



# Identification of *UBAP1* mutations in juvenile hereditary spastic paraplegia in the 100,000 Genomes Project

Thomas Bourinaris<sup>1</sup> · Damian Smedley<sup>2</sup> · Valentina Cipriani<sup>2</sup> · Isabella Sheikh<sup>1</sup> · Alkyoni Athanasiou-Fragkouli<sup>1</sup> · Patrick Chinnery<sup>3,4</sup> · Huw Morris<sup>5</sup> · Raquel Real<sup>5</sup> · Victoria Harrison<sup>6</sup> · Evan Reid<sup>7</sup> · Nicholas Wood<sup>5</sup> · Genomics England Research Consortium · Jana Vandrovцова<sup>1</sup> · Henry Houlden<sup>1</sup> · Arianna Tucci<sup>2</sup>

Received: 17 January 2020 / Revised: 29 July 2020 / Accepted: 15 August 2020 / Published online: 15 September 2020  
© The Author(s) 2020. This article is published with open access

## Abstract

Hereditary spastic paraplegia (HSP) is a group of heterogeneous inherited degenerative disorders characterized by lower limb spasticity. Fifty percent of HSP patients remain yet genetically undiagnosed. The 100,000 Genomes Project (100KGP) is a large UK-wide initiative to provide genetic diagnosis to previously undiagnosed patients and families with rare conditions. Over 400 HSP families were recruited to the 100KGP. In order to obtain genetic diagnoses, gene-based burden testing was carried out for rare, predicted pathogenic variants using candidate variants from the Exomiser analysis of the genome sequencing data. A significant gene-disease association was identified for *UBAP1* and HSP. Three protein truncating variants were identified in 13 patients from 7 families. All patients presented with juvenile form of pure HSP, with median age at onset 10 years, showing autosomal dominant inheritance or de novo occurrence. Additional clinical features included parkinsonism and learning difficulties, but their association with *UBAP1* needs to be established.

These authors contributed equally: Thomas Bourinaris, Damian Smedley, Valentina Cipriani, Henry Houlden, Arianna Tucci

Members of the Genomics England Research Consortium are listed in Supplementary information.

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41431-020-00720-w>) contains supplementary material, which is available to authorized users.

✉ Arianna Tucci  
a.tucci@qmul.c.uk

- <sup>1</sup> Department of Neuromuscular Disorders, UCL Institute of Neurology, Queen Square, London, UK
- <sup>2</sup> William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK
- <sup>3</sup> Department of Clinical Neurosciences, School of Clinical Medicine, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK
- <sup>4</sup> Medical Research Council Mitochondrial Biology Unit, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK
- <sup>5</sup> Department of Clinical and Movement Neurosciences, UCL Institute of Neurology, Queen Square, London, UK
- <sup>6</sup> Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton SO16 5YA, UK
- <sup>7</sup> Department of Medical Genetics and Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

## Introduction

Hereditary spastic paraplegia (HSP) represents a wide spectrum of rare inherited neurodegenerative conditions, typically characterized by progressive limb spasticity. HSP is traditionally classified into pure and complicated forms, depending on the presence of additional clinical features, such as cognitive decline, cerebellar ataxia, peripheral neuropathy, or parkinsonism. To date, more than 80 genes have been associated with HSP under various modes of inheritance, including autosomal recessive, autosomal dominant (AD), or X-linked [1].

Recent advances in molecular analysis techniques have led to the identification of new genetic loci associated with HSP. Yet, for up to 70% of cases clinically diagnosed as HSP the genetic cause remains unknown [2].

The 100,000 Genomes Project (100KGP) was established to deliver advanced genomic testing for NHS patients with rare disease and cancer. One of the aims of the rare disease arm of the project was to make molecular diagnoses and novel gene discoveries through whole genome sequencing. Over 60,000 participants with genetically undiagnosed rare disease have been recruited [3], including over 400 HSP families.

