *TUBA4A* [9, 21-24]. It was immediately apparent that only two variants appeared in the list of candidates for more than one family; the well characterized pathogenic M337V mutation in *TARDBP* [25] was found in two North American families, and a novel p.D40G variant in *ANXA11* (Refseq NM_145869) was found in two British families (an uncle/niece pair, and two cousins). The full list of candidate variants found in both UK *ANXA11* p.D40G families are listed in Supplementary Table 2. One additional Italian proband from the extended exome FALS cohort also carried the same *ANXA11* D40G variant.

We then extended the analysis to include 694 unrelated European FALS probands (including only the 50 probands from our multiplex families) and sought novel protein-changing variants that were shared by three or more probands. This approach also identified *ANXA11* D40G, and the following well-characterised pathogenic ALS mutations: 10x *SOD1* I114T [26], 6x *TARDBP* A382T [22], 5x *SOD1* A5V [27] 4x *FUS* R521C [23], 3x *TARDBP* M337V, 3x *SOD1* G94D [28], 3x *FUS* P525L [29].

We then Sanger sequenced the coding exons of *ANXA11* in a separate set of 180 British apparent SALS cases and identified one further heterozygous p.D40G carrier, bringing the total number to four out of a combined cohort of 874 probands.

As the two p.D40G UK families were underpowered to conduct a linkage analysis (Simulated LOD score of 0.63 using MERLIN [30]) we sought an alternate method to ascertain the significance of the p.D40G mutation. It is possible that a variant could be present 4 times in a sample of 874 cases and absent from 72,000 other individuals yet still be unrelated to disease. We therefore tested the null hypothesis: that any equal sized cohort of Europeans could also contain a novel variant shared by at least 4 people. We achieved this by running simulation studies using the aggregated variant call counts from the Non-Finnish European (NFE) subset of the ExAC database (n=33,370). In brief, all ExAC NFE variants were randomly distributed across 33,370 simulated individuals and random cohorts of 874 people extracted. The cohort was deemed to have 'passed' if it contained at least one protein-changing variant found 4 or more times within the cohort, but absent from the remainder of ExAC, UK10K, EVS, and 1000genomes. After 250,000 iterations, only 2,550 simulated cohorts contained such a variant, which demonstrates that the presence of the D40G variant is statistically significant (p=0.0102). Although not population-matched to our cohort, we consider ExAC to be suitable for this purpose as it contains a high proportion of Swedes who are on average more genetically homogeneous than UK individuals (Supplementary Table 3) [31]. Therefore ExAC would be expected to produce more shared non-pathogenic variants than our cohort, and so we expect this estimate of significance to be a conservative one. In conjunction with the family-based study the simulation analysis provides additional evidence that the *ANXA11* D40G mutation is associated with ALS.

(ii) The D40G mutation has a common European founder

Sanger sequencing of DNA from 17 family members across the two multi-generational British kindreds confirmed the presence of the p.D40G mutation in all four affected individuals identified from the exome capture data (Figs 1A-B). Four unaffected individuals also carried the mutation, but incomplete penetrance of ALS

mutations is well recognized and three of these were in their forties, whilst the average age of disease onset in *ANXA11* p.D40G affected carriers is 72 years of age. DNA was available from five British p.D40G carriers who share a common haplotype on the disease allele defined by 4 exonic SNPs and two polymorphic microsatellites spanning the locus with phase determined by a cluster of carriers in UK Family 2, confirming a common founder (Supplementary Figure 1) (primers Supplementary Table 4). The minimal haplotype is defined by a physical stretch of 2.5MB of genomic DNA spanning the *ANXA11* locus, common to all p.D40G carriers. The core 4 SNP haplotype, located in exons of genes flanking *ANXA11* is present in ~5% of our extended FALS cohort (n=787, including *C9orf72* expansion positive cases) and ~5% of Europeans from 1000genomes (n=514). This suggests that the mutation arose on a European background. The maximal recombination region defining the limits of the p.D40G locus is 7.1MB and contains 23 genes (Supplementary Figure 2). Interrogation of exome sequencing data found that no p.D40G carriers shared any additional protein-changing variants in the 23 genes within this region. 70.3–97.2% of coding bases were covered to a read depth ≥ 10 for each sample, and 97.9% of bases were covered by ≥ 10 reads in at least one sample (Supplementary Table 5). This evidence indicates that p.D40G is the sole causal variant within this locus.

(iii) ANXA11 mutations in FALS and SALS cluster in the N-terminus

Additional *ANXA11* variants identified in the FALS cohort were two unrelated individuals carrying a c.112G>A, p.G38R variant (rs81930608) that was absent from local UK and Italian exome controls (n=4505). Although the p.G38R variant is present in 5/31804 NFE ExAC exomes it is still significantly associated with FALS in our cohort (p=0.004, 2-tailed Fishers exact test). Novel p.G175R (c.523G>A) and p.R346C (c.1036C>T) variants were also detected in two index cases. The p.G175R was also present in an affected sibling with ALS (Fig 1C) confirming segregation with disease. All FALS cases possessing *ANXA11* variants were negative for exonic coding mutations in all known ALS genes including *SOD1*, *TARDBP*, *FUS*, *PFN1*, *UBQLN2*, *MATR3*, *CHCHD10*, *TBK1*, *OPTN*, *VAPB*, *ANG*, *SQSTM1*, *CCNF*, *TUBA4A*, *C21orf2*, *NEK1* and *VCP*.

All coding exons of *ANXA11* were Sanger sequenced in an additional 180 British SALS cases (primers listed in Supplementary Table 4). As previously mentioned, an additional D40G case was identified as well as an individual carrying a p.G189E (c.566G>A) variant, present once in local controls (n=4505) and 14 times in the ExAC NFE sub-population (n=33,140). Interestingly p.G189E is also present once in

ALSdB (<http://chgv.org/alsdb/index.jsp>). The last variant, located in the first annexin domain, was a p.R235Q (c.704G>A) change absent in local controls (n=4505) but present once as a low quality variant call in ExAC. Of note, ALS cases carrying *ANXA11* mutations did not carry the R230C risk allele (rs1049550) that is associated with Sarcoidosis [19]. All variants found in FALS and SALS cases are listed in Supplementary Table 6 with control database frequencies and various computational predictions of pathogenicity.

Four of the six mutations in Annexin A11 cluster with p.G38R and p.D40G within the long N-Terminus implying that this region has functional importance. All mutation positions are completely conserved in mammals and those in the annexin domains are also conserved in all currently sequenced birds, amphibians and reptiles (Fig 1D).

(iv) ANXA11 D40G cases have late disease onset and cytoplasmic immuno-reactive inclusions in post mortem tissue

All patients with *ANXA11* mutations had late disease onset (average 67 yrs) with a classical ALS phenotype without features of dementia (Supplementary Table 7). Five of six patients carrying the p.D40G variant had bulbar onset. Post mortem tissue was available from a SALS case carrying the p.D40G variant showing classical pathological features of ALS with a unique feature of large Annexin A11 immunoreactive inclusions that were absent from other ALS cases or controls (Fig 2). Spinal cord sections revealed marked neuronal loss within the anterior horns. There was marked myelin pallor and astrogliosis in the anterior and lateral corticospinal tracts. Many surviving motor neurons contained cytoplasmic inclusions positive for p62 and phospho-TDP-43 (Fig 2A). Phospho-TDP-43 inclusions were also found in the medulla, temporal neocortex and hippocampus. Annexin A11 staining of spinal cord sections showed numerous neuronal cytoplasmic inclusions (NCIs) in the perikaryon and proximal axon (Figs 2B-E). These included skein-like (Fig 2B), large-caliber tubular-shaped structures (Figs 2C-D) as well as filamentous and more complex basket-like inclusions (Fig 2E). Annexin A11 positive NCIs were also evident in small numbers in the motor cortex, dentate gyrus of the hippocampus and temporal neocortex sometimes accompanied by abundant torpedo-like neuritic structures in the neuropil (Fig 2F). Occasional, sparse Annexin A11 positive NCIs and neurites were also seen in the occipital lobe but no staining was observed in cerebellum. Staining for the amyloid precursor protein, Neurofilament (light, medium and heavy) and β 3 tubulin was negative. SALS (n=13) devoid of known mutations

were negative for Annexin A11 NCIs in spinal cord (Figure 2G), as were 2 *C9orf72* expansion positive cases and a *SOD1* positive case with a p.D101G mutation (Figs 2H-J). Additionally, tissue from 3 ALS/FTD cases (Fig 2K), 3 Alzheimer's (Fig 2L) and 3 Parkinson's were negative for Annexin A11 positive NCIs in the cortex (n=2) and spinal cord (n=1) for ALS/FTD and in cortical and brainstem tissue for AD and Parkinson's respectively. Spinal cord Annexin A11 staining was negative in 13 controls (Figure 2M) (including frontal and temporal cortex and medulla regions in 3 controls). Double fluorescent labelling of phospho-TDP-43 and annexin A11 in spinal cord neurons showed no evidence of co-localization (Fig 2N and O). Some Annexin A11 positive aggregates were ubiquitinated (Fig P). The cytoplasmic aggregation seen in post-mortem tissue from the D40G patient may be due to enhanced Annexin A11 expression due to an additional promoter mutation. However, screening of the *ANXA11* promoter identified no novel, rare or common variants (MAF<0.05) in any of the cases carrying *ANXA11* mutations. Furthermore, protein expression of lysates made from frontal cortex tissue of the SALS case possessing the p.D40G change did not show any changes in expression due to the mutant D40G allele compared to 3 age and sex matched controls (Supplementary Figure 3).

(v) Over-expressed Annexin-A11^{R235Q} forms foci in mouse Primary Motor Neurons that are insoluble

In order to understand the functional significance of these variants we undertook a series of cellular studies of mutations identified in our exome set. We generated Annexin-A11^{WT}, Annexin-A11^{G38R}, Annexin-A11^{D40G} and Annexin-A11^{R235Q} tagged with hemagglutinin (HA) or GFP at the C-terminus. Constructs were expressed in mouse primary motor neurons (PMNs) and human embryonic kidney 293 (HEK) cells. PMNs transfected with Annexin A11-HA^{WT} constructs showed nuclear and cytoplasmic localization. In the cytoplasm Annexin A11 was present in larger vesicle-like structures, smaller foci structures and diffusely distributed throughout the soma, axons and dendrites (Fig 3A). Apart from Annexin A11-HA^{R235Q}, which was elevated in the cytoplasm, there was no evidence that the other mutations affected the distribution between the nucleus and cytoplasm (Fig 3B) [32]. The Annexin A11-HA^{R235Q} mutant appeared to aggregate into foci, was never associated with vesicle-like structures and diffuse staining was absent (p<0.001) (Fig 3C). Annexin A11-HA^{D40G} had a similar pattern to WT but Annexin A11-HA^{G38R} displayed significantly less association with the vesicle-like structures (p=0.005). Annexin A11-GFP^{R235Q} in HEK cells showed marked aggregation into ubiquitin and p62 positive inclusions that were also specific for endogenous Annexin A11 antibody and confirmed by western

blot (Supplementary Figure 4). Annexin A11-GFP^{R235Q} also produced high molecular weight species on western blot that were insoluble (Figs 3D-E) (p=0.007).

(vi) Annexin-A11^{R235Q} sequesters wild-type Annexin-A11

We also investigated whether the aggregation-prone R235Q mutation recruits WT Annexin A11. We did this by co-transfecting SH-SY5Y cells with either Annexin A11-GFP^{WT} and Annexin A11-HA^{WT} or Annexin A11-GFP^{R235Q} and Annexin A11-HA^{WT} constructs (Figure 4a). The merge of Annexin A11-HA^{WT} and Annexin A11-GFP^{WT} showed nuclear and diffuse cytoplasmic staining (top panel), whereas Annexin A11-GFP^{R235Q} cytoplasmic puncta and WT-HA showed co-localization (bottom panel). This suggests that insoluble, mutant Annexin A11 recruits wild-type Annexin A11. This was confirmed by firstly conducting a solubility assay on HEK cells co-transfected with Annexin A11-HA^{WT} and either Annexin A11-GFP^{WT} or Annexin A11-GFP^{R235Q}, and secondly by immunoprecipitation (IP) and analyzed by Western blot. For the solubility assay, only the insoluble urea fraction of the Annexin A11-HA^{WT} / Annexin A11-GFP^{R235Q} co-transfection contained Annexin A11-HA^{WT} compared to the WT-Annexin A11^{HA/GFP} co-transfection (Figure 4b), demonstrating that the insoluble R235Q mutant can sequester wild-type Annexin A11. Furthermore, co-transfection of HEK cells with Annexin A11-HA^{WT} and Annexin A11-GFP^{R235Q}, and pulldown with rabbit-GFP antibody and blotting with mouse-HA, demonstrated co-precipitation of Annexin A11-HA^{WT} with clear banding of higher molecular weight insoluble species Figure 4c). This provides evidence that mutant Annexin A11 species recruits WT Annexin A11 and thus may act in a dominant negative manner.

(vii) Annexin A11 mutants abolish binding to Calcyclin

Residues 50-62 in Annexin A11 bind calcyclin (encoded by *S100A6*) [33] and are in close proximity to the p.G38R and p.D40G mutations, therefore we sought evidence for any effect on their interaction. Calcyclin is a 10kDa protein that contains 2 EF-hand calcium-binding motifs [34]. Jpred4 secondary structure predictions [35] on an alignment of all mammalian Annexin A11 orthologue sequences predicted that residues 40-44 and 51-59 form dual amphipathic helices (Supplementary Figure 5A). These helices mirror those seen in Annexin A1 which in the absence of calcium are embedded within the annexin core but are released on calcium activation, facilitating its binding to S100A11 and also mediating calcium-dependent protein-membrane interactions [36, 37] (Supplementary Figure 5B-C). Whilst the N-terminus of Annexin 11 is much longer than, and highly divergent from the N-terminus of Annexin 1, the high conservation of the first 4 residues in the second helix (QEYV in

Annexin 1 versus QDYL in Annexin 11) imply a possible conserved binding site in the annexin core and is additional evidence for a similar mechanism of action. Both the Annexin A11 p.G38R and p.D40G variants are predicted to prevent or severely reduce formation of the first amphipathic helix (Supplementary Fig 5D). Extending this analysis to 22 rare and common polymorphisms listed in ExAC, spanning the D40 residue, predicted no abolition of the first or second N-terminal amphipathic helices (Supplementary Figure 6) suggesting that this disruption of structure is disease specific.

