OPEN ACCESS

Original research

Identification of a pathogenic mutation in *ARPP21* in patients with amyotrophic lateral sclerosis

Oriol Dols-Icardo ,[1,2] Álvaro Carbayo ,[3,4,5] Ivonne Jericó,[6,7] Olga Blasco-Martínez,[8] Esther Álvarez-Sánchez,[1,2] Maria Angeles López Pérez,[8] Sara Bernal,[4,9] Benjamín Rodríguez-Santiago,[4,9,10] Ivon Cusco,[9] Janina Turon-Sans,[3,4,5] Manuel Cabezas-Torres,[3,4] Marta Caballero-Ávila ,[3,4,5] Ana Vesperinas,[3,4,5] Laura Llansó,[3,4,5] Inmaculada Pagola-Lorz,[6,7] Laura Torné,[6,7] Natalia Valle-Tamayo,[1,2] Laia Muñoz,[1,2] Sara Rubio-Guerra ,[1,2] Ignacio Illán-Gala ,[1,2] Elena Cortés-Vicente ,[3,4,5] Ellen Gelpi,[11] Ricard Rojas-García [3,4,5]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jnnp-2024-333834).

For numbered affiliations see end of article.

Correspondence to
Dr Ricard Rojas-García, Neurology, Hospital de la Santa Creu i Sant Pau, Unidad de Enfermedades Neuromusculares, Barcelona, Spain; rrojas@santpau.cat
Dr Oriol Dols-Icardo; odols@santpau.cat

OD-I and ÁC contributed equally.

Received 18 March 2024
Accepted 17 June 2024

To cite: Dols-Icardo O, Carbayo Á, Jericó I, *et al*. *J Neurol Neurosurg Psychiatry* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/jnnp-2024-333834

ABSTRACT

Background and objective Between 5% and 10% of amyotrophic lateral sclerosis (ALS) cases have a family history of the disease, 30% of which do not have an identifiable underlying genetic cause after a comprehensive study of the known ALS-related genes. Based on a significantly increased incidence of ALS in a small geographical region from Spain, the aim of this work was to identify novel ALS-related genes in ALS cases with negative genetic testing.

Methods We detected an increased incidence of both sporadic and, especially, familial ALS cases in a small region from Spain compared with available demographic and epidemiological data. We performed whole genome sequencing in a group of 12 patients with ALS (5 of them familial) from this unique area. We expanded the study to include affected family members and additional cases from a wider surrounding region.

Results We identified a shared missense mutation (c.1586C>T; p.Pro529Leu) in the cyclic AMP regulated phosphoprotein 21 (*ARPP21)* gene that encodes an RNA-binding protein, in a total of 10 patients with ALS from 7 unrelated families. No mutations were found in other ALS-causing genes.

Conclusions While previous studies have dismissed a causal role of *ARPP21* in ALS, our results strongly support *ARPP21* as a novel ALS-causing gene.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Between 5% and 10% of amyotrophic lateral sclerosis (ALS) cases have a family history of the disease. 30% of familial ALS cases do not have an identifiable underlying genetic cause after a comprehensive study of the known ALS-related genes. Incidence rates of ALS are fairly homogeneous across European and American populations.

WHAT THIS STUDY ADDS

⇒ This work supports the causative role of cyclic AMP regulated phosphoprotein 21 (*ARPP21)* in ALS and adds to the extensive list of RNA-binding proteins involved in ALS pathophysiology. Further studies are required to elucidate the contribution and frequency of this variant in *ARPP21* to the genetic basis of the disease. Peaks of incidence of the disease in small geographical areas may help identify the underlying genetic causes of ALS.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ *ARPP21* should be reclassified as an ALS-causing gene, and included in genetic screenings along with other known genes.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a devastating motor neuron disease, characterised by muscle weakness and atrophy secondary to upper and lower motor neuron degeneration, leading to respiratory insufficiency and death within a mean of 3–5 years.[1] Over the last decades, an evident overlap of clinical, genetic and neuropathological features has been established between ALS and frontotemporal dementia (FTD).[2–5]

Although most cases are considered sporadic (sALS), approximately 5–10% of ALS cases have a previous family history of ALS and are classified as familial ALS (fALS).[6] To date, over 40 genes have been related to ALS, most showing a Mendelian autosomal dominant inheritance pattern. Several ALS-causing mutations have an impact on genes encoding RNA-binding proteins, such as transactive response DNA-binding protein 43 (*TDP-43*), fused in sarcoma (*FUS*) and T cell-restricted intracellular antigen-1 (*TIA1*), among others, highlighting the central role of RNA metabolism in the disease.[7] However, despite the rapidly growing number of ALS-related genes, genetic diagnosis remains elusive in approximately 30% of patients with fALS.[8 9] The disease has a complex genetic architecture, and oligogenic and polygenic models of inheritance have gained relevance.[10] In this era of protein-targeted therapies and genetically oriented treatments, discovery and detection of new genetic variants and implicated genes is paramount.[10]