Here, we describe 13 patients from 7 families with HSP recruited to the 100KGP, who carry 3 truncating variants in

*UBAPI*, a newly identified gene that causes pure forms of juvenile HSP [4]. Variant interpretation analysis had been initially carried out via the 100KGP automated pipeline in known HSP genes [5], and failed to identify a genetic diagnosis in all families. An ongoing re-analysis of the data in the research environment using gene-based burden testing for rare, predicted damaging variants allowed the rapid identification of the three pathogenic *UBAPI* variants in seven families.

## Materials and methods

### Patients

Following ethical approval (14/EE/1112), consenting participants were recruited to the 100KGP cohort. Standardized baseline clinical data were recorded at the time of recruitment using the Human Phenotyping Ontology, according to the HSP data model (Supplementary information). Further, additional information was collected retrospectively by contacting the recruiting clinician for each patient (summarized in Table 1). Patients had been previously tested negative for single nucleotide variants, small deletions/insertions and CNVs in genes from the PanelApp [5] curated virtual gene panel for HSP. CNV calls were generated by Genomics England using Manta and Canvas software (Illumina) and processed using custom R scripts to detect CNVs overlapping the *SPAST* gene [5]. No likely pathogenic variants were identified.

### Gene-based burden testing

A statistical analysis framework was developed to detect enrichment of rare, predicted pathogenic variants in novel genes for specific diseases. The analysis was run on all rare, coding variants that segregated with disease as expected for each possible mode of inheritance (minor allele frequencies,  $MAF < 0.1\%$  dominant,  $< 1\%$  recessive), identified from Exomiser [6, 7] analysis of each case's whole genome data. The following case-control association analysis framework was used: "case" sets were defined as all 100KGP probands recruited under HSP, while corresponding "control" sets were all recruited 100KGP probands except those recruited under a neurological disorder. We tested for enrichment of rare variants using Fisher's exact test under four scenarios: (1) rare, predicted pathogenic variants (Exomiser variant score  $> 0.8$  corresponding to rare variants that are predicted to be pathogenic by in silico prediction tools REVEL [8] and/or MVP [9]), (2) rare variants in a constrained coding region [10], (3) rare, predicted loss of function (LoF) variants, (4) rare, de novo variants. For the latter, only trios or larger families where de novo calling was possible were considered. The Benjamini and Hochberg method was used

to correct for multiple testing; an overall false discovery rate adjusted  $q$  value threshold of 0.10 was used for claiming significant gene-disease associations.

All variants described in this study have been submitted to ClinVar.

## Results

In the context of an ongoing gene-based burden analysis of the rare disease component of the 100KGP data, a significant gene-disease association was identified for *UBAPI* and HSP. A significant excess of LoF variants was observed in 5 out of 417 HSP cases compared to 16,449 non-neurological 100KGP probands used as controls (odds ratio,  $OR = 66.5$ ;  $q$  value = 0.002). Through further inspection of newly available 100KGP data and other neurological probands that were excluded from the case-control analysis dataset, we identified two additional patients with spastic paraplegia and *UBAPI* truncating variants. One of these patients had been recruited under HSP diagnosis and the other as familial Parkinson's disease (PD), but was also affected by HSP.

In total, seven families carrying three different truncating variants in *UBAPI* were identified in the 100KGP data (Table 1, ClinVar SUB7456898). Families 1 and 2 were consistent with de novo occurrence of HSP (Fig. 1a), while the pedigrees of families 3–7 suggested an AD pattern of inheritance (Fig. 1a) that was confirmed by segregation analysis in family 5 (Fig. 1b). The identified variants included: c.426\_427del; p.(Lys143SerfsTer15) in two families, c.535G > T; p.(Glu179Ter) in four families and c.373C > T; p.(Gln125Ter) in one family (NM\_016525.4; NP\_057609.2) (Fig. 1a).