In order to assess the effects of mutations in Annexin A11 on calcyclin binding we performed binding assays of Annexin A11-GFP^{WT} and the Annexin A11^{G38R/D40G/G189E/R235Q} mutants using a previously published *in-vitro* IP assay [38] (also methods). As anticipated, Annexin A11-GFP^{WT} co-immunoprecipitated with calcyclin, confirming calcyclin–Annexin A11 binding. Annexin A11-GFP^{D40G} severely disrupted calcyclin binding, as did Annexin A11-GFP^{G189E} and Annexin A11-GFP^{R235Q} mutants (Figure 5A). Annexin A11-GFP^{G38R} binding however was markedly increased compared to WT. To assess the specificity of the loss of calcyclin binding due to ANXA11 mutants, we conducted binding assays with four ANXA11 polymorphisms present in ExAC. These were a rare p.D40H variant (once in ExAC, rs368751524) not detected in our ALS cases and 2 common NFE N-terminal polymorphisms flanking p.D40G ie. p.P8L (rs147334030, MAF 0.011) and p.R191Q (rs2229554, MAF 0.07) and the p.R230C SNP associated with Sarcoidosis (rs1049550, MAF 0.44). These polymorphisms behaved in the same manner as wild-type and did not abolish calcyclin binding (Supplementary Figure 7) when compared to the p.D40G and p.R235Q mutations. Therefore, disruption to calcyclin binding in ALS cases harbouring ANXA11 mutations is disease specific. No differences were observed in Annexin A11 binding to other known binding partners ALG2 and Sorcin (Supplementary Figure 8).

(viii) Calcyclin is over-expressed in astrocytes of the corticospinal tract in ALS patients

Spinal cord sections from the ALS patient carrying the p.D40G mutation were stained with anti-calcyclin antibody. Although there was no observable difference in neuronal staining between the ALS case and controls, very high levels of calcyclin expression were detected in the cytoplasm of astrocytes in the lateral corticospinal tracts compared to controls (Figure 5Bi-ii). This is not a mutation-specific event as elevated calcyclin expression was also seen in apparent sporadic ALS cases without known

ALS gene mutations (Figure 5Biii). Increased expression of calcyclin in astrocytes of the corticospinal tract has previously been reported in sporadic ALS cases as well as SOD1^{G93A} transgenic mice, but the functional significance of this observation is unknown [39, 40].

(ix) Calcyclin co-overexpression in HEK cells clears R235Q aggregates

We also explored the impact of calcyclin interaction on Annexin A11 solubility. FLAG-Calcyclin^{WT} was co-transfected with Annexin A11-GFP^{WT} or Annexin A11-GFP^{R235Q} in HEK cells and assessed by NP40 insolubility assay and western blot. Remarkably, the increase in calcyclin expression reduced levels of soluble WT and p.R235Q but it completely cleared insoluble Annexin A11-GFP^{R235Q} (Figure 5C). This suggests that increasing calcyclin may inhibit Annexin A11 expression or enhance Annexin A11 clearance via a degradation pathway. Treating the cells with the ubiquitin proteasome inhibitor MG132 treatment for 24hrs significantly restored Annexin A11-GFP^{R235Q} insolubility (Figure 5D, $p < 0.001$).

Discussion

Using a stringent filtering strategy on a large cohort of FALS exomes (n=694) in conjunction with a shared variant analysis of 50 Caucasian kindreds we identified a non-synonymous mutation p.D40G in *ANXA11*, shared in two UK families but absent from >70,000 control exomes. The mutation segregated with disease in the two UK families and an additional Italian FALS proband also carried the p.D40G variant. Sequencing of 180 unrelated UK SALS identified a fourth p.D40G individual, providing further evidence that p.D40G is causative for ALS. An independent simulation strategy designed to ascertain the significance of mutations identified multiple times in our ALS cohort (n=874) found the D40G mutation (n=4) to be associated with ALS, $p = 0.0102$. All p.D40G individuals had a common founder p.D40G haplotype, which adds to the growing list of published founder mutations that are abundant in ALS [41-43]. In addition a novel p.G175R mutation segregated with disease in one kindred and several other *ANXA11* mutations clustered in the N-terminus of the molecule were also associated with ALS.

Our study was limited by several factors. The primary shared variant analysis that formed the basis of the study relied on the pathogenic variant for each family being novel, nonsense, mis-sense or splicing, and located in a region sequenced to an adequate depth in all family members. The lack of power when conducting linkage in the D40G kindreds necessitated us devising a new simulation-based approach to the

determination of significance, applied to the case of novel variants found multiple times in a disease cohort. Furthermore the high underlying heterogeneity of ALS, highlighted by discovering one novel gene from 50 families, is an on-going challenge in the identification of new causative genes.

Mutations in *ANXA11* present a distinct clinical phenotype characterized by late-onset classical ALS with 5 of 6 p.D40G cases having bulbar onset disease. Post mortem tissue from a p.D40G ALS case displayed abundant Annexin A11-positive aggregates within motor neurons in spinal cord and neurons and neuropil in the neocortex and hippocampus in addition to the classical features of neuronal loss, phospho-TDP-43 inclusions and astrogliosis.

Annexin A11 is a phospholipid binding protein, which may participate in vesicle trafficking as it has been reported to play a role in apoptosis, exocytosis and cytokinesis and forms phospholipid vesicles *in-vitro* in a calcium dependent manner [18, 44]. Annexin A11 co-localises with calyculin in the nuclear envelope during prophase of mitosis [45], however little is known about its role in post-mitotic neurons and its trafficking within the cell. From our study Annexin A11 does form vesicle-like structures in PMNs (Figure 3), which suggests a role for Annexin A11 in vesicular transport. Recent evidence has found that Annexin A11 plays an important structural role in regulating the delivery of vesicular cargoes from the endoplasmic reticulum to the Golgi apparatus which is mediated through its binding to ALG2 to stabilize Sec31A [46]. In our study, Annexin A11 mutations did not directly affect ALG2 binding and it is therefore more likely that the mutant alleles disrupt vesicular trafficking by altered binding to calyculin.

Annexin A11 mutations may behave in a dominant negative manner and interfere with the normal function of Annexin A11 in neurons, which is currently unknown. The sequestering of wild-type Annexin A11 by the R235Q mutation, located in the first annexin domain, offers this plausible mechanism (Figure 4). However most *ANXA11* mutations are N-terminal and do not aggregate or form insoluble aggregates in our *in-vitro* functional studies. In spite of this, the N-terminal p.D40G mutation under endogenous, long-lived conditions can aggregate as evidenced by abundant immuno-reactive inclusions seen in post mortem tissue from the p.D40G SALS case (Figure 2). Although the R235Q mutant failed to bind to calyculin, over-expression of calyculin completely cleared insoluble p.R235Q aggregates by facilitating proteasomal degradation (Figure 5). Calyculin is known to play a role in proteostasis

when it forms a functional complex with calyculin binding protein (CACYBP), and the RING type E3 Ubiquitin Ligase SIAH-1 that regulates ubiquitination and degradation of beta-catenin [47]. More subtle differences in calyculin binding were also observed for other Annexin A11 mutants. Calyculin binding was inhibited by p.D40G and p.G189E mutations and consistently increased above wild-type for the p.G38R mutant. Therefore loss of calyculin binding may result in an accumulation of cytoplasmic Annexin A11 promoting its aggregation. As most N-terminal mutations abolish Calyculin binding, this suggests Calyculin may also be necessary for Annexin A11 structure and/or function and N-terminal mutations may therefore act by a loss of function mechanism. As the p.R235Q mutation sequesters WT Annexin A11, it would also lead to aborted calyculin binding in the same manner as the p.D40G mutation, therefore a possible common outcome would be a loss of function.

We and others have shown that the levels of calyculin are greatly increased in astrocytes in apparent sporadic ALS cases as well as the one carrying the p.D40G mutation [39, 40]. Even though this phenomenon is not solely linked to *ANXA11* mutations, it raises the possibility that increased astrocytic calyculin expression is a response to defective proteostasis. Interestingly, cells in which TDP43 is knocked-down demonstrate a significant increase in the expression of calyculin transcripts [48, 49]. Additionally, elevated calyculin concentrations can form oligomers that have amyloidogenic properties that have been shown to seed SOD1 aggregation [50].

In summary, we have identified 6 novel, or extremely rare, *ANXA11* variants in 13 individuals, which account for ~1% of familial and 1.7% of apparent sporadic ALS cases in our cohort. Functional investigation of four of these variants (p.G38R, p.D40G, p.G189E and p.R235Q) has yielded initial insights into a causative role in ALS biology. However, the contribution of the remaining *ANXA11* variants to disease, if any, is yet to be determined. *ANXA11* variants alter binding to calyculin, which can lead to cytoplasmic aggregation in transfected cells and neuronal cytoplasmic inclusions in a patient carrying the D40G mutation at post mortem. The identification of *ANXA11* variants presents a new field of ALS investigation and highlights the significance of calcium binding proteins and intracellular trafficking in ALS pathobiology [51]. Further work investigating how mutations affect the folding of Annexin A11, its binding to calyculin and the effect on vesicular transport and TDP-43 aggregation will clarify the underlying disease mechanisms.

Materials and Methods

DNA Samples

Full patient consent was provided by FALS and SALS index patients and control individuals for research purposes. All patients had a diagnosis of definite or probable ALS based on revised El Escorial criteria [52] with at least one relative known to have ALS+/-FTD. All FALS were pre-screened for the *C9orf72* intronic hexanucleotide expansion, and any positive samples excluded from this analysis. The discovery cohort was comprised of 694 FALS probands sourced from the United States (266), UK (193), Italy (138), Spain (33), Germany (25), Ireland (17), Canada (9), the Netherlands (9), Belgium (3) and New Zealand (1). An additional 60 affected relatives of these probands, representing 50 families, were also exome sequenced in tandem for segregation analysis. Sequence data for 102 of the cases in the cohort were obtained, with permission, from the dbGAP repository (NIH Exome Sequencing of Familial Amyotrophic Lateral Sclerosis, NINDS, phs000101.v4.p1, Traynor).

Exome Sequencing and Bio-informatic Analysis

Paired-end FASTQ files from Illumina exome sequencing were aligned to the hg19 human reference using NovoCraft Novoalign and variants called with samtools v1.1 mpileup. VCF files were filtered at DP \geq 10, GQ \geq 20, MQ \geq 50, and indels normalized with bcftools v1.1 norm. Functional annotation, pathogenicity predictions and matches to 1000 genomes were added with table_annovar.pl [53], whilst all other annotation was added via custom perl scripts. Variants were filtered out if they were not novel (defined as being present in either 1000 genomes (<http://www.1000genomes.org>), ExAC (<http://exac.broadinstitute.org/>), ESP (<http://evs.gs.washington.edu/EVS/>), UK10K (<http://www.uk10k.org/>) or 672 In-house exome controls), or if they were located at positions with a read depth <10 in >25% of ExAC samples. Insertions and deletions were excluded due to the increased error rate in matching variants to control databases assembled and called by alternative pipelines. Synonymous and intronic variants were assessed by Netgene2 and Genesplicer, and excluded if no changes in scores compared to the reference allele were observed at locations matching to known Refseq acceptor or donor splice sites. Common ancestry between samples was taken from existing familial annotation where available, and also deduced from IBD analysis in Plink v1.07 [54]. Two parallel approaches were then taken to narrow down this set of high quality, novel, protein-changing variants to those most likely to be pathogenic in ALS. Firstly, we identified the variants that were shared between all exome-sequenced cases of the 50 families

for which data was available for more than one family member, an approach used successfully in many Mendelian disorders [55]. Secondly, we identified all the variants that were shared by three or more probands, selected as the sample from each of the 694 families with the highest proportion of the exome covered at a depth $\geq 10\times$. The read alignments were manually inspected for all resulting variants and clearly-identifiable false positives removed, resulting in an effective revised filter cut-off of $GQ \geq 90$.

Simulation studies using ExAC variant calls

In order to assess the null hypothesis, that finding a novel protein-changing variant ≥ 4 times would also be likely to occur in a non-disease cohort, a simulation strategy was chosen based on variant calls from the ExAC database. The VCF files for the r.0.3 release of ExAC were annotated and filtered by the identical pipelines described above, with the exception that only variants present within the NFE (Non-Finnish European) population of ExAC samples were considered. ExAC NFE heterozygote and homozygote counts were added together to give NFE carrier counts. The data was then further filtered to only retain variants shared by ≥ 4 carriers ($n=67,154$). These variants were then randomly distributed across 33,370 simulated ExAC European individuals. The FALS + SALS cohort was then simulated by randomly selecting 874 of these individuals (694 FALS probands + 180 SALS) from the total pool. To replicate the variant being novel across the remainder of the dataset, each variant in the filtered ExAC dataset was rejected if all NFE carriers were not within the simulated cohort. A cohort was deemed to have passed if at least one variant met the above criteria. The p-value was calculated as the proportion of cohorts that passed, after sufficient permutations of random cohorts had been generated to achieve a stable result to 2 significant figures. All random integers were generated using the perl CPAN Math::Random::Secure irand function.

Genetics Screening

All coding exons of *ANXA11* (Refseq ID NM_145869) were amplified using standard polymerase chain reaction procedures. At least 100bp of flanking intronic sequence was included to detect splice site mutations (Primers listed in Supplementary Table 4). Amplicons were directly sequenced with Big-Dye Terminator v1.1 on an ABI3130 genetic analyzer (Applied Biosystems Pty Ltd, Warrington, UK) and sequence chromatograms analyzed for mutations using Sequencher 4.10 directly by eye (Gene Codes Corporation, Ann Arbor, Michigan, USA). Reconfirmation of novel mutations was conducted by re-dilution of stock DNA and re-PCR and direct sequencing. Rare

variant positions in *ANXA11* were also filtered against 3596 Italian exome controls. Nine hundred and nine local control UK samples matched for sex and age were assayed for *ANXA11* genomic SNP positions for the G38R, D40G, G189E and R235Q mutations using KASPar genotyping methods by LGC Genomics (LGC Genomics, Teddington, UK).

Haplotype Study

Microsatellites surrounding the *ANXA11* locus were identified using Hg19 build of the UCSC Genome Browser (<http://www.genome.ucsc.edu/>) and amplified by standard PCR with the forward primer incorporating a fluorescent FAM label. 1ul of PCR product was run using Fragment Analysis on an ABI3130 genetic analyzer. Allele sizing was conducted using GeneMapper V4. SNP markers rs41291392, rs17617713, rs72821609 and rs72805713 were genotyped using standard PCR and direct sequencing (primers listed in Supplementary Table 4). The genomic haplotype locus figure (Supplementary Figure 2) was drawn using the 'Dalliance' program [56].

Plasmids and Cloning

A cDNA encoding *ANXA11* was amplified from pEGFP-C3/*ANXA11* [57] with the following pair of oligonucleotides with a *Bam*HI site (underlined): 5'-TAGGATCCACCATGAGCTACCCTGGCTATCC-3' (sense) and 5'-GCGGATCCGAGTCATTGCCACCACAGATCTT-3' (antisense). The DNA fragment obtained was subcloned into the pCR-BluntII-TOPO (Invitrogen). To construct p*ANXA11*-GFP, a *Bam*HI fragment from pCR-Blunt II-TOPO/*ANXA11* was inserted into the *Bam*HI site of pGFP-N-SGG [58]. Single amino acid changes (G38R, D40G and R235Q) were introduced into p*ANXA11*-GFP by PCR-based site-directed mutagenesis. To construct pFLAG-S100A6, an *Eco*RI fragment of pCR2.1 TOPO/S100A6 (kindly provided by Dr. Kiyotaka Hitomi, Nagoya University, Japan) was inserted into the *Eco*RI site of pCMV3xFLAG-B [59]. A cDNA encoding sorcin was amplified from human fetus cDNA library (Clontech) and subcloned into the pCMV-Tag 2A (Stratagene) to construct pFLAG-sorcin. An expression vector encoding ALG-2 N-terminally tagged with FLAG (pFLAG-ALG-2 RNAi^R) was described previously [60].