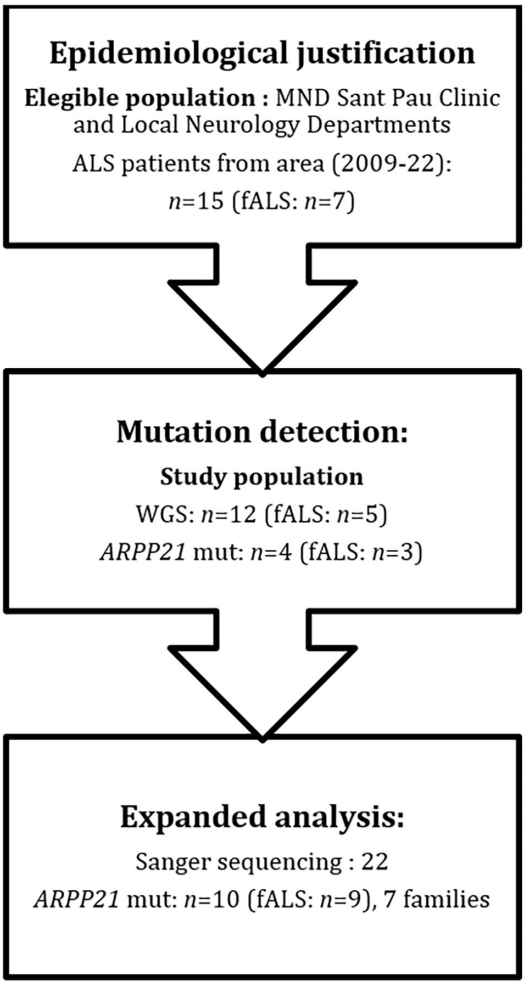

Figure 1 Flowchart indicating the steps followed in the study and the number of patients at each phase. ALS, amyotrophic lateral sclerosis; ARPP21, cyclic AMP regulated phosphoprotein 21; fALS, familial ALS; MND, motor neuron disease; WGS, whole genome sequencing.

Prospective, population-based studies show that incidence rates of ALS are relatively uniform across European and American populations, ranging between two and three cases per year per 100 000 inhabitants.[11 12] Peaks of incidence in small regions, especially when several cases aggregate in a family, might suggest the disorder has an underlying genetic substrate.[13]

Here we report the identification of a mutation in the cyclic AMP-regulated phosphoprotein 21 (*ARPP21*) gene, supporting its role as a novel ALS-causing gene and expanding the repertoire of RNA-binding proteins associated with ALS. We describe the clinical phenotype of 10 patients from 7 unrelated families carrying the same mutation in *ARPP21*.

MATERIALS AND METHODS

Case selection and study design

Since 1998 our register has prospectively included cases of ALS seen at the Motor Neuron Clinic at the Neuromuscular Diseases Unit of Hospital Sant Pau (Barcelona), a national referral centre for neuromuscular diseases. Between 2009 and 2022 we observed an unusual number of cases of ALS (especially familial cases) at our outpatient clinic, from a small region in the province of La Rioja in Spain. Figure 1 illustrates the study design.

For epidemiological justification (figure 1 – box 1) we reviewed our database and contacted neurology departments from local centres in the region. Eligible population: We selected new cases with diagnosis of ALS from the region, visited between January 2009 and December 2022.

Study population: For genetic analysis (figure 1 – box 2) we selected cases from the study area (see figure 2) with available DNA samples, performing whole genome sequencing (WGS). Patients fulfilled ALS-El Escorial revised criteria for probable, probable laboratory-supported, or definite ALS.[14] Both sporadic and fALS cases were included. fALS was defined as having a family member with a diagnosis of ALS in first or second-degree relatives, or having a first-degree family member with a confirmed FTD in the formal neuropsychological examination, according to current proposed criteria.[6] All patients went through a targeted ALS-gene panel prior to WGS (online supplemental table 1 in supplementary materials). Patients with a previously identified ALS-causing mutation in this custom panel were excluded from this part of the study (WGS analysis), as they already had a known underlying genetic substrate for the disease.

After identifying the candidate variant, we expanded the analysis with the aim of increasing the series of mutation carriers. We performed Sanger sequencing in other cases of sALS and fALS from the surrounding geographical areas and also in their affected family members (figure 1 – box 3). In cases carrying

Figure 2 (A) Map of Spain and its regions. (B) Map of La Rioja province and its regions. Regions belonging to the south-eastern area of La Rioja, where an increased frequency of amyotrophic lateral sclerosis cases was detected, are shaded in green. White dots mark the location on the map of the six towns of origin of the seven families.

the candidate variant, mutations in other known ALS-FTD genes were ruled out through exome sequencing (see online supplemental table 1 in supplementary materials).

Demography and epidemiology

The expected number of cases in the area during the study period was calculated using incidence rates from previous studies and demographic data of those in the catchment area over 18 years.

Demographic information on the geographical region was obtained from the official demographic records between 2009 and 2022 in the Spanish National Institute of Statistics (INE—Instituto Nacional de Estadística).[15 16]

The minimum incidence was calculated using the numbers of new cases identified from the area and the population at risk in the study period. We compared the observed cases with the expected number of cases in the area, to obtain a ratio.

Clinical features

We recorded demographic and clinical variables. Demographic data included sex, date and place of birth and place of origin of the family. Clinical variables included age at disease onset, disease duration, region of onset of motor symptoms (spinal or bulbar). The clinical phenotype was categorised depending on the presence of upper (primary lateral sclerosis), lower motor neuron signs (progressive muscular atrophy) or both (ALS). The presence of cognitive and/or behavioural impairment and whether or not they met clinical criteria for FTD and its variants, were also noted.[17–19]

Family history

We recorded previous cases of ALS in the family as part of the routine clinical evaluation. As possible indicators of an underlying FTD we also recorded previous history of FTD, early-onset cognitive impairment (defined as younger than 65 years of age at symptom onset), cognitive impairment with marked behavioural impairment but no definite diagnosis, and diagnosis of other neurodegenerative diseases.

Genetic analyses

Library preparation and sequencing

WGS was performed at the National Centre for Genomic Analysis (Centro Nacional de Análisis Genómico—CNAG-CRG, Barcelona, Spain). Paired-end multiplex libraries were prepared following the manufacturer's instructions. Libraries were loaded to Illumina flow cells for cluster generation prior to producing 150 base read pairs on a NovaSeq 6000 instrument following the Illumina protocol. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis and followed by generation of FASTQ sequence files.