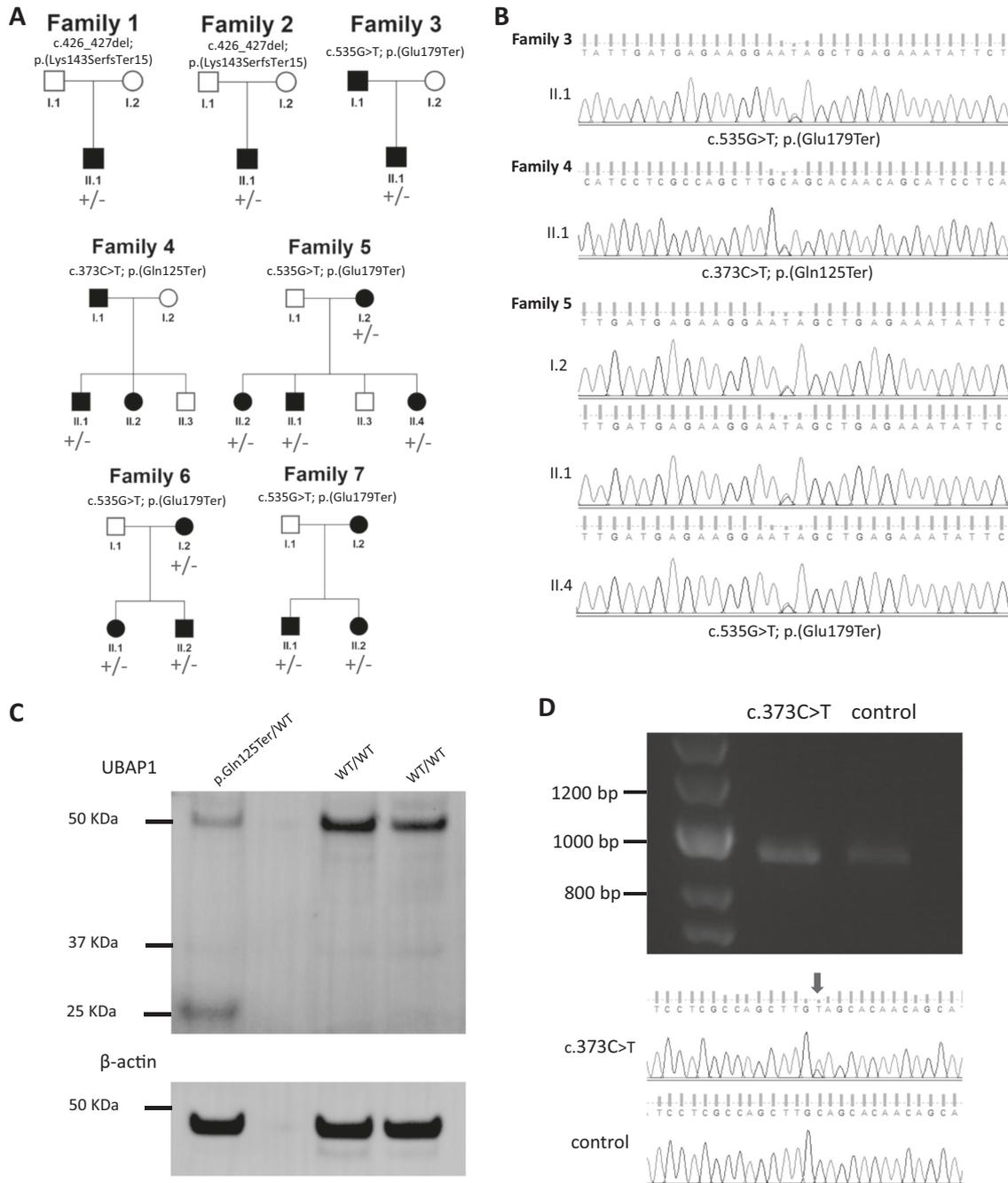
The median age at onset was 10 years (interquartile range 9–11 years). In most cases the onset occurred during childhood, with some individuals having a later age at onset, including early adulthood. The latest age at onset was 30 years (Table 1).