ANXA11 expression vectors (Gateway pcDNA3.1/nV5-DEST, Invitrogen) encoding c-terminal HA tagged *ANXA1* wild type (WT), G38R, D40G and R235Q mutants were used in this study. Site-directed mutagenesis was performed according to the manufacturer's protocol (Quickchange II Site-Directed Mutagenesis Kit, Stratagene) using a *ANXA11*-HA tagged pDONR221 entry clone plasmid to produce constructs

harboring the novel mutants identified in this study. The pDONR221-ANXA11 mutant constructs were then recombined with pcDNA3.1/nV5-DEST to create the final mutant expression constructs. All constructs were verified by sequencing.

Antibodies

Mouse monoclonal anti-GFP at 1/2000 (cat.no sc-9996, Santa Cruz, Dallas, USA), mouse monoclonal anti-GAPDH (cat.no G8795, Gillingham, UK) and mouse monoclonal Histone H3 (cat.no 96C10, New England Biolabs Hitchin, UK) were used for detecting lysate, soluble and insoluble fractions on Western blot from NP40 insolubility assays. In the Annexin A11:EF-hand proteins (calcyclin, ALG-2 and sorcin) binding assays, mouse monoclonal anti-FLAG (cat.no. F3165, Sigma-Aldrich) was used to immunoprecipitate FLAG-tagged EF-hand proteins, and mouse monoclonal anti-GFP (cat.no. sc-9996, Santa Cruz) and mouse monoclonal anti-FLAG or rabbit polyclonal anti-FLAG (cat.no. F7425, Sigma-Aldrich) were used to detect GFP-fused proteins and FLAG-tagged proteins, respectively. Polyclonal Rabbit Anti-ANXA11 (cat.no. 10479-2-AP, Proteintech, Manchester, UK) was used for ANXA11^{R235Q} staining of HEK cells (Supplementary Figure 4) and spinal cord of the D40G SALS patient, SALS patient devoid of known ALS causing mutation, other neurodegenerative disorders and controls. Polyclonal rabbit Anti-calcyclin (cat.no. 10245-1-AP, Proteintech) was utilized for IHC of patient and control post mortem tissue and detection of FLAG-tagged calcyclin by Western blot in HEK cells.

Transfection of HEK and SH-SY5Y cells

HEK293T and SH-SY5Y cells were maintained in DMEM with high glucose plus Glutamax and DMEM/F12 respectively (Life Technologies, Paisley, UK) with 10% fetal bovine serum, 100U/mL penicillin, and 100 mg/mL streptomycin in a water jacketed incubator at 5% CO₂. For solubility fractionation, transfections were performed in 12-well plates with 500 ng of plasmid DNA and 1.5ml Eugene HD (Promega, Southampton, UK) per well, as per the manufacturer's protocol. For immunofluorescence, HEK293T cells were plated at 25,000 cells/cm² on 13 mm diameter, 1.5-mm-thickness coverslips coated with poly-D-lysine (Sigma-Aldrich, Dorset, UK) in 24-well plates and transfected with 250ng of DNA and 0.75ml of Eugene HD. SH-SY5Y cells were transfected in the same manner except the transfection reagent used was Lipofectamine (Fisher Scientific, Loughborough, UK).

FLAG-S100A6 and ANXA11-GFP Binding Assays

HEK293 YS14 cells (a sub-cloned HEK293 cell line) [61] which had been plated on the previous day, at a density of 1.8×10^6 cells per 10 cm dish were transfected with expression plasmids using Polyethyleneimine (150 μ g) (Polysciences) and then cultured for 24 hr. To attain equivalent expression level, the following amounts of plasmid DNA for SGFP2-fused proteins were used: pSGFP2-N-SGG, 3 μ g; pANXA11^{WT}-SGFP2, 5 μ g; G38R, 6 μ g; D40G, 7 μ g; G189E, 6 μ g; R235Q, 10 μ g; P8L, 5 μ g; D40H, 5 μ g; R191Q, 5 μ g; R230C 5 μ g. The cells were washed and harvested with PBS, suspended in 210 μ L per 10 cm dish (for cells expressing SGFP2-fused proteins in the case of interaction with calcyclin), 470 μ L per 10 cm dish (for cells expressing SGFP2-fused proteins in the case of interactions with ALG-2 or sorcin) or in 350 μ L per 10 cm dish (for cells expressing FLAG-tagged proteins) of lysis buffer (20 mM HEPES-KOH, pH 7.2, 142.5 mM KCl, 2.5 mM MgCl₂) containing 0.2% Triton X-100, 10 μ M EGTA, protease inhibitors (1 μ M E64, 3 μ g/mL leupeptin, 0.1 mM pefabloc, 2 μ M pepstatin A, 0.2 mM PMSF) and phosphatase inhibitors (50 mM NaF, 10 mM β -glycerophosphate, 1 mM Na₃VO₄). After 30 min on ice, the cell lysates were centrifuged at 15,000 *g* for 10 min at 4°C. The cleared lysates of the cell expressing GFP-fused proteins were divided into aliquots of 200 μ L and mixed with 100 μ L of the cleared lysates of the cells expressing FLAG-tagged proteins. Twenty μ L of the mixture was taken as input. The samples were rotated at 4°C for 90 min in the presence of 100 μ M of CaCl₂, mixed with 0.8 μ g of mouse antibody against FLAG (M2) (cat.no F7425, Sigma, Poole, UK) and rotated at 4°C for further 60 min. Then the samples were incubated overnight at 4°C with 10 μ L of Dynabeads Protein G (Novex, Invitrogen). The beads were collected using a magnet and washed twice with 500 μ L of lysis buffer containing 0.1% Triton X-100 and 100 μ M CaCl₂. Washed beads were boiled for 5 min in 56 μ L of 1x SDS-PAGE sample buffer. Four μ L of input and 8 μ L of IP (for detection of FLAG-tagged proteins) or 1.2 μ L of input and 12 μ L of IP (for detection of GFP-fused proteins) were run on a western blot.

FLAG-Calcyclin and ANXA11-GFP over-expression in HEK cells

500ng of ANXA11-GFP^{WT} and ANXA11-GFP^{R235Q} constructs were co-transfected with 500ng of FLAG-Calcyclin in a 6 well dish of HEK cells plated at 25,000 cells/cm². Untreated ANXA11-GFP^{WT} and ANXA11-GFP^{R235Q} were co-transfected with empty pEGFP-C1 vector (Clontech Pty Ltd, Mountain View, California, USA). Duplicate wells of ANXA11-GFP^{WT} and ANXA11-GFP^{R235Q} transfected cells were also treated

with 0.5 μ M MG132 24hrs before harvesting (48 hours post transfection). Cells were harvested and processed as per the NP40 insolubility assay below.

Primary Motor Neuron (PMN) Cultures

Motor neurons were isolated from mouse embryos, cultured and transfected as described [62]. Cells were fixed 1 or 4 days after transfection and processed for immunofluorescence. After antigen retrieval in citrate buffer for 20 minutes and blocking in 5% bovine serum albumin for 1 hour, cells were hybridized overnight at 4°C with HA primary antibody (Covance, 1:1000) Alexa488 or Alexa594-conjugated secondary antibodies (Jackson Immuno-research) were incubated for 1 hour at room temperature. Vesicles were defined as structures with a diameter between 0.5 μ m to 1.9 μ m (mean 1 μ m) and foci 0.16 μ m to 0.5 μ m (mean 0.3 μ m).

Protein Fractionation, Insolubility Assay and Western Blot

ANXA11-GFP WT and mutant construct transfected HEK cells were harvested at 48 wells post-transfection and fractionated using a NP40 solubility assay as described previously [63]. Lysate and soluble fractions were quantified using a BSA standard protein assay and 5 μ g of lysate and soluble fractions analysed by Western blotting. The same volume of insoluble fraction was loaded as per cell lysate. Nitrocellulose membranes were probed with mouse monoclonal anti-GFP at 1/2000 dilution, mouse monoclonal anti-GAPDH 1/2000 dilution and mouse monoclonal Histone H3 1/1000 dilution in 1% Skim Milk and PBS-Tween. Band intensities were quantified using Image J (<http://imagej.nih.gov/ij/>). Membrane imaging was conducted with fluorescent secondary antibodies, Goat Anti-rabbit and Anti-mouse IgG (H+L) Dylight 680 Conjugate (Thermolifesciences cat #35568 and #35521) on a Licor Odyssey Imager. 15 μ g of frontal cortex lysate was run on PVDF membranes and probed with 1/400 Proteintech Rabbit ANXA11 and developed using ECL and visualized using anti-rabbit chemi-luminescent horse raddish peroxidase (HRP) (Millipore Immobilon cat# WBKLS0500).

Co-transfection Immuno-precipitation

Individual wells of a 6 well plate were seeded with HEK cells and transfected at 70% confluence with either Annexin A11-GFP^{R235Q} or Annexin A11-GFP^{R235Q} plasmid. A duplicate set of wells with Annexin A11-GFP^{WT} or Annexin A11-GFP^{R235Q} were co-transfected with Annexin A11-HA^{WT}. At 48hrs cells were harvested in immunopurification buffer (50mM Tris pH7.4, 150mM NaCl, 1% TritonTM X-100, 100uM CaCl₂ with protease and phosphatase inhibitor). The cell lysate was collected

and incubated with immunopurification antibody (rabbit anti-GFP, cat.no AB290 Abcam) and Dynabead protein G (Life Technologies) overnight at 4°C. The Dynabead protein Antibody- protein complex was purified using magnetic separation and washed with immunopurification buffer before elution in loading buffer (variation on protocol described in [64]. Lysates, IP and flow through fractions were run on a nitrocellulose membrane and membrane imaging was conducted with fluorescent secondary antibodies and a Licor Odyssey.

Statistical Analysis

Statistical analysis of PMN immunofluorescence data, NP40 insolubility data and FACs Survival Data was performed with Graphpad Prism software. Normality of the datasets was assessed with the D'Agostino-Pearson test. One-way ANOVA with Dunnett's post *hoc* test or Kruskal-Wallis with Dunn's post *hoc* test were performed depending on normality.

Mutation modelling and Jpred Analysis

The Refseq protein database was searched with human ANXA11 (NP_001148.1) via BlastP at the NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). The highest matching sequence per vertebrate species from the first 250 results was aligned with all others using Muscle (<http://www.ebi.ac.uk/Tools/msa/muscle/>). Sequences were removed if they had a better reciprocal match to a different human annexin gene or if they contained significant numbers of gaps, mismatching regions or consecutive "X"s, implying incomplete or mistranslated sequence. Fish were also excluded, as they consistently showed significant divergence along the majority of the N-terminus, and with the exception of the first 20 amino acids a satisfactory alignment to other taxonomic Classes could not be obtained. The resulting multiple alignment was hand edited in GeneDoc and submitted to JPred4 (<http://www.compbio.dundee.ac.uk/jpred4/>) for secondary structure prediction. Predictions were also made for the ALS-associated variants G38R and D40G, and 22 non-ALS variants from the ExAC database flanking these residues, by replacing every amino acid in the multiple alignment with the variant residue. Possible amphipathicity of any resulting alpha helices was tested by submitting the sequences to Helixator (http://www.tcdb.org/progs/helical_wheel.php). The procedure was then repeated for human ANXA1, but with this gene it was not necessary to exclude fish species from the alignment.

Neuropathology

Brain tissue samples in 10% formalin-fixed, paraffin-embedded tissue blocks were available from the London Neurodegenerative Diseases Brain Bank (King's College London, UK). Consent for autopsy, neuropathological assessment and research were obtained from all subjects and all studies were carried out under the ethical approval of the tissue bank. Block taking for histological and immunohistochemical studies and neuropathological assessment for neurodegenerative diseases was performed in accordance with standard criteria.

Immunohistochemistry:

Immunohistochemistry was carried out as per previously published protocols [65]. In brief, sections of 7µm thickness were cut from the paraffin-embedded tissue blocks, deparaffinised in xylene, endogenous peroxidase was blocked by 2.5% H₂O₂ in methanol and immunohistochemistry performed. To enhance antigen retrieval, sections were kept in citrate buffer for 10 minutes following microwave treatment. After blocking in normal serum primary antibody was applied overnight at 4°C.

Following washes, sections were incubated with biotinylated secondary antibody (DAKO), followed by avidin:biotinylated enzyme complex (Vectastain Elite ABC kit, Vector Laboratories, Peterborough, UK). Finally sections were incubated for 10–15 min with 0.5 mg/mL 3,3'-diaminobenzidine chromogen (Sigma-Aldrich Company Ltd, Dorset UK) in Tris-buffered saline (pH 7.6) containing 0.05% H₂O₂. Sections were counterstained with Harris' haematoxylin and immunostaining analysed using a Leica microscope (Leica, Wetzlar, Germany).

Double Immunofluorescence:

7µm sections were cut from formalin fixed paraffin embedded blocks, dewaxed in xylene and dehydrated in 99% industrial methylated spirit. Sections were then pretreated by microwaving in citrate buffer and blocked using normal goat serum (1:10 for 45min). Primary antibodies were then applied and sections incubated at 4°C overnight. Sections were washed and secondary Alexa Fluor antibody (Invitrogen, Paisley, UK) applied for 45 min (in dark). Autofluorescence was quenched by incubating the sections in Sudan black for 10min followed by numerous washes in phosphate buffered saline before coverslip mounting using hard set media with DAPI. Sections were visualised using a fluorescent microscope (Zeiss Axiovert S 100, Gottingen, Germany) and images captured using ImagePro Express (V6).

References

1. Abhinav, K., et al., *Amyotrophic lateral sclerosis in South-East England: a population-based study. The South-East England register for amyotrophic lateral sclerosis (SEALS Registry)*. Neuroepidemiology, 2007. **29**(1-2): p. 44-8.
2. Smith, B.N., et al., *The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder*. Eur J Hum Genet, 2013. **21**(1): p. 102-8.
3. Al-Chalabi, A., et al., *The genetics and neuropathology of amyotrophic lateral sclerosis*. Acta Neuropathol, 2012. **124**(3): p. 339-52.
4. White, M.A. and J. Sreedharan, *Amyotrophic lateral sclerosis: recent genetic highlights*. Curr Opin Neurol, 2016. **29**(5): p. 557-64.
5. Wu, C.H., et al., *Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis*. Nature, 2012. **488**(7412): p. 499-503.
6. Johnson, J.O., et al., *Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis*. Nat Neurosci, 2014. **17**(5): p. 664-6.
7. Bannwarth, S., et al., *A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement*. Brain, 2014. **137**(Pt 8): p. 2329-45.
8. Johnson, J.O., et al., *Exome sequencing reveals VCP mutations as a cause of familial ALS*. Neuron, 2010. **68**(5): p. 857-64.
9. Smith, B.N., et al., *Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS*. Neuron, 2014. **84**(2): p. 324-31.
10. Cirulli, E.T., et al., *Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways*. Science, 2015.
11. Freischmidt, A., et al., *Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia*. Nat Neurosci, 2015.
12. Kenna, K.P., et al., *NEK1 variants confer susceptibility to amyotrophic lateral sclerosis*. Nat Genet, 2016.
13. Brenner, D., et al., *NEK1 mutations in familial amyotrophic lateral sclerosis*. Brain, 2016. **139**(Pt 5): p. e28.
14. Williams, K.L., et al., *CCNF mutations in amyotrophic lateral sclerosis and frontotemporal dementia*. Nat Commun, 2016. **7**: p. 11253.
15. van Rheenen, W., et al., *Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis*. Nat Genet, 2016. **48**(9): p. 1043-8.
16. Gerke, V. and S.E. Moss, *Annexins: from structure to function*. Physiol Rev, 2002. **82**(2): p. 331-71.
17. Rezvanpour, A. and G.S. Shaw, *Unique S100 target protein interactions*. Gen Physiol Biophys, 2009. **28 Spec No Focus**: p. F39-46.
18. Wang, J., et al., *Annexin A11 in disease*. Clin Chim Acta, 2014. **431**: p. 164-8.
19. Hofmann, S., et al., *Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis*. Nat Genet, 2008. **40**(9): p. 1103-6.
20. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans*. Nature, 2016. **536**(7616): p. 285-91.