Bioinformatic and genetic analyses

Reads were mapped to human build GRCh38 with BWA-MEM V.0.7.17.[20] Alignment files containing only properly paired, unique mapping reads without duplicates were processed using Picard V.2.20 (http://broadinstitute.github.io/picard/) to add read groups and to remove duplicates. The Genome Analysis Tool Kit (GATK V.4.1.9.0)[21] was used for local realignment and base quality score recalibration. Joint variant calling was done using HaplotypeCaller, CombineGVCFs and GenotypeGVCFs from GATK following the GATK best practices pipeline. Variants were hard filtered using GATK VariantFiltration (BaseQRankSum>4.0 || BaseQRankSum<−4.0, FS>60,000, FS>200,000, ReadPosRankSum<−8.0, ReadPosRankSum>20.0, MQRankSum<−12.5, QD<2.0, MQ<40.0). We also retained variants where at least one sample contained the variant with a Depth ≥10.

Functional annotations were added using SnpEff V.5.0 with the hg38.[22] Variants were annotated with SnpSift V.5.0[23] using population frequencies, conservation scores and deleteriousness predictions from dbNSFP V.4.1a.[24] We also used other sources of annotations, such as gnomAD (V.3.1.2), CADD (V.1.6), InterVar and Clinvar (V.20200602).[25–27] In order to check the frequency of the *ARPP21* variant in the Spanish population, we used the Collaborative Spanish Variant Server which comprises exomes and genomes from 2105 unrelated Spanish individuals.[28]

Somalier software[29] was used with WGS and whole exome sequencing (WES) data to assess relatedness among all probands included in the study (coefficient of relationship as a measure of the degree of consanguinity). Briefly, the software considers for each comparison: (1) the number of sites where one sample is hom-ref and another is hom-alt; (2) the number of sites where the samples have the same genotype; (3) the number of sites where both samples are heterozygotes; and (4) the number of sites where both samples are homozygous alternate.

WGS data was used to identify shared genotypes that might suggest a common haplotype in the four mutation carriers with available WGS data. A merged VCF file of *ARPP21* gene ±1 megabase combining variants from the four patients was obtained to explore shared blocks of genotypes. LDhap web-based tool (https://ldlink.nih.gov/?tab=ldhap) was used to obtain specific population haplotype frequencies of all haplotypes observed for a list of 30 query genetic variants (the maximum accepted as input) flanking the shared *ARPP21* candidate variant and part of the shared haplotype (23 of them were available in LDhap). We selected the Iberian population in Spain (IBS) population to obtain allelic frequencies, in addition to the other populations in Europe and all other super populations available (African (AFR), American (AMR), East Asian (EAS) and South Asian (SAS)).

The presence of the disease-causing mutation in *ARPP21* was validated using Sanger sequencing in all mutation carriers using the following forward (5' GACAGTGGTGTGCATCTTGTG 3') and reverse (3' GAGACATTCGCACATACCCC 5') primers.

Once the *ARPP21* variant was confirmed, other ALS and FTD causing genes were screened using WES and mutations in 58 ALS-FTD-related genes were ruled out (see online supplemental table 1 in supplementary materials for the complete list of genes). WES was used to study gene coding exons and flanking regions (±50 base pairs). Briefly, DNA libraries were prepared using kapa reagents (Roche) and sequenced in NextSeq 500 or NextSeq 1000 instruments (Illumina) to obtain 2×150 base pair reads. Fastq files were mapped to the human genome using BWA,[30] variants were called through GATK variant calling,[21] and annotated using ANNOVAR.[31] The *C9orf72* hexanucleotide repeat expansion was discarded by means of rpPCR (repeat-primed Polymerase Chain Reaction), as previously described.[32]

Ethics

The study was conducted in accordance with the Declaration of Helsinki. All participants gave their written informed consent to participate in the study.

RESULTS

Epidemiology

The south-eastern region of the province of La Rioja is a 1219.42 km^2 area in north-eastern Spain (figure 2). Between

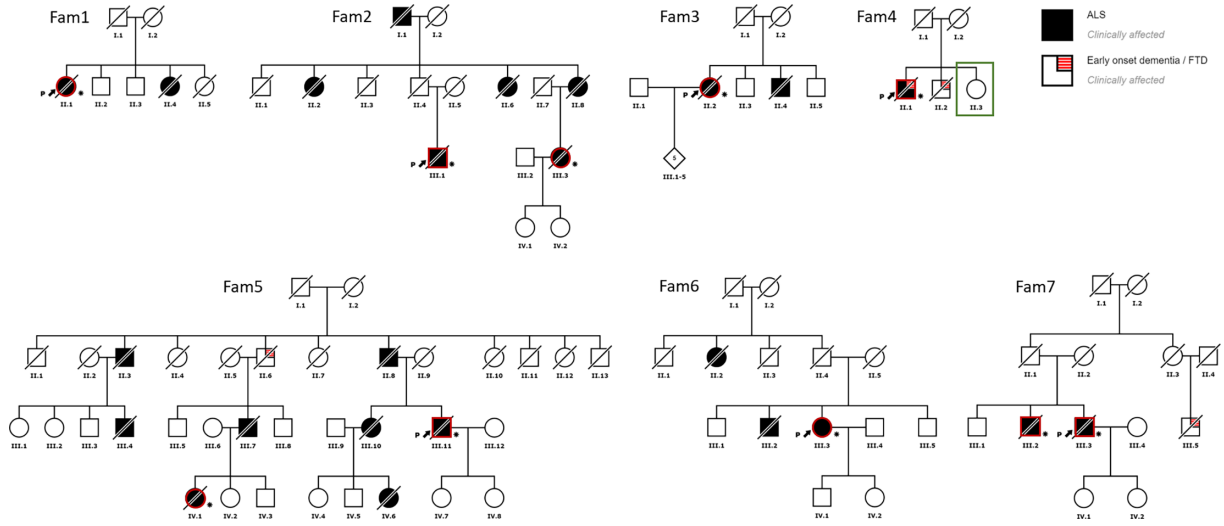

Figure 3 Family trees of the seven families in which the p.Pro529Leu mutation in *ARPP21* was detected. Patients with ALS are shaded in black. Cases presenting an early onset dementia or dementia with prominent behavioural features are marked with a red square at the top right corner. The proband in each family is marked with an arrow and a 'P'. Cases that harbour the p.Pro529Leu mutation in *ARPP21* are outlined in red and marked with an asterisk. The healthy individual (Family 4— II.3) with negative genetic testing is framed in a green box. No DNA was available from the other patients with ALS or patients with early-onset cognitive-behavioural impairment in the pedigrees. ALS, amyotrophic lateral sclerosis; ARPP21, cyclic AMP regulated phosphoprotein 21; FTD, frontotemporal dementia.