Overall, clinical history and examination of the 13 individuals were consistent with uncomplicated HSP phenotype. The first symptoms included difficulties with running and walking, with frequent falls and leg scissoring also reported. Disease course was similar in all families, with slowly progressive lower limb spasticity and weakness, except proband of family 1, who progressed over 2–3 years to wheelchair use. Additional symptoms included learning difficulties in family 5, reduced light touch in the lower limb distally and upper limb areflexia in one of the affected members of family 6, although no results of nerve conduction studies were available. Intrafamilial heterogeneity was observed, with a milder phenotype usually associated with a later age of onset. At the time of examination, the most frequent signs included lower limb weakness, extensor

**Table 1** Detailed clinical characterization of the individuals included in the study and carrying truncating mutations in *UBAP1* (NM\_016525.4; NP\_057609.2).

Family/ID	F1-IL1	F2-IL1	F3-IL1	F4-IL1	F5-IL2	F5-IL1	F5-IL2	F5-IL4	F6-IL2	F6-IL1	F6-IL2	F7-IL1	F7-IL2
Variant	c.426,427del; p. (Lys143SerfsTer15)	c.426,427del; p. (Lys143SerfsTer15)	c.535G>T; p. (Glu179Ter)	c.373C>T; p. (Gln125Ter)	c.535G>T; p. (Glu179Ter)	c.535G>T; p. (Glu179Ter)	c.535G>T; p. (Glu179Ter)	c.535G>T; p. (Glu179Ter)	c.535G>T; p. (Glu179Ter)	c.535G>T; p. (Glu179Ter)	c.535G>T; p. (Glu179Ter)	c.535G>T; p. (Glu179Ter)	c.535G>T; p. (Glu179Ter)
CADD score (phred)	29.8	29.8	37	36	37	37	37	37	37	37	37	37	37
Sex	M	M	M	M	F	M	F	F	F	F	M	M	F
Ethnicity	Caucasian	Caucasian	Caucasian	Ashkenazi	Caucasian	Caucasian	Caucasian	Caucasian	Caucasian	Caucasian	Caucasian	Caucasian	Caucasian
Family history	De novo	De novo	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive
Age at examination (years)	16	69	33	43	42	23	24	18	n/a	36	n/a	10	11
Age at onset (years)	9	9	18	11	11	10	10	9	n/a	30	11	9	7
Development	Normal. Toe walker	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal
Presenting symptom(s)	Gait difficulty, tiptoe walking	Difficulty running	Mild stumbling	Gait difficulty	Gait difficulty	Gait difficulty	Difficulty running	Gait difficulty, frequent falls	n/a	n/a	n/a	Gait difficulty, falls, foot pain	Unsteady gait, falls, ankle pain
Lower limb power	4+/5 proximal, 4/5 distal	Reduced in a pyramidal distribution	Mild pyramidal weakness	Pyramidal weakness	Reduced	Moderate weakness	Reduced	Normal	n/a	Normal	Normal	Normal	Normal
Lower limb spasticity	Markedly present	Present	Present	Markedly present	Present	Present	Present	Present	n/a	No	Present	Present	Present
Lower limb reflexes	Very brisk, sustained ankle clonus	Brisk, sustained ankle clonus	Increased, sustained ankle clonus	Increased	Increased	Increased, sustained ankle clonus	Increased	Increased	n/a	Very brisk	Very brisk	Increased, mild clonus	Increased, clonus
Plantar reflex	Withdrawal	Extensor	Extensor	Extensor	n/a	Extensor	n/a	Extensor/withdrawal	n/a	Extensor	Extensor	n/a	Extensor
Upper limb power	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	n/a	Normal	Normal	Normal	Normal
Upper limb tone	Normal	Extrapyramidal rigidity	Normal	Normal	Normal	Normal	Normal	Normal	n/a	Normal	Normal	Normal	Normal
Upper limb reflexes	Normal	n/a	Normal	Brisk	Normal	Normal	Normal	Normal	n/a	Normal	Absent	Normal	Normal
Bladder dysfunction	No	Suprapubic catheter	Urinary urgency	No	Urinary urgency	No	No	No	n/a	no	no	no	Urinary urgency
Sensory dysfunction	No	No	No	No	No	No	No	No	n/a	No	distal numbness	No	No
Cognitive impairment	No	No	No	No	No	No	No	No	n/a	No	No	No	No
Ataxia	No	No	No	No	No	No	No	No	n/a	No	No	No	No
Dysarthria	No	No	No	No	No	No	No	No	n/a	No	No	No	No
Additional features	None	Parkinsonism	Occasional cramps	None	Cramps	Learning difficulties, spasms	Learning difficulties	None	n/a	Strabismus	n/a	Severe bilateral valgus foot deformity	None
Progression	Wheelchair use after 2–3 years	Wheelchair dependant from age 55	Walks without aids	Walks without aids	Walks with a single crutch	Walks without aids	Walks without aids	Walks without aids	n/a	Walks without aids	Walks without aids	Walks with crutches, occasionally uses wheelchair	Walks with crutches
Imaging	MRI brain and spine normal	MRI brain: mild generalized cortical subcortical volume loss	MRI brain normal	MRI brain normal	n/a	n/a	n/a	n/a	n/a	MRI brain normal	n/a	n/a	n/a