21. Enayat, Z.E., et al., *Two novel mutations in the gene for copper zinc superoxide dismutase in UK families with amyotrophic lateral sclerosis*. Hum Mol Genet, 1995. **4**(7): p. 1239-40.
22. Kabashi, E., et al., *TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis*. Nat Genet, 2008. **40**(5): p. 572-4.
23. Kwiatkowski, T.J., Jr., et al., *Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis*. Science, 2009. **323**(5918): p. 1205-8.
24. Puls, I., et al., *Mutant dynactin in motor neuron disease*. Nat Genet, 2003. **33**(4): p. 455-6.
25. Sreedharan, J., et al., *TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis*. Science, 2008. **319**(5870): p. 1668-72.
26. Rosen, D.R., et al., *Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis*. Nature, 1993. **362**(6415): p. 59-62.
27. Deng, H.X., et al., *Amyotrophic lateral sclerosis and structural defects in Cu,Zn superoxide dismutase*. Science, 1993. **261**(5124): p. 1047-51.
28. Esteban, J., et al., *Identification of two novel mutations and a new polymorphism in the gene for Cu/Zn superoxide dismutase in patients with amyotrophic lateral sclerosis*. Hum Mol Genet, 1994. **3**(6): p. 997-8.
29. Chio, A., et al., *Two Italian kindreds with familial amyotrophic lateral sclerosis due to FUS mutation*. Neurobiol Aging, 2009. **30**(8): p. 1272-5.
30. Abecasis, G.R., et al., *Merlin--rapid analysis of dense genetic maps using sparse gene flow trees*. Nat Genet, 2002. **30**(1): p. 97-101.
31. Ralph, P. and G. Coop, *The geography of recent genetic ancestry across Europe*. PLoS Biol, 2013. **11**(5): p. e1001555.
32. Mizutani, A., et al., *The long amino-terminal tail domain of annexin XI is necessary for its nuclear localization*. Arch Biochem Biophys, 1995. **318**(1): p. 157-65.
33. Lesniak, W., L.P. Slomnicki, and A. Filippek, *S100A6 - new facts and features*. Biochem Biophys Res Commun, 2009. **390**(4): p. 1087-92.
34. Sudo, T. and H. Hidaka, *Characterization of the calyculin (S100A6) binding site of annexin XI-A by site-directed mutagenesis*. FEBS Lett, 1999. **444**(1): p. 11-4.
35. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server*. Nucleic Acids Res, 2015.
36. Gerke, V., C.E. Creutz, and S.E. Moss, *Annexins: linking Ca²⁺ signalling to membrane dynamics*. Nat Rev Mol Cell Biol, 2005. **6**(6): p. 449-61.
37. Rosengarth, A. and H. Luecke, *A calcium-driven conformational switch of the N-terminal and core domains of annexin A1*. J Mol Biol, 2003. **326**(5): p. 1317-25.
38. Osako, Y., et al., *Autolytic activity of human calpain 7 is enhanced by ESCRT-III-related protein IST1 through MIT-MIM interaction*. FEBS J, 2010. **277**(21): p. 4412-26.
39. Hoyaux, D., et al., *S100A6 overexpression within astrocytes associated with impaired axons from both ALS mouse model and human patients*. J Neuropathol Exp Neurol, 2002. **61**(8): p. 736-44.

40. Hoyaux, D., et al., *S100A6, a calcium- and zinc-binding protein, is overexpressed in SOD1 mutant mice, a model for amyotrophic lateral sclerosis*. Biochim Biophys Acta, 2000. **1498**(2-3): p. 264-72.
41. Al-Chalabi, A., et al., *Recessive amyotrophic lateral sclerosis families with the D90A SOD1 mutation share a common founder: evidence for a linked protective factor*. Hum Mol Genet, 1998. **7**(13): p. 2045-50.
42. Nishimura, A.L., A. Al-Chalabi, and M. Zatz, *A common founder for amyotrophic lateral sclerosis type 8 (ALS8) in the Brazilian population*. Hum Genet, 2005. **118**(3-4): p. 499-500.
43. Chio, A., et al., *Large proportion of amyotrophic lateral sclerosis cases in Sardinia due to a single founder mutation of the TARDBP gene*. Arch Neurol, 2011. **68**(5): p. 594-8.
44. Lecona, E., et al., *Structural and functional characterization of recombinant mouse annexin A11: influence of calcium binding*. Biochem J, 2003. **373**(Pt 2): p. 437-49.
45. Tomas, A., C. Futter, and S.E. Moss, *Annexin 11 is required for midbody formation and completion of the terminal phase of cytokinesis*. J Cell Biol, 2004. **165**(6): p. 813-22.
46. Shibata, H., et al., *A new role for annexin A11 in the early secretory pathway via stabilizing Sec31A protein at the endoplasmic reticulum exit sites (ERES)*. J Biol Chem, 2015. **290**(8): p. 4981-93.
47. Fukushima, T., et al., *Critical function for SIP, a ubiquitin E3 ligase component of the beta-catenin degradation pathway, for thymocyte development and G1 checkpoint*. Immunity, 2006. **24**(1): p. 29-39.
48. Polymenidou, M., et al., *Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43*. Nat Neurosci, 2011. **14**(4): p. 459-68.
49. Ling, J.P., et al., *NEURODEGENERATION. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD*. Science, 2015. **349**(6248): p. 650-5.
50. Botelho, H.M., et al., *S100A6 amyloid fibril formation is calcium-modulated and enhances superoxide dismutase-1 (SOD1) aggregation*. J Biol Chem, 2012. **287**(50): p. 42233-42.
51. Prell, T., J. Lautenschlager, and J. Grosskreutz, *Calcium-dependent protein folding in amyotrophic lateral sclerosis*. Cell Calcium, 2013. **54**(2): p. 132-43.
52. Brooks, B.R., et al., *El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis*. Amyotroph Lateral Scler Other Motor Neuron Disord, 2000. **1**(5): p. 293-9.
53. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data*. Nucleic Acids Res, 2010. **38**(16): p. e164.
54. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses*. Am J Hum Genet, 2007. **81**(3): p. 559-75.
55. Yu, Y., et al., *Whole-exome sequencing identifies rare, functional CFH variants in families with macular degeneration*. Hum Mol Genet, 2014. **23**(19): p. 5283-93.

56. Down, T.A., M. Piipari, and T.J. Hubbard, *Dalliance: interactive genome viewing on the web*. Bioinformatics, 2011. **27**(6): p. 889-90.
57. Tomas, A. and S.E. Moss, *Calcium- and cell cycle-dependent association of annexin 11 with the nuclear envelope*. J Biol Chem, 2003. **278**(22): p. 20210-6.
58. Shibata, H., et al., *The ALG-2 binding site in Sec31A influences the retention kinetics of Sec31A at the endoplasmic reticulum exit sites as revealed by live-cell time-lapse imaging*. Biosci Biotechnol Biochem, 2010. **74**(9): p. 1819-26.
59. Katoh, K., et al., *The ALG-2-interacting protein Alix associates with CHMP4b, a human homologue of yeast Snf7 that is involved in multivesicular body sorting*. J Biol Chem, 2003. **278**(40): p. 39104-13.
60. Okumura, M., et al., *Penta-EF-hand protein ALG-2 functions as a Ca²⁺-dependent adaptor that bridges Alix and TSG101*. Biochem Biophys Res Commun, 2009. **386**(1): p. 237-41.
61. Shibata, H., et al., *ALG-2 directly binds Sec31A and localizes at endoplasmic reticulum exit sites in a Ca²⁺-dependent manner*. Biochem Biophys Res Commun, 2007. **353**(3): p. 756-63.
62. Fallini, C., G.J. Bassell, and W. Rossoll, *High-efficiency transfection of cultured primary motor neurons to study protein localization, trafficking, and function*. Mol Neurodegener, 2010. **5**: p. 17.
63. Smith, B.N., et al., *Novel mutations support a role for Profilin 1 in the pathogenesis of ALS*. Neurobiol Aging, 2015. **36**(3): p. 1602 e17-27.
64. Chen, H.J., et al., *The heat shock response plays an important role in TDP-43 clearance: evidence for dysfunction in amyotrophic lateral sclerosis*. Brain, 2016. **139**(Pt 5): p. 1417-32.
65. Maekawa, S., et al., *TDP-43 is consistently co-localized with ubiquitinated inclusions in sporadic and Guam amyotrophic lateral sclerosis but not in familial amyotrophic lateral sclerosis with and without SOD1 mutations*. Neuropathology, 2009. **29**(6): p. 672-83.

List of Supplementary Materials

Supplementary Tables

Supplementary Table 1. Numbers of novel shared candidate variants in 50 ALS families.

Supplementary Table 2. List of candidate variants found in both UK ANXA11 p.D40G families.

Supplementary Table 3. Population breakdown of the FALS cohort used in this study and ExAC.

Supplementary Table 4. Primer sequences used for Sanger sequencing of ANXA11 exons, and SNPs and microsatellite markers in the p.D40G haplotype study.

Supplementary Table 5. Exon coverage data of the 23 genes located in the maximal p.D40G haplotype region spanning ANXA11.

Supplementary Table 6. List of all *ANXA11* variants identified in FALS and SALS cases in this study.

Supplementary Table 7. Clinical information for FALS and SALS patients harbouring *ANXA11* variants.

Supplementary Figures

Supplementary Figure 1. The *Annexin A11* p.D40G mutation shares a common haplotype comprised of 4 exonic SNPs and 2 microsatellites.

Supplementary Figure 2. Haplotype gene map of the *ANXA11* D40G locus.

Supplementary Figure 3. Annexin A11 western blot of lysates made from post-mortem tissue of the sporadic patient harbouring the p.D40G mutation and control individuals.

Supplementary Figure 4. R235Q *ANXA11*-GFP co-localises with ubiquitin and p62 in HEK cells.

Supplementary Figure 5. Modelling of Annexin A11 identifies two amphipathic helices in the N-terminus of Annexin A11 that overlap the G38 and D40 residues.

Supplementary Figure 6. Twenty two ExAC polymorphisms spanning the D40 locus do not disrupt the formation of amphipathic helices.

Supplementary Figure 7. Rare and common *ANXA11* polymorphisms do not disrupt calyculin binding compared to ALS specific variants.

Supplementary Figure 8. N-terminal Annexin A11 mutations do not alter binding of Sorcin and ALG2.

Acknowledgements

We would like to thank people with ALS and their families for their participation in this project. We acknowledge sample management undertaken by Biobanking Solutions funded by the Medical Research Council at the Centre for Integrated Genomic Medical Research, University of Manchester.

Funding

Funding for this work was provided by The Middlemass family, Heaton-Ellis Trust, Motor Neurone Disease Association, Medical Research Council, The Psychiatry Research Trust of the Institute of Psychiatry, Guy's and St Thomas' Charity, the Wellcome Trust and the Noreen Murray Foundation. This is a EU Joint Programme - Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organisations under the aegis of JPND - www.jpnd.eu (United Kingdom, Medical Research Council and Economic and Social Research

Council). This project is supported by the Medical Research Foundation with salary provided for BNS (MRF-060-0003-RG-SMITH). CES and AAC receive salary support from the National Institute for Health Research (NIHR) Dementia Biomedical Research Unit at South London and Maudsley NHS Foundation Trust and King's College London. CF received salary support from the ALS Association (ALSA). Japan Society for the Promotion of Science (15K07384) to Hideki Shibata (HS) Japan Society for the Promotion of Science (26292050) to Masatoshi Maki (MM). P.C.S. was supported through the auspices of Dr. H. Robert Horvitz, an Investigator at the Howard Hughes Medical Institute in the Department of Biology at the Massachusetts Institute of Technology. AriSLA cofinanced with support of "5 x 1000" Healthcare Research of the Ministry of Health (grant NOVALS 2012 [N.T., CT., C.G., V.S., J.E.L.]) and the Italian Ministry of Health (grant GR-2011-02347820 - IRisALS [N.T., C.T]). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. Funding was provided by the National Institutes of Health (NIH)/National Institute of Neurological Disorders and Stroke (NINDS) (R01NS073873 (JEL)), and the American ALS Association, and Project MinE. Funding was also provided by Instituto de Salud Carlos III (grant PI14/00088) and FUNDELA (Spanish foundation for the development of ALS research). The work leading up to this publication was funded by the European Community's Health Seventh Framework Programme (FP7/2007–2013; grant agreement number 259867). Samples used in this research were in part obtained from the UK National DNA Bank for MND Research, funded by the MND Association and the Wellcome Trust.

Author Contributions

B.N.S, S.T, C.F, H.S, H.C, C.T, A.K, S.A.S, J.L, M.W, V.S and C.E.S designed the experiments. B.N.S, S.T, C.F, H.S, H.C, C.T, A.K N.T, K.K, C.V, D.M.D, A.S, A.S.G, J.M, C.H.W, M.M, E.S, W.K, P.S, C.M, N.P and M.N performed the experimental procedures and analyzed the data. P.N, F.B, V.J, A.A, A.R, J.B, J.P, C.T, F.V, S.D, N.L, H.P, K.M, A.A.C, P.J, J.K, M.T, K.T, O.H, J.G, J.D.B, C.G, A.R, R.B, S.A.S, V.S, J.L and C.E.S collected and/or contributed DNA samples or exome sequences. B.N.S, S.D.T and C.E.S wrote the manuscript with contributions from the authors.

Competing Interests

None

Data availability

FALS exome variants generated from the cohort used in this study have been deposited for public access into the University of Massachusetts Medical School ALS Variant Server (<http://als.umassmed.edu>).