2009 and 2022, it had a mean population of 43 433, of whom 31 324 were over 18 years of age. The population density was 35.62 persons/km^2. Figure 1 illustrates the study design and flowchart of patients studied in each phase.

Considering a mean ALS incidence of 1.4–2.47 cases/100 000 person/year,[11 12] we calculated an expected number of cases of 0.44–0.77 cases per year in this area, equivalent to 5–10 patients during the study period (2009–2022). In reference to fALS, assuming a frequency of 5–10% of fALS, the expected number of cases in the area would be 0.02–0.08 cases/year or one new case every 12.5–50 years.

Between 2009 and 2022, we visited 15 patients from the study area who met the diagnostic criteria for ALS. 7 of the 15 (46.6%) had a family history of ALS and were considered as fALS. In fALS cases, known ALS-causative mutations were ruled out through exome sequencing analysis or a custom gene panel.

The comparison between the observed frequency and the expected frequency resulted in an increased frequency of ALS cases in the area, with a ratio of 1.5–2.6 for total ALS cases and 7–24.5 for fALS cases.

12 of the 15 patients were eligible for further genomic analysis. One patient was excluded from this part of the study because an alternative genetic cause was found (non-sense pathogenic mutation in *NEK1*—c.3107C>G; p.Ser1036X—NM_001199397.3). No DNA was available for analysis in the remaining two patients.

Genetics

Detection of the *ARPP21* mutation

WGS was performed in 12 unrelated patients (5 fALS and 7 sALS). We found a shared heterozygous missense mutation in *ARPP21* (c.1586C>T; p.Pro529Leu (NM_001385486.1 or ENST00000417925.5)) in four unrelated patients with ALS (three fALS) from four different families (figure 3—Families 1–4) (table 1). The high frequency of this almost unique variant, together with its predicted deleteriousness (*CADD_phred=26.7; Polyphen HDIV and HVAR=Deleterious and Mutation Taster=Deleterious*) and high GERP (Genomic Evolutionary Rate Profiling) score (6.07), which predicted a huge evolutionary constraint, led us to prioritise this variant in *ARPP21* as a

Table 1 Clinical features of patients with ALS carrying the ARPP21 mutation

	Family	Patient	Region of onset	Motor phenotype	Cognitive symptoms	Disease duration	Study
1	Fam 1	II.1*	Spinal	ALS	No	13	WGS
2	Fam 2	III.3	Spinal	ALS	No	–	WES
3	Fam 2	III.1*	Spinal	ALS	No	16	WGS
4	Fam 3	II.2*	Spinal	ALS	No	–	WGS
5	Fam 4	II.1*	Bulbar	ALS-FTD	bvFTD	31	WGS
6	Fam 5	III.11*	Bulbar	ALS	No	16	WES
7	Fam 5	IV.1	Spinal	ALS	No	15	WES
8	Fam 6	III.3*	Spinal	ALS	No	–	WES
9	Fam 7	III.2	Spinal	ALS	No	17	WES
10	Fam 7	III.3*	Spinal	ALS	No	4	WES

Probands of each family are marked with an asterisk (*). 'Disease duration' is expressed in months (M).
ALS, amyotrophic lateral sclerosis; ARPP21, cyclic AMP regulated phosphoprotein 21; bvFTD, behavioural variant of FTD; FTD, frontotemporal dementia; MND, motor neuron disease; WES, whole exome sequencing; WGS, whole genome sequencing.

potential causative mutation. The coefficients of the relationship obtained using Somalier suggest that the probands with available WGS and WES data are unrelated (data not shown). This variant is absent in the population database gnomAD (V.3.1.2) and carried by 4 out of 1 384 130 chromosomes in the latest updated version of gnomAD (V.4.0.0), where the mutation is annotated as p.Pro563Leu (NM_001385562.1 or ENST00000684406.1). We did not find any mutation carrier among the 2105 unrelated Spanish individuals included in the Collaborative Spanish Variant Server.[28]

Our analysis disclosed a common haplotype shared by the four mutation carriers with available WGS data, which included 442 SNPs (Single Nucleotide Polymorphisms), encompassing 256 Kb (chr3:35605944–35862077). This haplotype was also found in WGS data from two of the eight non-mutation carriers included in this study. We further assessed its frequency in the general population using LDhap. The frequency of the haplotype in the IBS was 7.48%, the same as in the Toscani in Italy. Interestingly, this haplotype was observed at a higher frequency in the British in England and Scotland (GBR) population (11.5%) or at a lower frequency in the Finnish in Finland population (5.6%). The haplotype was almost absent in AFR or EAS populations (<1%), and at low frequencies in the ad mixed AMR population (2.7%). Notably, 9% of the SAS population carried this haplotype.

Expanded study in the region

We subsequently broadened the analysis to affected family members of mutation carriers (hence new fALS cases) and other new patients with sALS and fALS from a wider surrounding geographical area.

Using Sanger sequencing we analysed 22 new patients with ALS, 8 of whom were fALS. We found the candidate variant in 6 novel individuals: 1 affected relative in Family 2, and 5 more cases from 4 unrelated families, resulting in a final count of 10 *ARPP21* mutation carriers from 7 unrelated families (see pedigrees in figure 3). We were able to study only one unaffected family member (Family 4—II.3), who did not harbour the candidate variant. We did not find any patient with ALS among the families who did not carry the candidate variant in *ARPP21*.