M male, F female, n/a not available.



**Fig. 1** Pedigrees, DNA, RNA and protein analysis of UBAP1. **a** Pedigrees, **b** Sanger Sequencing, **c** Western Blot, **d** RNA. **a** Pedigrees of families reported in this study. Squares represent males, circles represent females, shaded shapes represent individuals with HSP, and unshaded shapes represent individuals without HSP. **b** Electropherograms of Sanger sequencing performed for the validation of variants identified in families 3, 4, and 5. **c** Western Blot analysis for the proband of family 4 (p.Gln125Ter/WT) and two healthy controls (WT/WT), showing reduced expression of normal size ubiquitin-associated protein 1 (UBAP1) and the presence of a smaller additional band in the proband, indicating the

presence of truncated protein. Loading control with  $\beta$ -actin was performed to confirm that the same amount of the total protein was loaded for all three samples. **d** Gel electrophoresis of PCR products amplified from cDNA synthesized from RNA of the proband of family 4 (c.373C>T) and a healthy control, showing the presence of amplicons of target sequence in both samples. Sanger sequencing of the same PCR products confirm the result as well as the presence of the variant in the proband, as shown by the respective electropherograms. WT wild type; KDa kilodalton; bp base pairs.

plantars and brisk reflexes of the lower limbs. Only the proband of family 4 presented brisk upper limb reflexes, while none of the individuals of our cohort had spasticity or weakness of the upper limbs. Bladder dysfunction, mainly presenting with urinary urgency, was also reported in three individuals.

Interestingly, in two of the families an extrapyramidal syndrome consistent with typical Parkinson Disease presented later on in the disease course. The proband of family 2 was diagnosed with levodopa-responsive PD in his mid-1950s, with predominantly left-sided rest tremor, bradykinesia and cogwheel rigidity particularly in the upper limb, facial hypomimia, and normal eye movements. The DaTscan was reported abnormal for bilaterally reduced uptake of the tracer in the striatum. The father of the proband of family 3, who was also affected by HSP, was reported to have PD with onset in early 1970s. Each family carried a different pathogenic variant. The size of this analysis does not allow to correlate the presence of parkinsonism with the identified pathogenic variants in *UBAP1*.

A Western Blot for the proband of family 4, carrying a heterozygous c.373C > T; p.Gln125Ter variant, revealed the presence of truncated protein along with reduced amount of normal size protein (Fig. 1c). PCR of c.DNA, synthesized from RNA isolated from the same individual, showed that mRNA escapes nonsense mediated decay (Fig. 1d), as previously reported [4].