Figure Legends

Main Figures

Figure 1. Annexin A11 mutations identified in ALS patients following stringent filtering of exome sequencing data. (A and B) Pedigrees of UK Family 1 and 2 respectively, carrying the p.D40G mutation. Family members for whom DNA was available for segregation analysis are labelled with M/W for a heterozygous mutant (A>G allele) or W/W for a homozygous reference allele. The gender has been anonymised for each individual. Affected individuals with ALS are denoted by a solid black diamond, unaffected by white diamonds and unaffected carriers with a black dot. Affected family members exome sequenced as part of this study are marked with a blue star. (C) The G175R mutation segregates in both an index case and affected sibling. The ages of death, where data is available, is listed above each D40G carrier (affected and unaffected) (A), (B) or (C). (D) Schematic representation of the Annexin A11 molecule highlighting a clustering of mutations in the N-terminus (first 196 residues). Mutation positions in the N-terminus are fully conserved in mammals, and conserved in birds, amphibians and reptiles if in an annexin domain. Mutations in FALS are indicated by red boxes, and in SALS by blue boxes. The D40G mutation found in 2 index FALS and an affected sibling from each UK family, an Italian index case and a UK SALS (n=6) clusters with G38R in the N-terminus of the ANXA11 molecule. The binding site of calyculin (labelled CACY) is located in the N-terminus, residues 50-62.

Figure 2. TDP-43 and Annexin A11 immunohistochemistry from post-mortem spinal cord tissue of a SALS case harbouring the novel D40G mutation. (A) phospho-TDP43 positive cytoplasmic inclusion from anterior horn of the spinal cord. (B-E) Annexin A11 positive inclusions in motor neurons of the cord included skein-like (B), filamentous and tubular- shaped structures (C-D). Occasional more “basket-like” inclusions were seen in the cord (E). Abundant annexin A11 positive torpedo-like neuritic structures were also present in the neuropil here seen in the motor cortex (arrows) (F). Representative spinal cord staining in a SALS case (n=15) negative for

ANXA11 mutations (G) and two *C9ORF72* expansion positive cases (spinal cord and frontal cortex, H and I). *ANXA11* staining was also negative in a *SOD1* positive case harboring a p.D101G mutation (spinal cord, J). Similarly spinal cord cases of FTLD-TDP43 (K, n=3) and Alzheimer's cases (L, n=3) and control individuals (M, n=13) were also negative for annexin A11 inclusions. Double labelling for phospho-TDP43 (green) and Annexin A11 aggregates (red) (N-O). Co-staining for ubiquitinated aggregates (red= Annexin A11 and green=ubiquitin) showed occasional co-localisation (P) (indicated by arrows). Scale bars - A, E, G and H- 30µm; B, C- 20µm; D-15µm; F, I, J, K, L- 50µm, M, N- 25µm, O, P- 50µm.

Figure 3. Functional Annexin A11 studies in mouse Primary Motor Neurons and HEK cells at 4 days post transfection. (A) *ANXA11*-HA^{WT/G38R/D40G/R235Q} transfected into PMNs display cytoplasmic vesicle-like structures for WT, G38R and D40G and a significant localization of smaller cytoplasmic foci in the cytoplasm for R235Q, (B) $p=0.0004$ (One way ANOVA and Dunnett's post hoc test, Bars represent mean and SEM). Vesicles were defined as structures with a diameter between 0.5µm to 1.9 µm (mean 1 µm) and foci 0.16µm to 0.5 µm (mean 0.3 µm). (C) There was a significant proportion of foci associated with R235Q ($p<0.0001$, One way ANOVA and Dunnett's post hoc test) but a smaller proportion of vesicle like structures associated with G38R and a complete absence with R235Q ($p=0.0051$ and $p<0.0001$ respectively, One way ANOVA and Dunnett's post hoc test). (D) R235Q *ANXA11*-GFP forms high molecular weight detergent resistant aggregates in HEK cells. The lysate (L), soluble (S) and insoluble (I) fractions of HEK cells transfected with *ANXA11*-GFP^{WT/G38R/D40G/G189E/R235Q} show R235Q to form high detergent resistant insoluble aggregates. (E) The level of R235Q insoluble fraction is statistically significant when compared with WT ($p=0.007$, One-way ANOVA with Dunnett's post hoc test).

Figure 4. The R235Q mutation sequesters WT Annexin A11. (A) Co-expression of *ANXA11*-HA^{WT} and *ANXA11*-GFP^{R235Q} in SH-SH5Y cells demonstrated co-localisation suggesting sequestering of *ANXA11*-HA^{WT} by mutant *ANXA11*^{R235Q}. Scalebars 10µm. (B) Expression of *ANXA11*-GFP^{WT} and *ANXA11*-GFP^{R235Q} replicated that seen in Figure 3D, with insoluble aggregates in the urea fraction seen only in *ANXA11*-GFP^{R235Q} (Lanes 1-6). When co-expression of *ANXA11*-GFP^{WT} or *ANXA11*-GFP^{R235Q} with an equal amount of *ANXA11*-HA^{WT} was conducted in HEK cells, there was an accumulation of *ANXA11*-HA^{WT} in the insoluble urea fraction only when co-expressed with *ANXA11*-GFP^{R235Q} - indicated by a red star (n=3) (Lanes 7-

12). (C) Furthermore co-transfection of HEK cells with ANXA11-GFP^{R235Q} and ANXA11-HA^{WT} and immuno-precipitation (IP) with rabbit anti-GFP and probing with mouse-HA showed a direct sequestering of ANXA11-HA^{WT} in the IP fraction (Lane 6), including higher insoluble molecular weight species compared to vector and vector plus GFP (Top panel). Verification of the IP was shown by staining with mouse-GFP in Lanes 3 and 6 (Input and IP - Middle Panel) and positive staining of input ANXA11-HA^{WT} in the Input and bead flow-through (FT) in the top panel.

Figure 5. Calcyclin analysis: *in-vitro* binding assays with mutant Annexin A11 constructs and staining of post-mortem tissue. (A) Western blot demonstrating lack of binding of FLAG-calcyclin to ANXA11-GFP^{D40G}, ANXA11-GFP^{G189E} and ANXA11-GFP^{R235Q} compared to ANXA11-GFP^{WT} by immunoprecipitation (IP) (n=3). The top panel shows GFP intensities of input and precipitated fractions, with IgG heavy (50kDa*) and light (25kDa**) chains indicated. The bottom panel illustrates input and IP levels of FLAG-calcyclin. (B) Cross-section of spinal cord showing calcyclin expression in (i) controls and (ii) a SALS case with the p.D40G mutation. High calcyclin expression can be seen in the lateral corticospinal tracts (arrows). (iii) SALS case devoid of known ALS causing mutation but also displaying strong expression of calcyclin in the lateral corticospinal tracts (arrows). Scale bars (i) 40µm, (ii) 30µm and (iii) 25µm. (C) NP40 insolubility assay of co-expression of 500ng of ANXA11-GFP^{WT} and ANXA11-GFP^{R235Q} and 500ng of FLAG-calcyclin in HEK cells, demonstrating clearance of insoluble R235Q aggregates by calcyclin co-expression. An equivalent amount of empty GFP Vector (500ng) was added to untreated WT and R235Q HEK cells. (D) Quantification of ANXA11-GFP^{WT/R235Q} co-expression experiment. Bands were quantified with ImageJ (<http://imagej.nih.gov/ij/>) and intensity levels calculated with respect to untreated ANXA11-GFP^{WT}. A one-way ANOVA test demonstrated significance of the untreated R235Q and R235Q treatment with calcyclin followed by 24hours MG132 treatment (p<0.05 and p<0.001 respectively, One-way ANOVA and Dunnett's post-hoc test) (n=3).

Figure 1. *Annexin A11* mutations identified in ALS patients following stringent filtering of exome sequencing data.

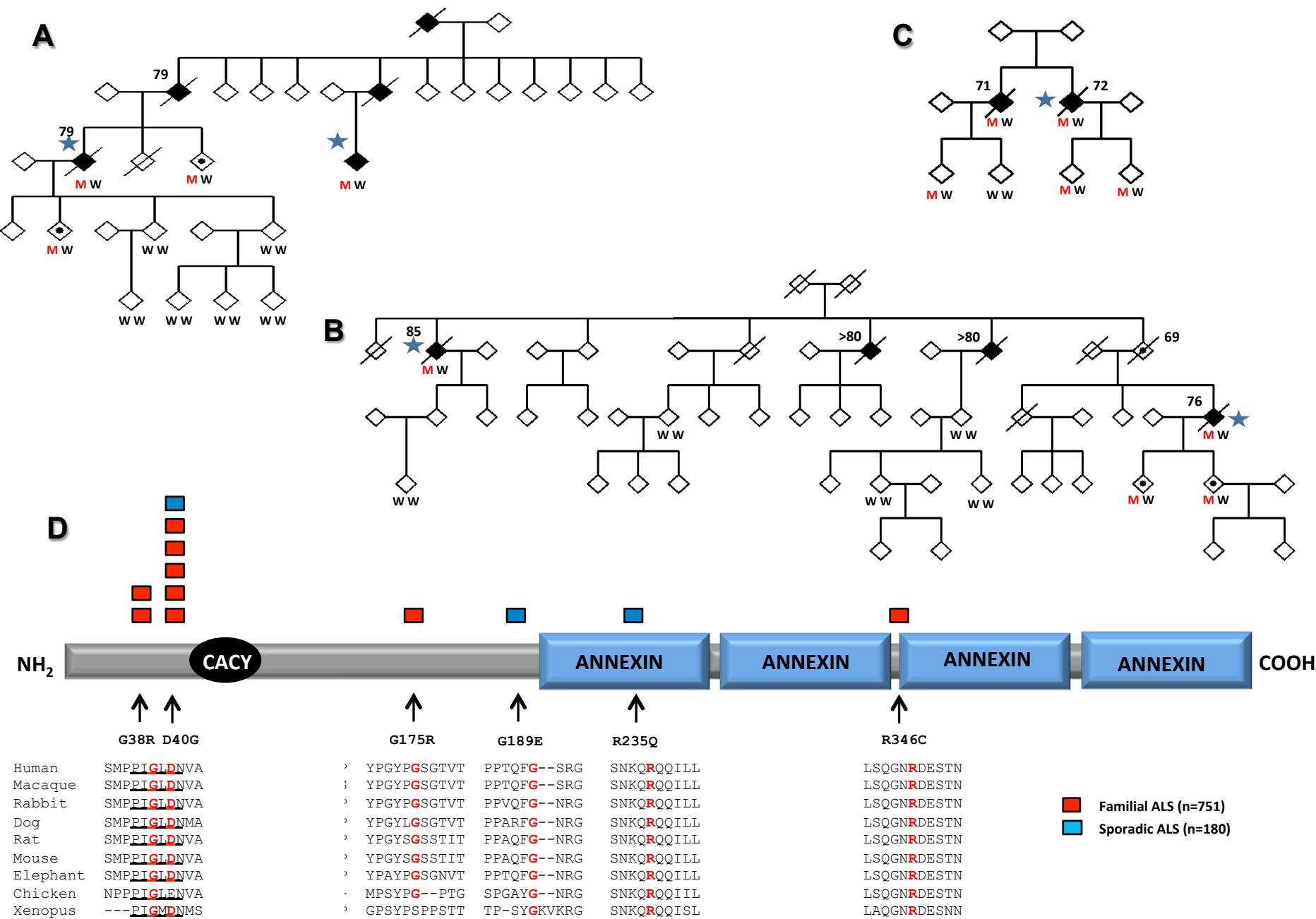


Figure 2. ANXA11 Immunohistochemistry of post mortem tissue from D40G sporadic patient

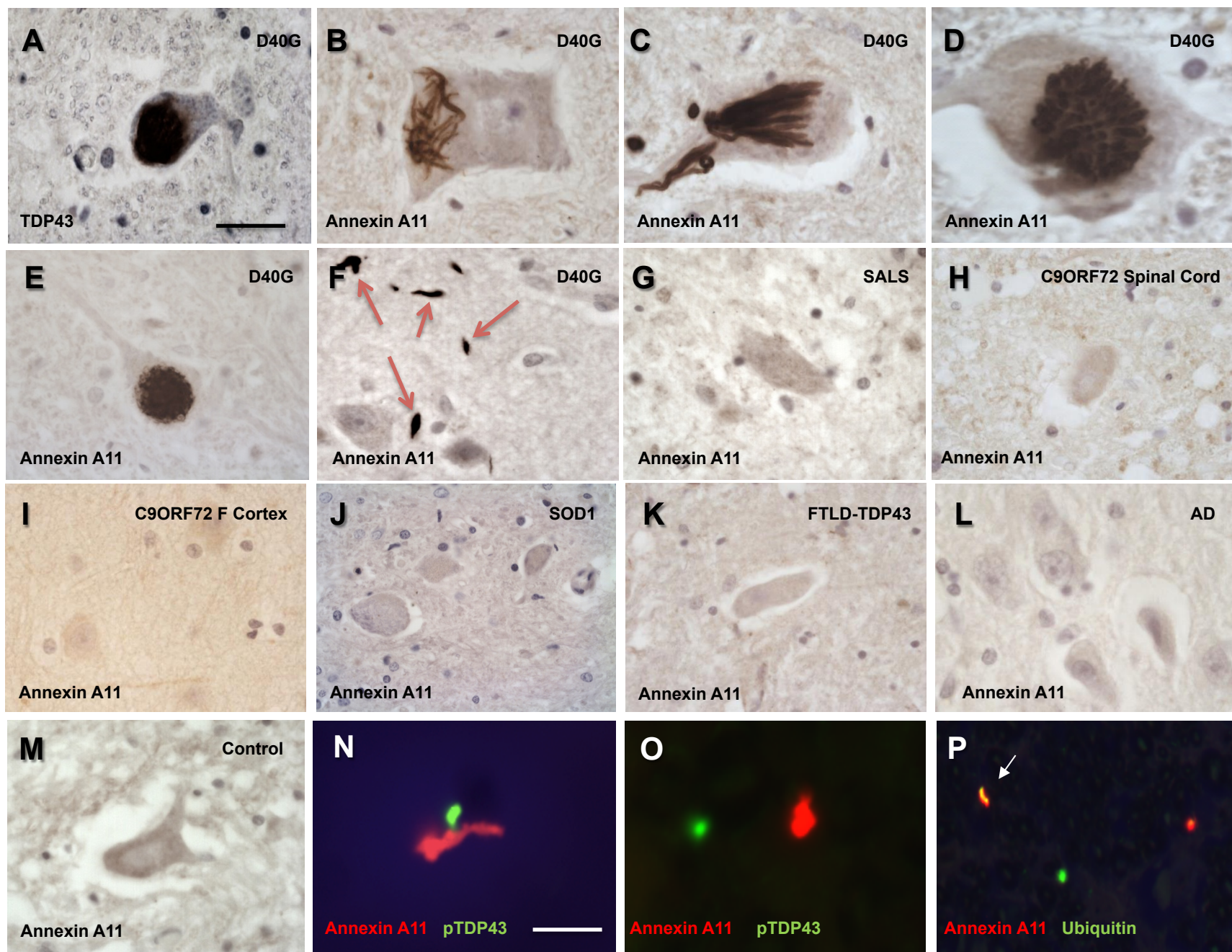


Figure 3. Functional Annexin A11 studies in mouse Primary Motor Neurons and HEK cells