Other ALS-FTD-related mutations were ruled out through WES, including the p.Arg92Cys variant in *GLT8D1* (see online supplemental table 1). No other rare variants were found in the *GLT8D1* gene in *ARPP21* mutation carriers. In the proband of Family 5 (III.11), we detected a heterozygous variant in *SQSTM1* (c.1175C>T; p.Pro392Leu), for which there is no consensus regarding its pathogenicity for ALS-FTD. The rare variant in *SQSTM1* was not carried by his affected family member (Family 5—IV.1) who also carried the mutation in *ARPP21* described in this study, thus demonstrating that did not segregate with disease and making any potential association with ALS unlikely.

Clinical features and family history

Table 1 shows the main clinical features of the series. Five patients were men (50%). The mean age at onset was 59.5 years (SD 11.37) and the mean disease duration was 16 months (SD 7.96). At the date of publication, three patients (30%) were alive. The region of onset was bulbar in two patients (20%) and spinal in the others (n=8; 80%). The clinical phenotype was classic ALS (ie, presence of upper and lower motor neuron signs) in all 10 patients (100%), with no cases of progressive muscular atrophy or primary lateral sclerosis in the series. Cognitive-behavioural impairment was present in one patient (10%).

Regarding family history, only one patient did not have a family history of ALS, but had a younger sibling with an early onset dementia (Family 4 in figure 3). Moreover, two other families (Families 5 and 7 in figure 3) had a family history of early-onset dementia.

Pedigrees and case descriptions

Table 1 summarises the main clinical features of the series. Figure 3 shows the family trees.

Family 1

Proband (II.1) presented with progressive paraparesis in their late 40s, with diffuse upper motor neuron signs. Family history was notable for a younger sibling (II.4) who had died from ALS in their 30s, and was hence diagnosed with fALS, with initial negative genetic testing, dying of respiratory insufficiency within 13 months from symptom onset. No clinical information was available concerning the proband's parents.

Family 2

Proband (III.1) was a patient in their late 30s with a family history of ALS in the grandfather on the father's side (I.1) and in three of his aunts (II.2; II.6; II.8). The patient presented with spinal onset ALS and died within 16 months from symptom onset. Conventional genetic testing was negative but the heterozygous variant in *ARPP21* was identified. A cousin (III.3) was diagnosed with ALS and carries the *ARPP21* variant.

Family 3

Proband (II.2) is an adult patient with no relevant medical history, but a family history of a brother (II.4) who died of spinal onset ALS. There were no medical records available for their parents. The patient was diagnosed with ALS in their late 60s following progressive distal upper and later lower limb weakness. Despite the slow progression of the disease, neurophysiological testing confirmed a motor neuron disease. At the time of publication of the present work, the patient was alive.

Family 4

Proband (II.1) was a patient in their 60s who presented with a sporadic, bulbar onset ALS, with a disease duration of 3 years. The patient showed profuse behavioural and cognitive impairment, suggestive of an ALS-FTD, but no formal neuropsychological testing was available.

A younger brother (II.2) had been diagnosed with an early-onset dementia before 65 years of age, but no formal cognitive-behavioural evaluation was performed. Both their parents had died from unknown causes before 60 years of age. No further detailed information was available on their family. The patient's younger sibling (II.3), older than 70 years of age, who showed no signs or symptoms suggestive of motor neuron disease, was genetically tested and did not harbour the *ARPP21* mutation.

Family 5

Proband (III.11) was an elderly patient with an unremarkable personal medical history. This patient presented with a bulbar onset motor neuron disease with progressive limb weakness and respiratory failure, leading to death within 16 months.

There was a profuse history of ALS in the family (see family tree: II.3, II.8, III.4, III.7, III.10, IV.1, IV.6). Furthermore, an uncle of the proband (II.6), whose son and granddaughter died of ALS and was hence an obligate carrier, was diagnosed with unspecified dementia and died before 50 years of age. After the

ARPP21 mutation was detected in the proband, the only affected family member (IV.1) with available DNA was retrospectively confirmed to harbour the same mutation.

Family 6

Proband (III.3) was a patient in their late 60s, with a family history of two cases of ALS (III.2, II.2). The proband developed progressive paraparesis evolving to upper limb weakness and dysphagia, and neurophysiological examinations compatible with ALS. At the time of submission of this work, the patient was alive.

Family 7

Proband (III.3) was an adult in their 60s who presented with a rapidly progressive tetraparesis, followed by respiratory insufficiency and death within the first 4 months of disease onset. The patient's brother (III.2) died of ALS 10 years prior to the proband's diagnosis, and was retrospectively confirmed to carry the *ARPP21* mutation. A cousin (III.5) was also diagnosed with early-onset dementia before 60 years of age. Their father died of pancreatic cancer in his early 70s.

DISCUSSION

Following an increased frequency of cases in the small region of the south-eastern La Rioja province in Spain, we identified the p.Pro529Leu mutation in *ARPP21* in 10 patients from 7 unrelated families, showing a presumably autosomal dominant pattern. Familial cases (ie, fALS) accounted for almost 50% of the total cases from the area, none of whom harboured any other known ALS-causing mutations.

From a clinical point of view, patients carrying the p.Pro529Leu variant presented a shorter survival (mean 16 months) than that described in the literature for ALS,[12] with no differential clinical features. Although no systematic cognitive assessment was performed, one patient showed prominent cognitive and behavioural impairment. Furthermore, three relatives (one of them an obligate carrier— Family 5 II.6) from different families had been diagnosed with early-onset dementia with marked behavioural symptoms, but no apparent motor neuron disease. It cannot be ruled out that the *ARPP21* mutation could also be related to FTD, and clinical presentation at any point might fall within the ALS-FTD spectrum (ie, ALS, FTD or ALS-FTD), as has been observed in other ALS-FTD-related genes.[33] Unfortunately, no formal cognitive testing, neuroimaging or DNA samples were available from these patients. Further studies will be required to confirm this hypothesis.