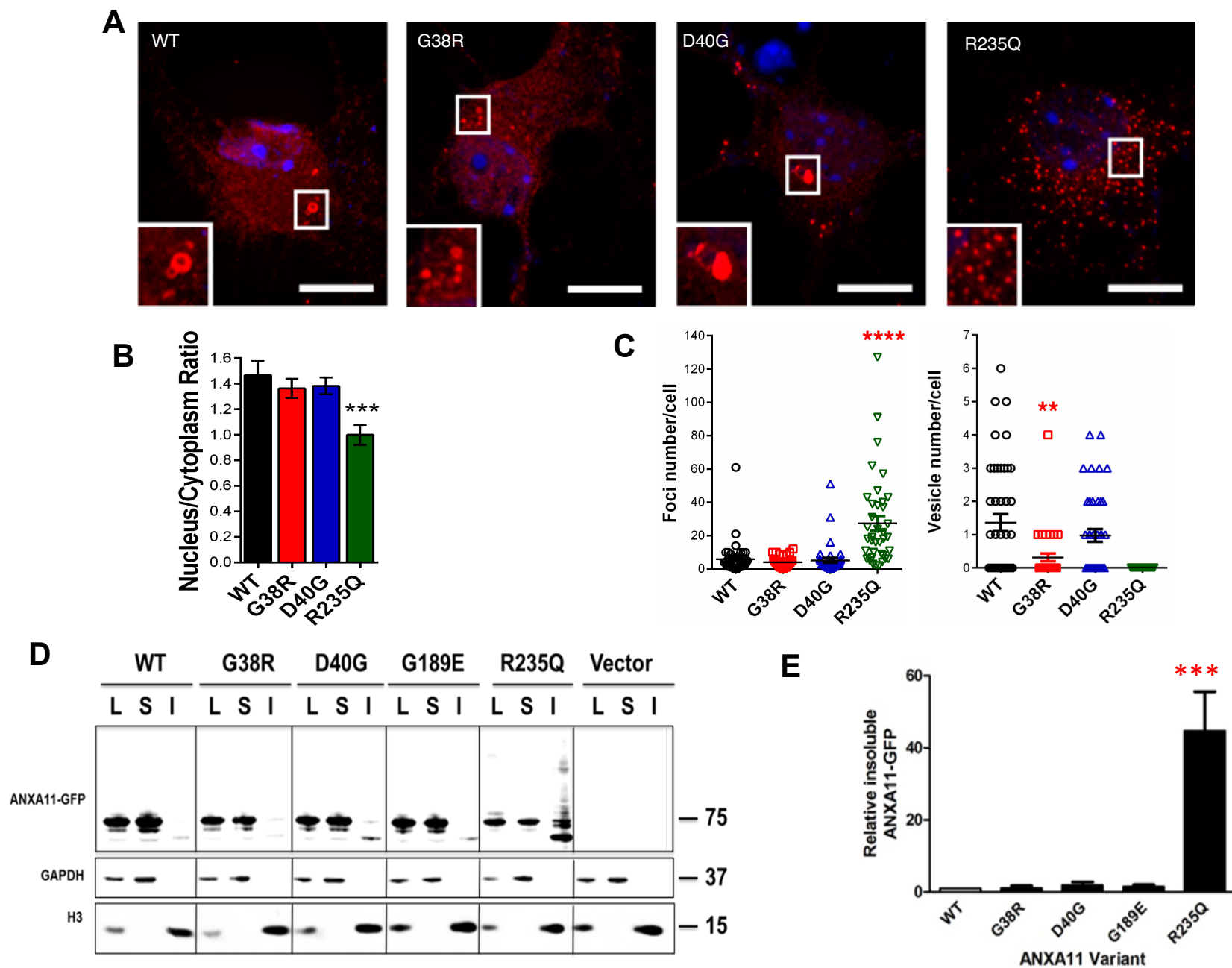
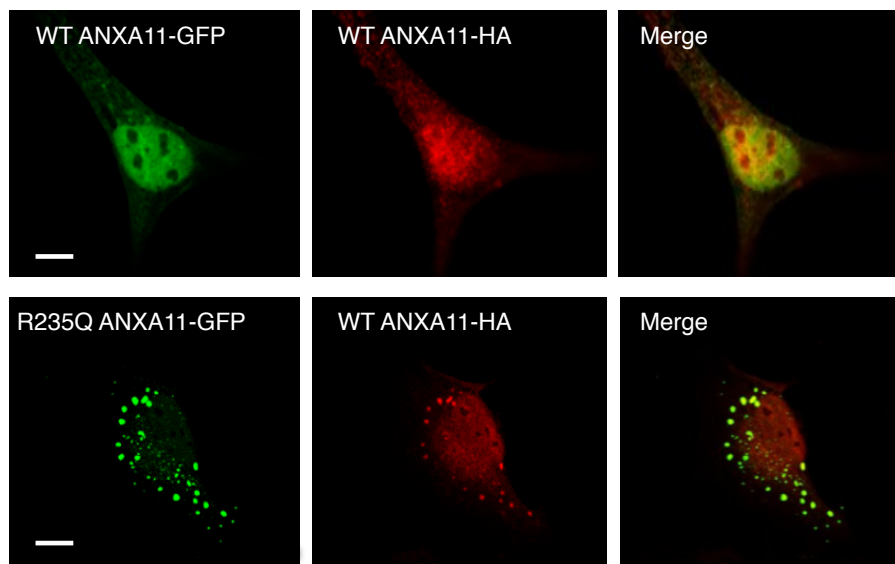
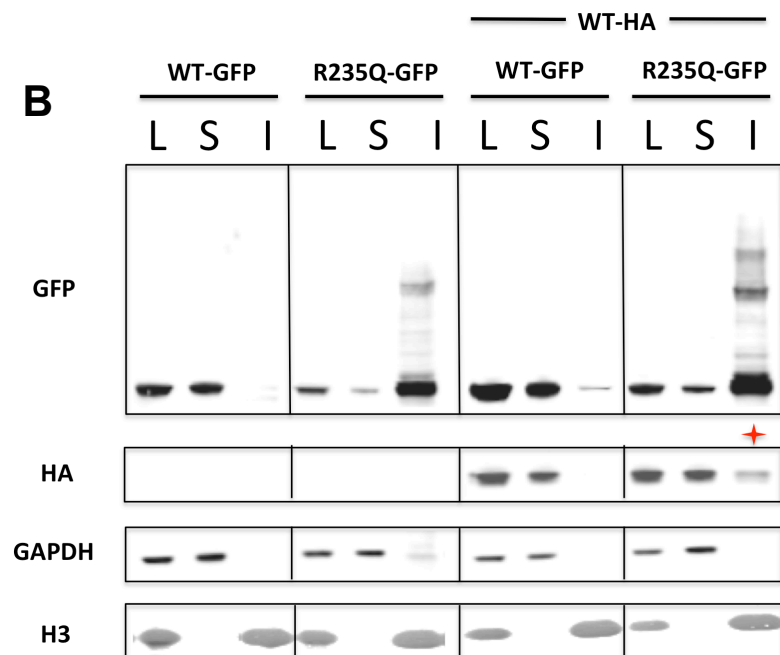


Figure 4. R235Q aggregates sequester WT Annexin A11

A



B



C

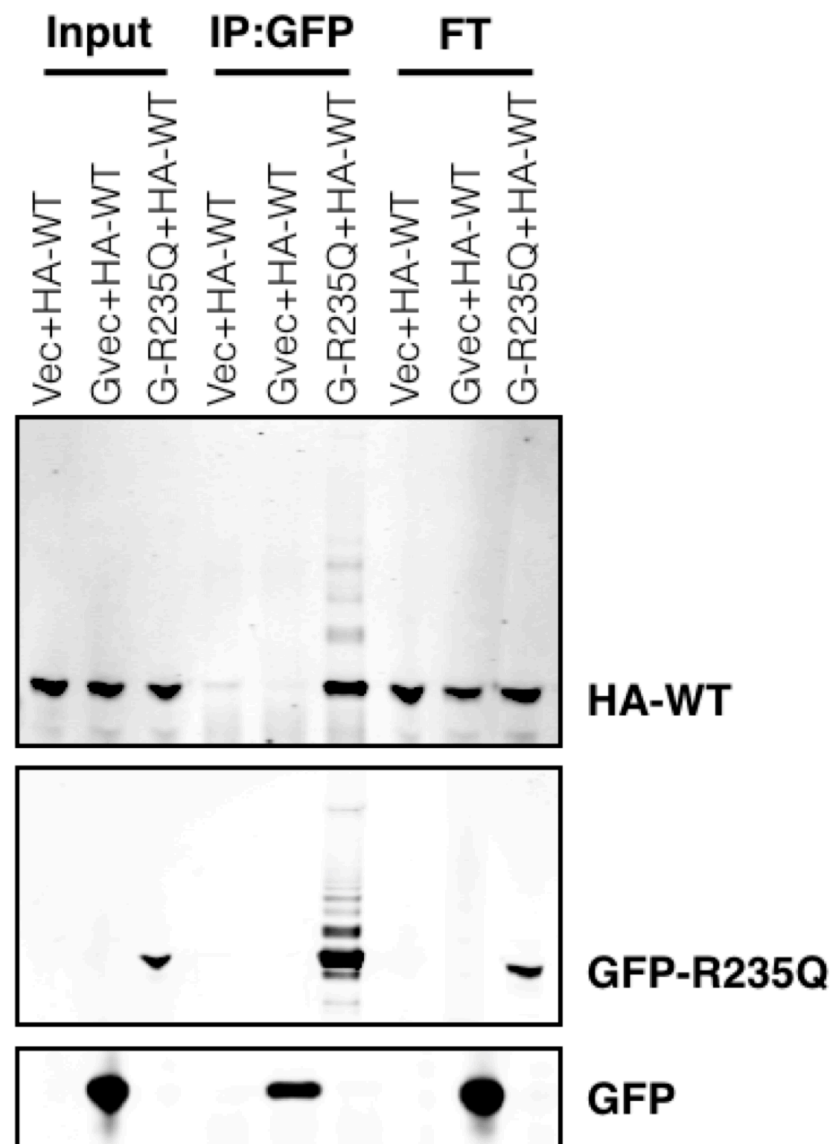
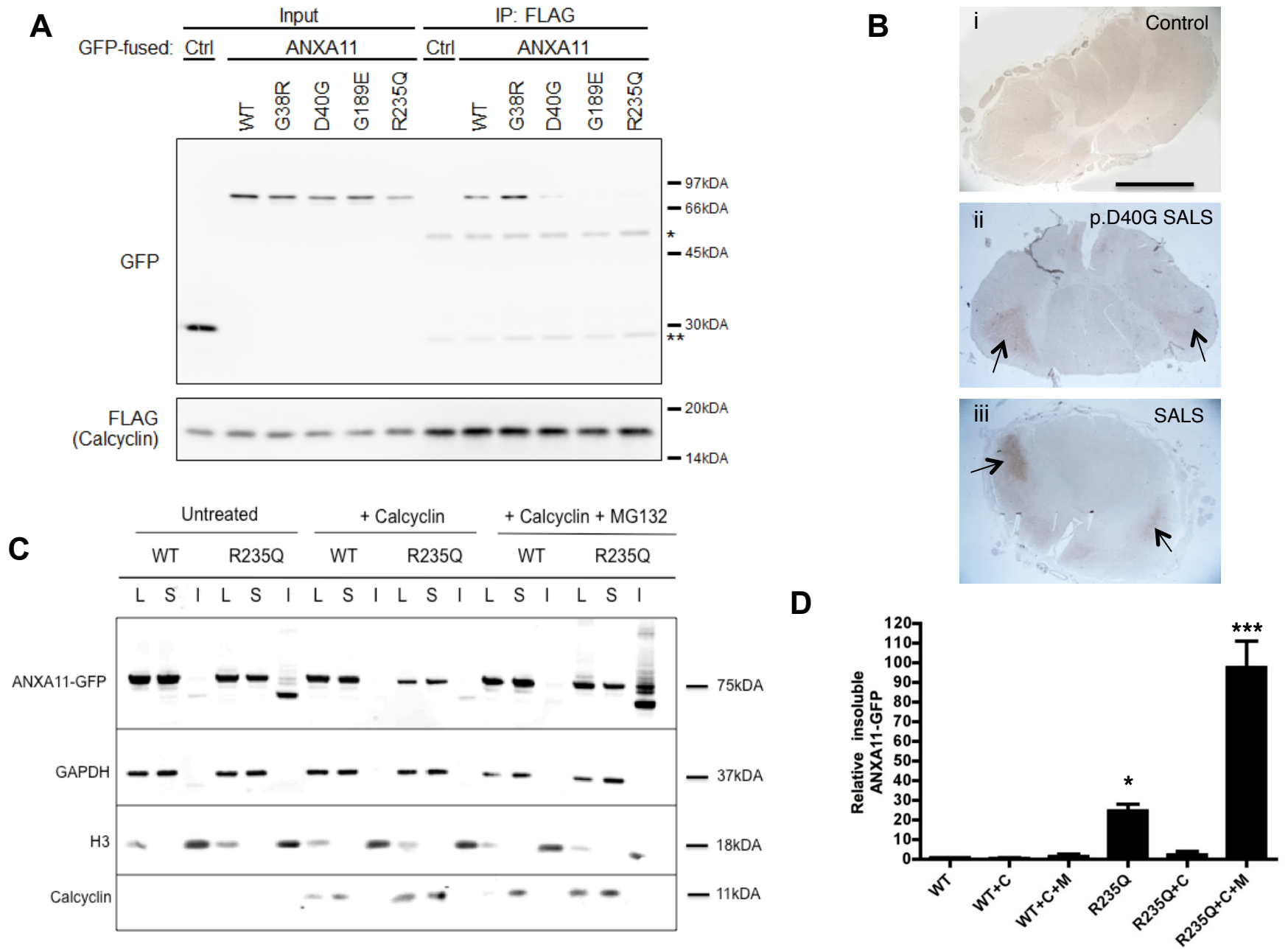


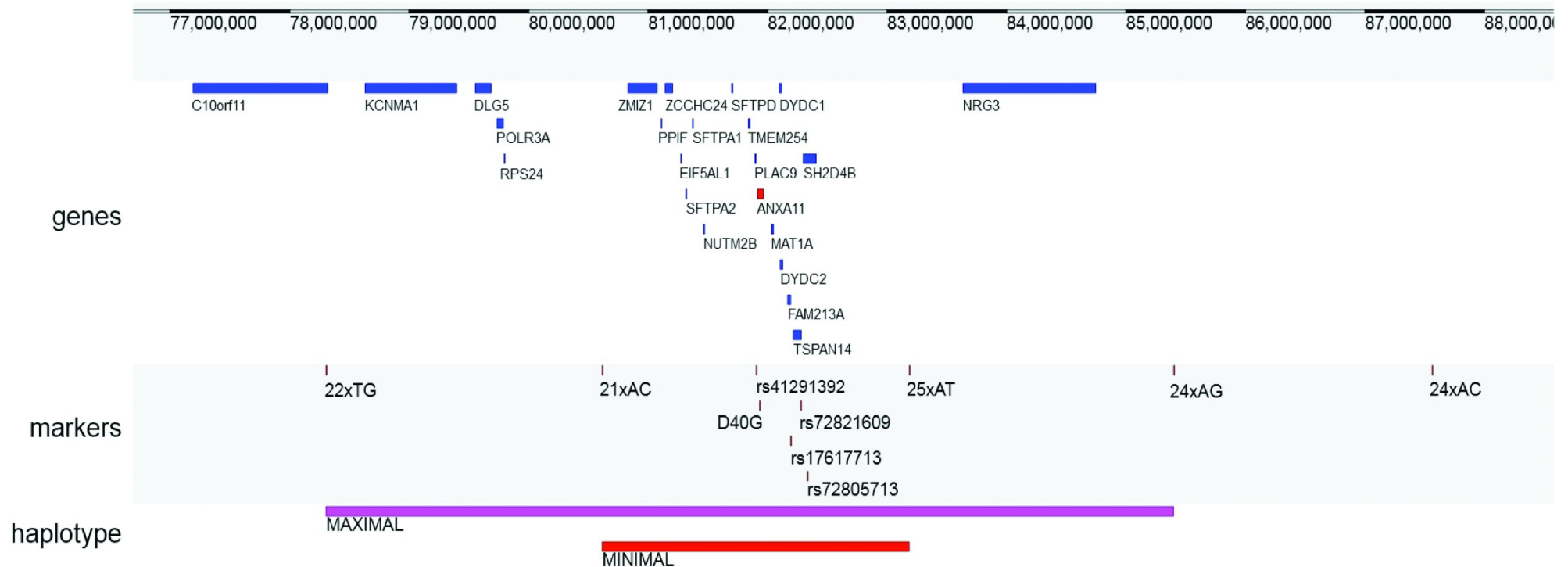
Figure 5. Calcyclin analysis: *in-vitro* binding assays with mutant ANXA11 constructs and staining of post-mortem tissue.