In 2019, Cooper-Knock *et al*[34] described the segregation of a mutation (p.Arg92Cys) in the glycosyltransferase 8 domain-containing protein 1 (*GLT8D1*) gene and the mutation described in the present study (p.Pro529Leu mutation in *ARPP21*) in an autosomal dominant ALS pedigree from the UK. Additional screening of both genes in 103 ALS cases disclosed 4 new *GLT8D1* mutation carriers, 3 of which also harboured the p.Pro529Leu mutation in *ARPP21*. However, after a rare variant burden analysis in the Project MinE data set, the authors prioritised mutations in the glycosyltransferase domain of *GLT8D1* as causative, suggesting only a secondary and putative synergistic effect for the mutation in *ARPP21* in double-mutation carriers. This conclusion was based on an observed shorter disease duration in patients carrying both mutations (<16 months),[34] similar to that of patients in our series who carry the p.Pro529Leu variant in isolation. Therefore, we show that the mutation described in our study is sufficient to cause ALS, and suggest that the shorter disease duration might be independently attributed to the p.Pro529Leu mutation in *ARPP21*.

Interestingly, the four mutation carriers identified through our WGS approach shared a common haplotype. This haplotype was present in two of the eight non-mutation carriers in our study with available WGS data, and it is also found in other worldwide populations. In this context, beyond the Spanish population, its high frequency in the GBR and SAS populations, together with the fact that this haplotype is found at very low frequencies in AMR, AFR or EAS populations, suggests that people carrying this mutation might be restricted to specific ancestries. In this sense, the *ARPP21* mutation, although not definitely linked to ALS, has previously been detected in patients of European ancestry, primarily from UK,[34] which is part of the GBR population. Indeed, the haplotype is almost null in the EAS population, which might explain why this mutation has not been found in large cohort studies from China[35] and Australia[34 36] that have previously described other rare variants in *ARPP21* or *GLT8D1*, but were unable to confirm their pathogenic role. Large ALS cohorts have been previously used to identify novel disease-causing variants, such as those included in project MiNE which comprises samples from diverse countries across the world (from Spain, but also Brazil, the USA, Australia and Russia, among others). As demonstrated by the differences in the presence of the *ARPP21* risk haplotype around the world, using such diverse populations may have precluded the identification of extremely rare disease-causing variants using rare genetic burden association analyses. It is important to note that our approach was based on the identification of a higher than expected burden of ALS cases in a remote region. For this reason, we implemented a unique strategy that diverged from previously used methods.

To date, the most recent reviews on the genetics of ALS do not consider *ARPP21* to be an ALS-causing gene,[8] nor was it included in a recent study comprising a set of 90 ALS-associated genes selected from multiple databases.[37] In addition, an OMIM number has not yet been assigned to the *ARPP21* gene, and in the ALS Online Database (ALSoD; https://alsod.ac.uk)[38] the gene is still categorised as 'tenuous evidence', meaning that it has been associated with ALS in small studies that have not stood up to replication. Our study demonstrates that *ARPP21*, or at least the p.Pro529Leu mutation in this gene, is consistently associated with ALS and should be considered in future genetic studies.

The ARPP21 protein is an RNA-binding protein widely expressed in the human brain[39] and based on the results of our study, adds to the long list of RNA-binding proteins implicated in ALS, such as TDP-43, TIA1 or FUS. It is involved in RNA metabolism, an event known to be deregulated and central to ALS pathophysiology.[40] Importantly, a common set of ARPP21 target RNAs with related functions have been identified through iCLIP. These RNAs have a prominent role in messenger RNA splicing and processing as previously demonstrated by gene ontology enrichment analyses.[39] In addition, in cells expressing ARPP21 treated with stress inductors (arsenite, clotrimazole or heat shock), the recruitment of ARPP21 to stress granules is induced. Importantly, the C-terminal sequence of the protein (where the p.Pro529Leu mutation is located) is necessary and sufficient to induce its recruitment to these membraneless dynamic structures. Although its domains have not been fully characterised, the mutation described in our study (p.Pro529Leu) is located in the C-terminal sequence and in close proximity to an intrinsically disordered region with compositional bias to Proline residues (low-complexity regions) of ARPP21. We speculate that this mutation might result in an increased propensity of ARPP21 to interact with and localise to stress granules, a mechanism

previously described in ALS-related mutations in other RNA-binding proteins such as FUS,[41] TIA1[42] or TDP-43, and directly related to the pathophysiology of ALS[43] and other neurodegenerative diseases.[44] Thus, beyond the implication of ARPP21 in RNA metabolism and processing, we suggest that this mutation could alter the dynamics of stress granules by increasing their aberrant formation and thus impacting on ALS pathogenesis.

Our study has some limitations. First, although our results suggest the causal role of the mutation described herein, DNA was available and genetic testing was consented to only in a few family members, thus precluding a complete segregation study. Given the variability in the age of onset of the disease, segregation analysis, especially involving some of the younger asymptomatic relatives, should be interpreted cautiously. Nevertheless, we strongly believe that the presence of the same variant in 10 patients with ALS (9 fALS) from 7 unrelated families in a small geographical area with a high incidence of ALS is a strong argument in favour of its pathogenicity. Further studies are required both in our country and worldwide to elucidate the contribution and frequency of this and other variants in *ARPP21* to the genetic basis of the disease, as well as functional validation of the mutation pathogenicity. Second, clinical information on patients' families was collected retrospectively and was often incomplete, especially regarding cognitive and behavioural symptoms, for which specific standardised cognitive-behavioural tests could have provided a wider phenotypical spectrum of the variant.

In conclusion, our results support the causative role of a mutation (p.Pro529Leu) in *ARPP21* in ALS. This work provides robust epidemiological and genetic arguments in favour of mutation pathogenicity. We also discuss disease-causing mechanisms associated with the mutation, supporting the role of *ARPP21* as an independent, novel ALS-causing gene, adding to the complex genetic architecture of the disease and expanding the landscape of RNA-binding proteins altered in ALS.