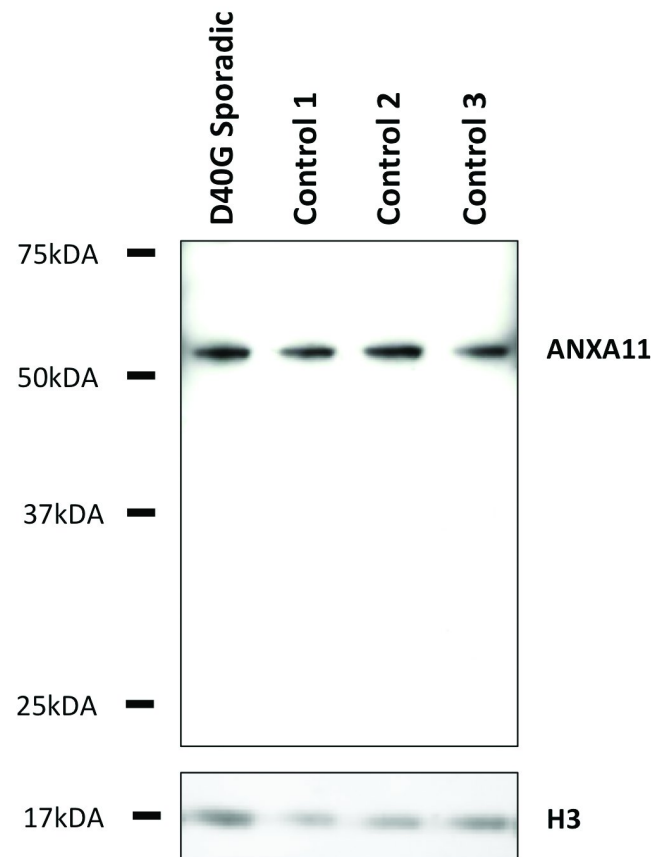
Marker	Position Hg19 (Chr10)	Gene	Allele Variant	Family 1				Family 2								Sporadic			
				III-2		III-5		III-19		III-18		IV-12		IV-13				II-2	
22xTG	78303074-78303119	NA	(TG) _n	272	292	272	290	288	274	274	278	274	274	274	274	274	274	272	290
21xAC	80615142-80615185	NA	(AC) _n	315	309	317	309	311	309	315	313	315	309	315	309	309	309	317	309
rs41291392	81904645	PLAC9	C>T	C	T	C	T	C	T	C	C	C	T	C	T	C	T	C	T
ANXA11 A>G	81930608	ANXA11	A>G	A	G	A	G	A	G	A	A	A	G	A	G	A	G	A	G
rs17617713	82191700	FAM213A	G>A	G	A	G	A	G	A	G	G	G	A	G	A	G	A	G	A
rs72821609	82276061	TSPAN14	C>T	C	T	C	T	C	T	C	C	C	T	C	T	C	T	C	T
rs72805713	82331416	SH2D4B	A>G	A	G	A	G	A	G	A	A	A	G	A	G	A	G	A	G
25xAT	83185144-83185195	NA	(AT) _n	248	250	248	250	250	250	248	240	248	250	248	250	250	250	272	250
24xAG	85399588-85399637	NA	(AG) _n	277	273	265	275	277	273	271	255	271	273	271	273	273	273	277	289
24xAC	87562595-87562642	NA	(AC) _n	278	288	276	276	278	278	276	280	276	278	278	280	280	280	278	282

Supplementary Figure 1. The *Annexin A11* p.D40G mutation shares a common haplotype comprised of 4 exonic SNPs and 2 microsatellites.

Genotyping of 5 polymorphic microsatellites spanning the *ANXA11* locus identified the 7.1MB maximal recombination region between microsatellites 22xTG and 24xAG (marked in red). The 2.5Mb minimal common region specific to all UK affected ALS D40G patients (4 FALS and 1 sporadic ALS case) was defined by two microsatellites (21xAC and 25xAT) and 4 exonic SNPs (rs41291392, rs17617713, rs72821609 and rs72805713), ie. 309-T-G-T-G-250. Genotyping of an unaffected spouse (Generation III-18) and two unaffected D40G carriers (Generation IV-12 and IV-13) from Family 2 (Figure 1C) determined haplotype phase.

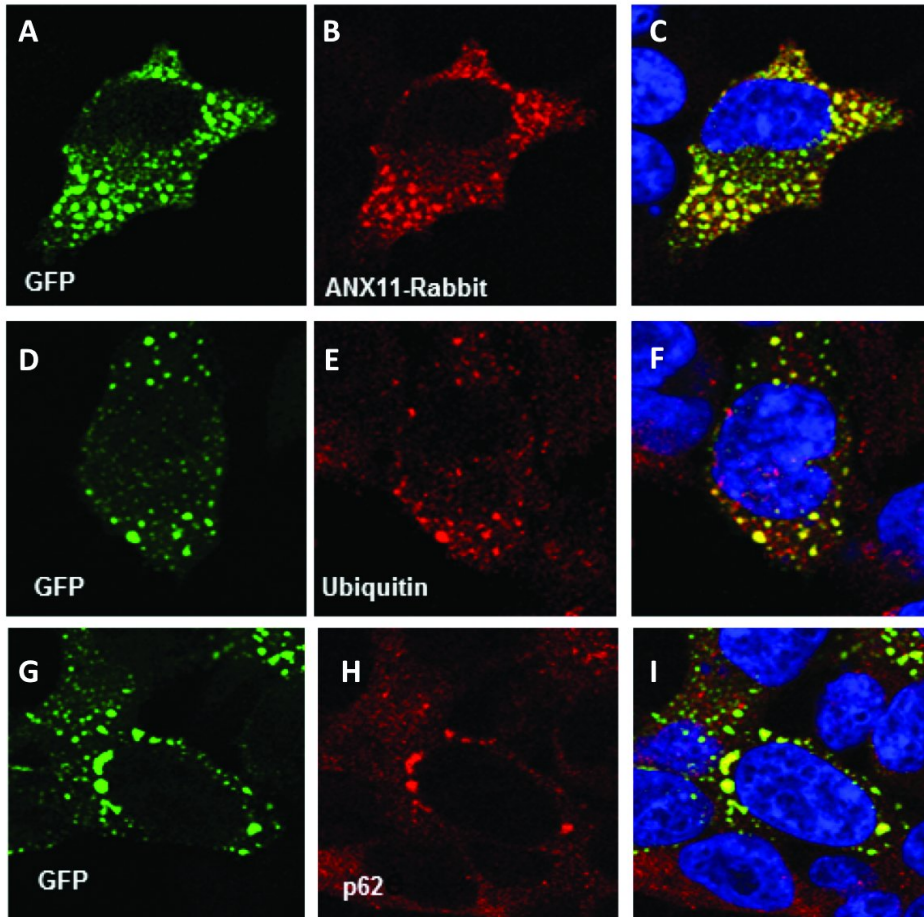


Supplementary Figure 2. Haplotype gene map of the *ANXA11* D40G locus. The maximal haplotype region (~7.1MB) is defined by the outer limit of recombination at the two flanking microsatellites 24xAG and 22xTG (pink region). The minimal common D40G haplotype block (~2.5MB) spans *ANXA11* and is shared by all 5 affected ALS UK cases and two unaffected ALS carriers. Exome sequencing of D40G carriers identified no additional novel or rare exonic or splicing variants in the 23 genes located within this maximal locus.

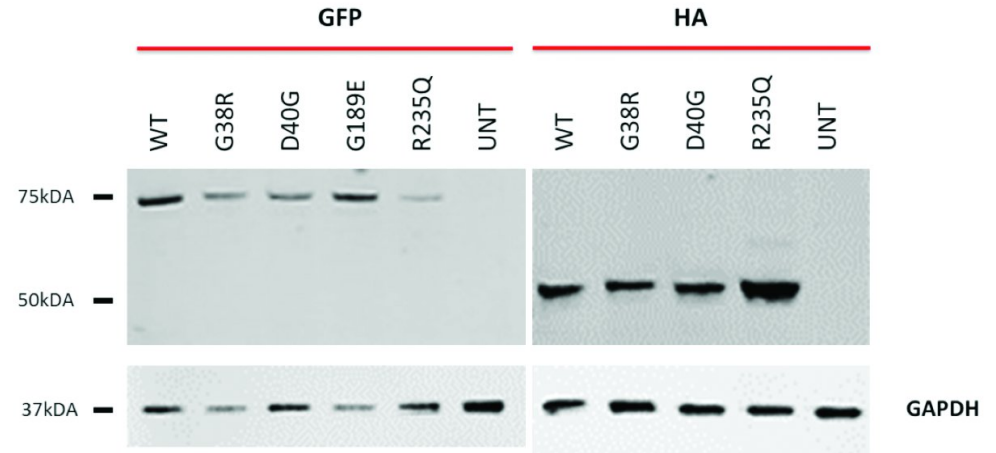


Supplementary Figure 3. Annexin A11 western blot of lysates made from post-mortem tissue of the sporadic patient harbouring the p.D40G mutation and control individuals. Western blotting of lysates made from frontal cortex tissue of the sporadic patient harbouring the D40G mutation, using a polyclonal Rabbit Annexin A11 antibody, yields a specific single band at ~55kDA. There was no difference in Annexin A11 expression due to the D40G mutation compared to lysates made from three control individuals.

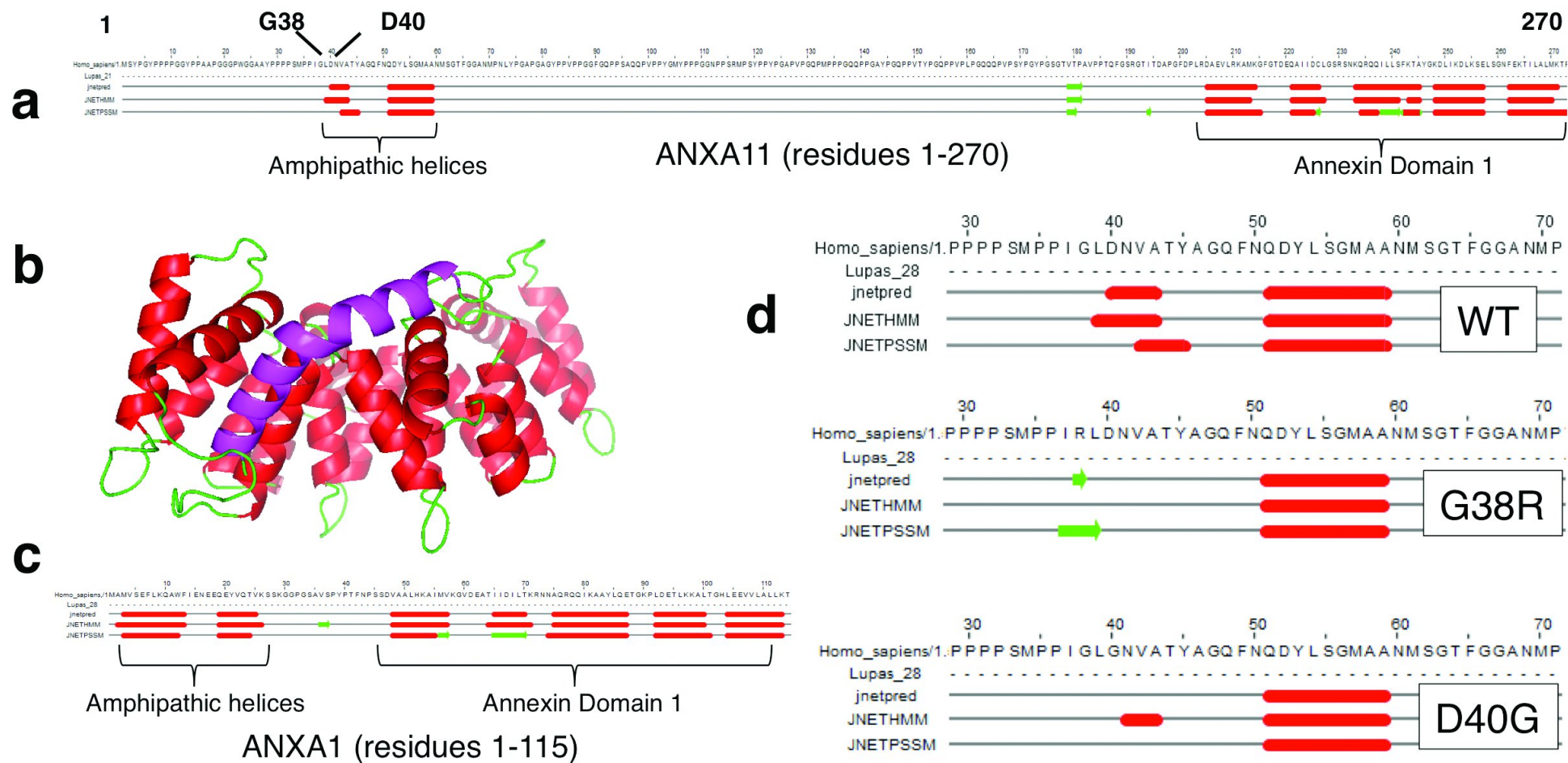
(i)



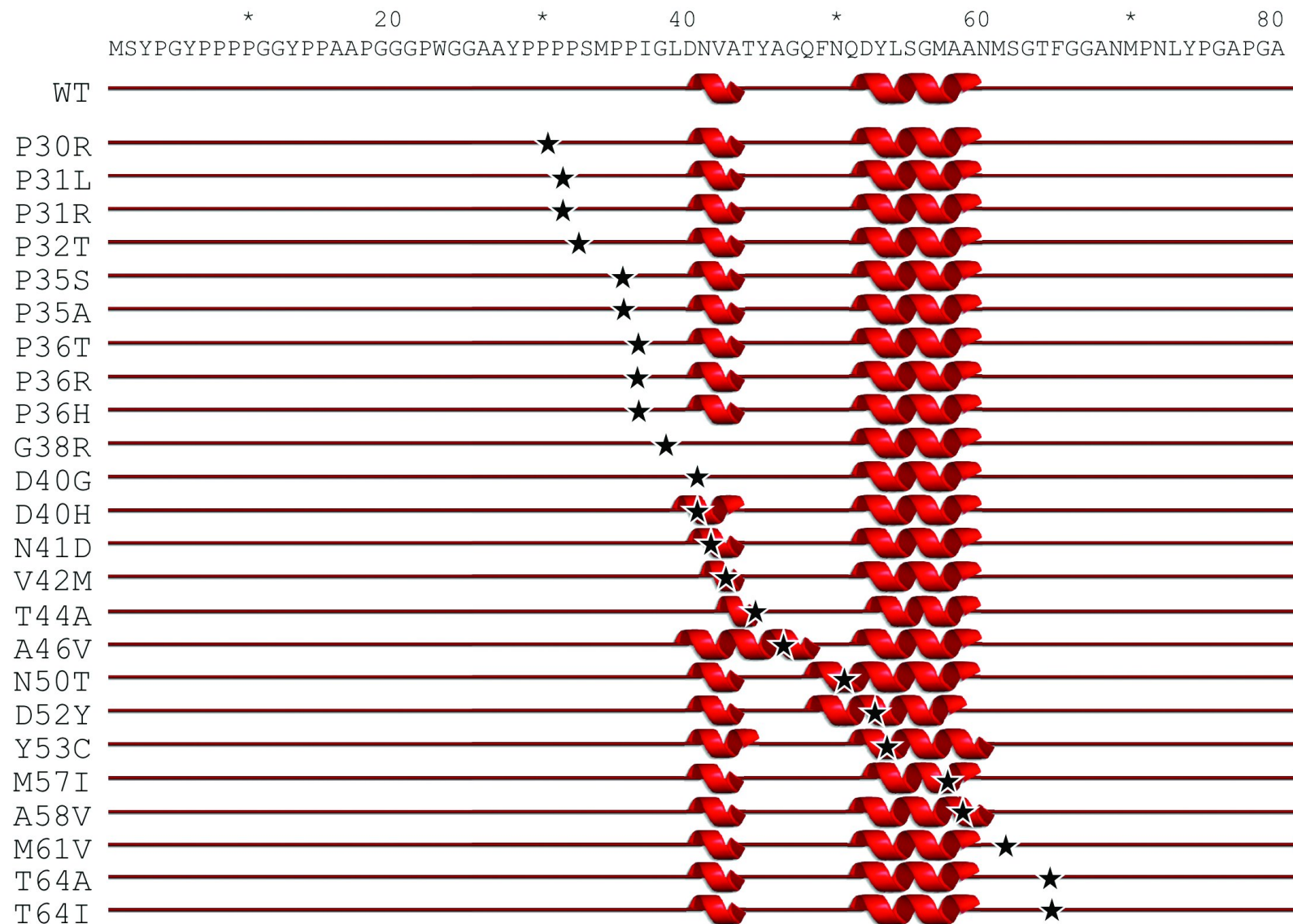
(ii)



Supplementary Figure 4. R235Q ANXA11-GFP co-localises with ubiquitin and p62 in HEK cells. (i) Cells were transfected with ANXA11-GFP^{R235Q} for 48hrs, fixed and stained with the following antibodies by immuno-cytohistochemistry (ICC) (A) mouse GFP - green. (B) ANXA11 Polyclonal rabbit Antibody (Proteintech) - red. (C) Merge of A, B and DAPI blue nuclear staining, showing specificity of ANXA11 specific antibody, (D) mouse GFP - green. (E) Rabbit K-48 Ubiquitin - red. (F) Merge of D, E and DAPI. (G) mGFP - green. (H) Rabbit p62 - red. (I) Merge of G, H and DAPI blue nuclear staining (ii) Confirmation of specificity of endogenous Rabbit ANXA11 antibody in ICC (as used in 4(i) A-C) for GFP and HA tagged ANXA11 constructs by Western blot.

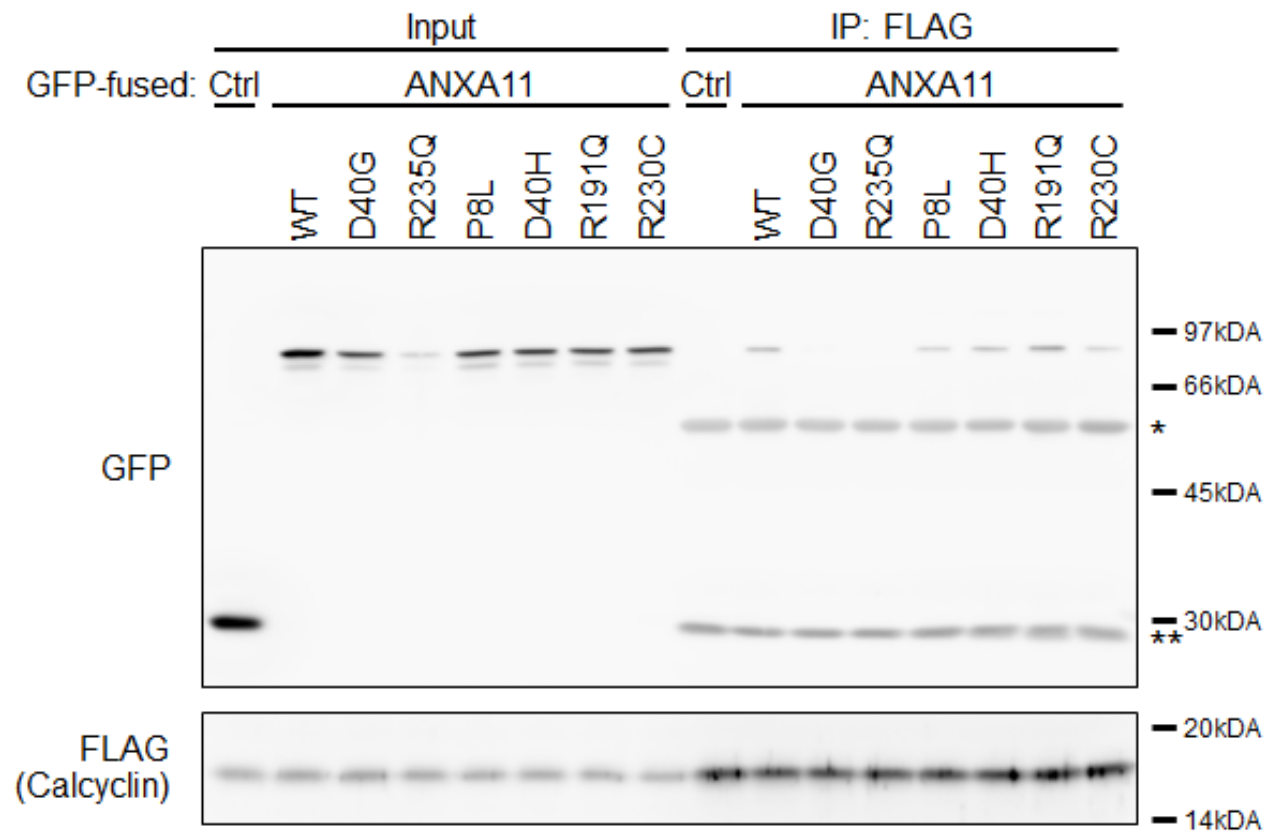


Supplementary Figure 5. Modelling of Annexin A11 identifies two amphipathic helices in the N-terminus of Annexin A11 that overlap the G38 and D40 residues. (A) Jpred secondary structure prediction of the N-terminus (residues 1-270) using a multiple alignment of 30 mammalian orthologue sequences identified two alpha helices, indicated in red (residues 40-44 and 51-59). Green represents potential beta strands. Known annexin domains start at ~205 residues. The locations of the G38 and D40 residues are indicated in the first helix. (B) Protein structure of Annexin A1 (pdb:1hm6) in the presence of calcium, with the annexin domain alpha helices in red and the N-terminal helices in magenta. (C) Jpred also predicts the N-terminus of Annexin A1 to possess two alpha helices (residues 3-13 and 19-25). (D) The Jnetpred consensus (top row) predicts that both the G38R and the D40G variants abolish the first alpha helix.



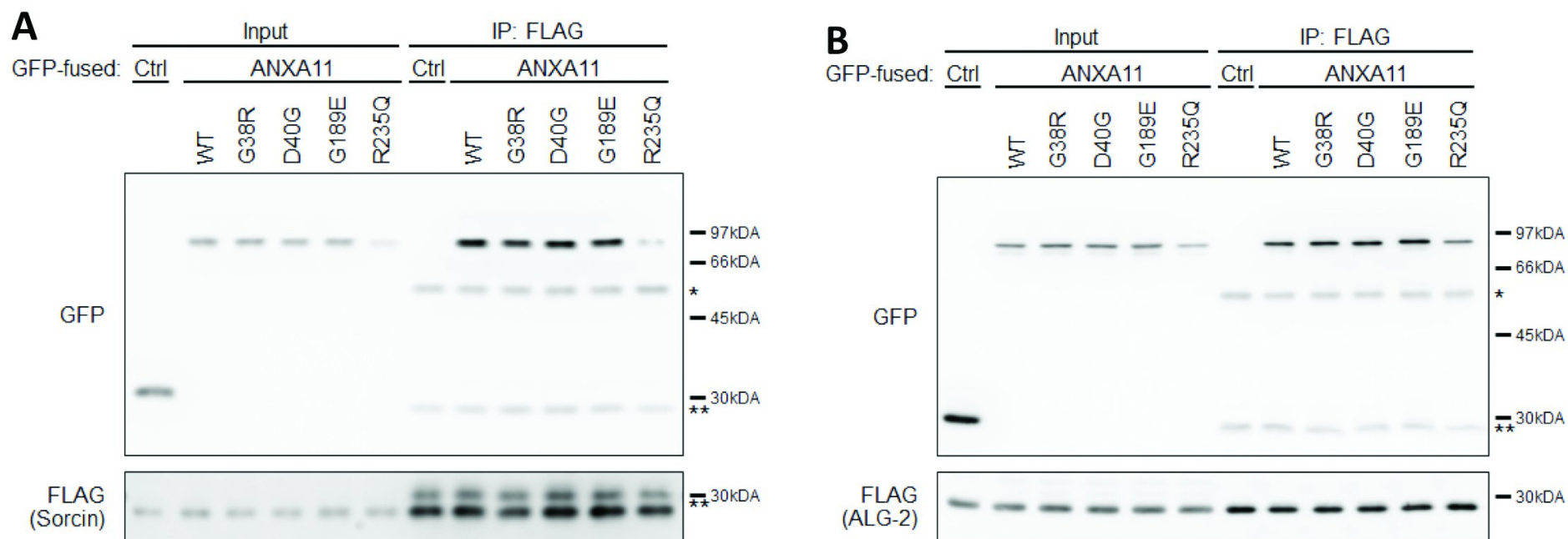
Supplementary Figure 6. Twenty two ExAC polymorphisms spanning the D40 locus do not disrupt the formation of amphipathic helices.

Secondary structure predictions on the N-terminus of Annexin 11 show that only the ALS-linked mutations G38R & D40G completely abolish the formation of the first alpha helix. None of the 22 ExAC variants flanking this region, including D40H, have the same effect. Secondary structures were derived from Jpred4 consensus, performed on a multiple alignment of 30 mammalian orthologues.



Supplementary Figure 7. Rare and common ANXA11 polymorphisms do not disrupt calcyclin binding compared to ALS specific variants.

Immunoprecipitation (IP) of FLAG-calcyclin from mixed lysates of HEK cells expressing GFP-fused proteins (as indicated) in the presence of calcium show similar binding affinities for the P8L, the D40H, the R191Q and the R230C to that of wild-type (WT) when probed for GFP (n=3). The top panel shows GFP intensities of input and pulldown fractions, with IgG heavy (*) and light chains (**) indicated. The bottom panel illustrates input and IP levels of FLAG-calcyclin.



Supplementary Figure 8. N-terminal Annexin A11 mutations do not alter binding of Sorcin and ALG2. (A) Immunoprecipitation (IP) of FLAG-Sorcini from mixed lysates of HEK cells transfected with ANXA11-GFP^{WT+G38R+D40G+G189E+R235Q} activated with calcium show no difference in binding between ANXA11 WT and mutants when probed for GFP. (B) The same was seen when IP was conducted with ALG2-sorcini. (n=3). The top panel shows GFP intensities of input and pulldown fractions, with IgG heavy and light chains indicated. The bottom panel for both (A) and (B) illustrates input and IP levels of FLAG-sorcini or FLAG-ALG2 respectively.

Figure Legends Supplementary Tables

Supplementary Table 1. Novel variants shared between all members of each Familial ALS family. Combined exome coverage and number of shared novel protein-altering variants for the 50 families in our FALS cohort. *The estimated proportion of the genome shared by all exome sequenced cases of each family was inferred from patient records or, when these were not available, calculated by Identity-By-Descent analysis performed in PLINK.

Supplementary Table 2. Novel variants shared between ANXA11 D40G Families. Filtering reduced the numbers of novel protein-altering variants in families 1 and 2 to 6 and 8 candidates, respectively.

Supplementary Table 3. Homogeneity of the FALS cohort and ExAC Non-Finnish Europeans. *An estimate of intranational genetic homogeneity, calculated as the average number of Identity-By-Descent blocks shared between two unrelated individuals [31]. This indicates that the most prevalent population in ExAC, the Swedish, are also the most homogeneous and are therefore likely to share more rare variants than the other European nationalities present in either cohort.

Supplementary Table 4. Primer sequences used in *ANXA11* D40G haplotype study and for Sanger sequencing of *ANXA11* exons. *Primers FAM labeled. Genomic locations are according to build hg19 co-ordinates.

Supplementary Table 5. Exon coverage data of genes within the minimal D40G haplotype region. Exome sequencing data achieved a read depth ≥ 10 across the minimal haplotype region for $>95\%$ of the coding bases in Family 1, $>85\%$ in Family 2 and $>70\%$ for the isolated index case. 97.9% of the bases were adequately covered by at least one sample, making the existence of an undetected, shared, coding variant highly unlikely.

Supplementary Table 6. All variants found within ANXA11 in FALS and SALS. Frequencies for each variant were assessed in dbSNP144, ESP, ExAC, 1000genomes, UK10K, and ALSdb (www.alsdb.org). An additional 909 UK and 3596 Italian exome controls were used to further assess frequencies for the p.G38R, D40G, P108L, G175R, G189E, R235Q and R346C variants. The protein-altering variants D40G, P108L, G175R & R346C are unique to ALS cases. Pathogenicity predictions are included from 20 different algorithms, obtained from Annovar (annovar.openbioinformatics.org). P108L did not segregate in an affected sibling therefore is deemed to be a private polymorphism.

Supplementary Table 7. Clinical details of patients with ANXA11 mutations. Most patients with N-terminal mutations have late disease onset (average of 72years). Patients carrying a D40G mutation have predominantly bulbar onset. There was a distinct absence of FTD symptoms in all *ANXA11* carriers. *No clinical information was available for the affected sibling of Patient 8 that carried the novel p.G175R variant.