Author affiliations
[1]Sant Pau Memory Unit, Department of Neurology, Hospital de la Santa Creu i Sant Pau, Institut d'Investigació Biomèdica Sant Pau (IIB Sant Pau), Barcelona, Spain
[2]Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain
[3]Motor Neuron Disease Clinic, Neuromuscular Diseases Unit, Department of Neurology, Hospital de la Santa Creu i Sant Pau, Institut d'Investigació Biomèdica Sant Pau (IIB Sant Pau), Barcelona, Spain
[4]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Madrid, Spain
[5]Department of Medicine, Universitat Autonoma de Barcelona, Barcelona, Spain
[6]Neuromuscular and Motor Neuron Diseases Research Group, Department of Neurology, Hospital Universitario de Navarra, Pamplona, Spain
[7]Health Research Institute of Navarra (IdisNa), Pamplona, Spain
[8]Department of Neurology, Hospital San Pedro, Logroño, La Rioja, Spain
[9]Genetics Department, Hospital de la Santa Creu i Sant Pau, Institut d'Investigació Biomèdica Sant Pau (IIB Sant Pau), Barcelona, Spain
[10]Genome Instability and DNA Repair Group, Department of Genetics and Microbiology, Universitat Autonoma de Barcelona, Barcelona, Spain
[11]Division of Neuropathology and Neurochemistry, Department of Neurology, Medical University of Vienna, Vienna, Austria

X Oriol Dols-Icardo @Uri713 @SantPauMemory and Álvaro Carbayo @NMDSantPau

Acknowledgements We are indebted to patients and their families for their cooperation in the study. We thank Carolyn Newey for her help with language supervision and correction, which greatly improved the quality of our work. This work is generated within the European Reference Network for Neuromuscular Diseases.

Contributors Study conceptualisation and design: OD-I, ÁC, IJ, EG and RR-G were actively involved in study conceptualisation and design. Data acquisition, analysis and interpretation: OD-I, ÁC, OB-M, EÁ-S, MALP, SB, BR-S, IC, JT-S, MC-T, MC-Á, AV, LL, IP-L, LT, NV-T, LM, RR-G. Drafting and critical revision: OD-I, ÁC, BR-S, JT-S, MC-T, MC-Á, AV, LL, IP-L, LT, NV-T, LM, SR-G, II-G, EC-V, EG and RR-G. Study guarantors: RR-G, OD-I.

Funding This study was funded by the Instituto de Salud Carlos III (Ministerio de Asuntos Económicos y Transformación Digital, Gobierno de España) through the projects PI19/01543, PI23/00845 to RR-G, and PI18/00326, PI21/01395 to OD-I, cofunded by the European Regional Development Fund/European Social Fund (ERDF/ESF), 'A way to make Europe'/'Investing in your future'. OD-I receives funding from the Fundación Española para el Fomento de la Investigación de la Esclerosis Lateral Amiotrófica (FUNDELA - 'Por un mundo sin ELA'), Fundación HNA ('Premio Investigación científica de salud') and the Alzheimer's Association (AARF-22-924456). II-G is a senior Atlantic Fellow for Equity in Brain Health at the Global Brain Health Institute (GBHI), and receives funding from the GBHI, the Alzheimer's Society (GBHI ALZ UK-21-720973) and the Alzheimer's Association (AACSF-21-850193). II-G was also supported by the Juan Rodés Contract (JR20/0018) and PI21/00791 from Instituto de Salud Carlos III (Ministerio de Asuntos Económicos y Transformación Digital, Gobierno de España).

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein), or of any geographical or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Consent obtained directly from patient(s).

Ethics approval This study involves human participants and was approved by Hospital de la Santa Creu i Sant Pau Ethics Comittee (Study ID: 20/016 HSCSP). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Oriol Dols-Icardo http://orcid.org/0000-0003-2656-8748
Álvaro Carbayo http://orcid.org/0000-0001-9282-8603
Marta Caballero-Ávila http://orcid.org/0000-0001-9850-8504
Sara Rubio-Guerra http://orcid.org/0000-0001-7652-8029
Ignacio Illán-Gala http://orcid.org/0000-0002-5418-2052
Elena Cortés-Vicente http://orcid.org/0000-0002-1428-1072
Ricard Rojas-García http://orcid.org/0000-0003-1411-5573

REFERENCES

1 van Es MA, Hardiman O, Chio A, *et al*. Amyotrophic lateral sclerosis. *Lancet* 2017;390:2084–98.
2 Borrego-Écija S, Turon-Sans J, Ximelis T, *et al*. Cognitive decline in Amyotrophic lateral sclerosis: neuropathological substrate and genetic determinants. *Brain Pathol* 2021;31:e12942.
3 Cortés-Vicente E, Turon-Sans J, Gelpi E, *et al*. Distinct clinical features and outcomes in motor neuron disease associated with behavioural variant frontotemporal dementia. *Dement Geriatr Cogn Disord* 2018;45:220–31.
4 Carbayo Á, Borrego-Écija S, Turon-Sans J, *et al*. Clinicopathological correlates in frontotemporal lobar degeneration: motor neuron disease spectrum. *Brain* 2024.:awae011.
5 Phukan J, Elamin M, Bede P, *et al*. The syndrome of cognitive impairment in amyotrophic lateral sclerosis: a population-based study. *J Neurol Neurosurg Psychiatry* 2012;83:102–8.
6 Byrne S, Bede P, Elamin M, *et al*. Proposed criteria for familial amyotrophic lateral sclerosis. *Amyotroph Lateral Scler* 2011;12:157–9.
7 Ling S-C, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* 2013;79:416–38.
8 Akçimen F, Lopez ER, Landers JE, *et al*. Amyotrophic lateral sclerosis: translating genetic discoveries into therapies. *Nat Rev Genet* 2023;24:642–58.

9 Chia R, Chiò A, Traynor BJ. Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications. *Lancet Neurol* 2018;17:94–102.
10 Feldman EL, Goutman SA, Petri S, *et al*. Amyotrophic lateral sclerosis. *Lancet* 2022;400:1363–80.
11 Pradas J, Puig T, Rojas-García R, *et al*. Amyotrophic lateral sclerosis in catalonia: a population based study. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 2013;14:278–83.
12 Jericó I, Elizalde-Beiras I, Pagola I, *et al*. Clinical features and incidence trends of amyotrophic lateral sclerosis in Navarre, Spain, 2007–2018: a population-based study. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 2021;22:401–9.
13 Logroscino G, Traynor BJ, Hardiman O, *et al*. Descriptive epidemiology of amyotrophic lateral sclerosis: new evidence and unsolved issues. *J Neurol Neurosurg Psychiatry* 2008;79:6–11.
14 Brooks BR, Miller RG, Swash M, *et al*. El escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders* 2000;1:293–9.
15 Instituto Nacional de Estadística (INE). 2024. Available: https://www.ine.es [Accessed 16 Feb 2024].
16 Instituto Nacional de Estadística (INE). Indicators for census tracks. Population and Housing Census La Rioja 2009-2022, 2024. Available: https://www.ine.es/jaxiT3/Tabla.htm?t=2879&L=0] [Accessed 16 Feb 2024].
17 Strong MJ, Grace GM, Freedman M, *et al*. Consensus criteria for the diagnosis of frontotemporal cognitive and behavioural syndromes in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler* 2009;10:131–46.
18 Rascovsky K, Hodges JR, Knopman D, *et al*. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 2011;134:2456–77.
19 Gorno-Tempini ML, Hillis AE, Weintraub S, *et al*. Classification of primary progressive aphasia and its variants. *Neurology* 2011;76:1006–14.
20 Li H. Aligning sequence reads, clone sequences and assembly Contigs with BWA-MEM. *arXiv* 2013;1303.
21 McKenna A, Hanna M, Banks E, *et al*. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
22 Cingolani P, Platts A, Wang LL, *et al*. A program for annotating and predicting the effects of single nucleotide polymorphisms, Snpeff. *Fly* 2012;6:80–92.
23 Cingolani P, Patel VM, Coon M, *et al*. Using drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, Snpsift. *Front Genet* 2012;3:35.
24 Liu X, Jian X, Boerwinkle E. dbNSFP V2.0: a database of human non-synonymous Snvs and their functional predictions and annotations. *Hum Mutat* 2013;34:E2393–402.
25 Landrum MJ, Lee JM, Benson M, *et al*. Clinvar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;46:D1062–7.
26 Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, *et al*. Predicting splicing from primary sequence with deep learning. *Cell* 2019;176:535–48.
27 Lek M, Karczewski KJ, Minikel EV, *et al*. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
28 Peña-Chilet M, Roldán G, Perez-Florido J, *et al*. CSVS, a Crowdsourcing database of the Spanish population genetic variability. *Nucleic Acids Res* 2021;49:D1130–7.
29 Pedersen BS, Bhetariya PJ, Brown J, *et al*. Somalier: rapid relatedness estimation for cancer and Germline studies using efficient genome sketches. *Genome Med* 2020;12:62.
30 Li H, Durbin R. Fast and accurate short read alignment with burrows–Wheeler transform. *Bioinformatics* 2009;25:1754–60.
31 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
32 Dols-Icardo O, García-Redondo A, Rojas-García R, *et al*. Analysis of known Amyotrophic lateral sclerosis and Frontotemporal dementia genes reveals a substantial genetic burden in patients manifesting both diseases not carrying the C9Orf72 expansion Mutation. *J Neurol Neurosurg Psychiatry* 2018;89:162–8.
33 Abramzon YA, Fratta P, Traynor BJ, *et al*. The overlapping genetics of amyotrophic lateral sclerosis and frontotemporal dementia. *Front Neurosci* 2020;14:42.
34 Cooper-Knock J, Moll T, Ramesh T, *et al*. Mutations in the glycosyltransferase domain of Glt8D1 are associated with familial amyotrophic lateral sclerosis. *Cell Rep* 2019;26:2298–306.
35 Li W, Liu Z, Sun W, *et al*. Mutation analysis of Glt8D1 and Arpp21 genes in amyotrophic lateral sclerosis patients from Mainland China. *Neurobiol Aging* 2020;85:156.
36 Chan Moi Fat S, McCann EP, Williams KL, *et al*. Genetic analysis of Glt8D1 and Arpp21 in Australian familial and sporadic Amyotrophic lateral sclerosis. *Neurobiol Aging* 2021;101:297.
37 Van Daele SH, Moisse M, van Vugt JJFA, *et al*. Genetic variability in sporadic amyotrophic lateral sclerosis. *Brain* 2023;146:3760–9.
38 Wroe R, Wai-Ling Butler A, Andersen PM, *et al*. ALSOD: the amyotrophic lateral sclerosis online database. *Amyotroph Lateral Scler* 2008;9:249–50.
39 Rehfeld F, Maticzka D, Grosser S, *et al*. The RNA-binding protein Arpp21 controls dendritic branching by functionally opposing the miRNA it hosts. *Nat Commun* 2018;9:1235.
40 Ito D, Hatano M, Suzuki N. RNA binding proteins and the pathological cascade in ALS/FTD neurodegeneration. *Sci Transl Med* 2017;9:eaah5436.
41 Baron DM, Kaushansky LJ, Ward CL, *et al*. Amyotrophic lateral sclerosis-linked FUS/TLS alters stress granule assembly and dynamics. *Mol Neurodegener* 2013;8:30.
42 Mackenzie IR, Nicholson AM, Sarkar M, *et al*. Tia1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. *Neuron* 2017;95:808–16.
43 Aulas A, Vande Velde C. Alterations in stress granule dynamics driven by TDP-43 and FUS: a link to pathological inclusions in ALS *Front Cell Neurosci* 2015;9:423.
44 Wolozin B, Ivanov P. Stress granules and neurodegeneration. *Nat Rev Neurosci* 2019;20:649–66.