### Discussion

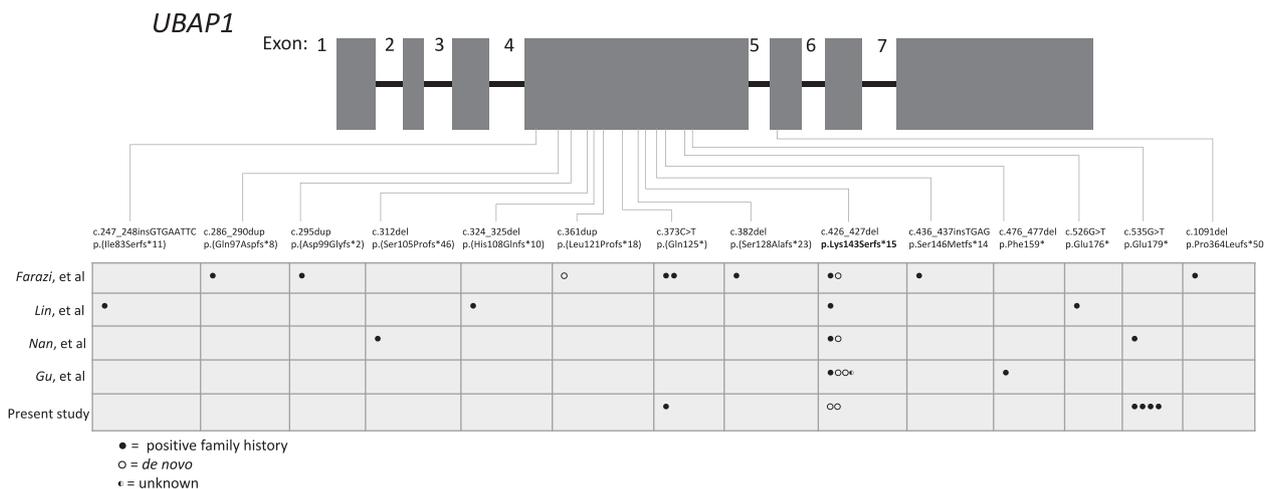
Here, we describe the genetic and clinical features of a cohort of UK HSP patients due to *UBAP1* protein truncating variants. The 100KGP is a research project to implement WGS in a national healthcare system. Patient with a clinical diagnosis of HSP were recruited from local hospitals in England. In most cases, the relevant HSP genes were sequenced prior to recruitment. This dataset represents one of the largest HSP cohort with WGS data from a single

country. In this dataset the prevalence of *UBAP1* in HSP is 1.7% (7 families out of 417).

Ubiquitin-associated protein 1 (encoded by *UBAP1*) constitutes a subunit of mammalian endosomal sorting complex required for transport I (ESCRT-I) and is involved in endosomal dynamics in neurons. Furthermore, it is involved in proteasomal degradation of ubiquitinated proteins. Both cellular pathways have been involved in HSP and our study further confirms the importance of such pathways in this disease [11].

While we were conducting this study, truncating *UBAP1* variants were reported in HSP families of diverse geographic origin, including the three variants we are reporting in this study [4, 12–14]. Thus far, and including our study, 30 families with AD HSP have been reported (Supplementary Table 1). The majority of the patients present with a pure form of juvenile-onset HSP. Of all cases reported, the average age at onset of symptoms was 9 years (interquartile range 7–11 years). Adult onset has been reported in two cases only (onset at age 30 and 61, respectively). Although a pure HSP phenotype has been reported in most families, two of the previously reported families additionally presented with cerebellar symptoms, while two of the families reported in this study presented with symptoms typical of an extrapyramidal disorder, this was later on in the disease course. While most families present with positive family history, de novo occurrence is reported in seven families. Incomplete penetrance has been described in three families.

In total, 14 different truncating variants have been reported so far. All but one of these variants occur within exon 4 of *UBAP1* (Fig. 2). The most frequent mutation identified so far by all studies is the p.(Lys143Serfs\*15), reported in 11 families out of 30 across different populations. In the current study, representing HSP in the British population, the most



**Fig. 2 Schematic representation of *UBAP1* with all variants reported in this study and previously.** The table indicates the number of families carrying each variant. Each circle represents one family.

Shaded circles represent positive family history, unshaded circles represent de novo occurrence, and half-shaded circles indicate unknown family history. Exons are numbered like in Fard et al. 2019 [4].

frequent mutation is p.(Glu179Ter), only previously described in one family in the Japanese population [13].

Our study provides further evidence that *UBAP1* causes mostly pure forms of AD HSP. Among the known genes that cause pure forms of HSP, with 30 families described so far, *UBAP1* is a common cause of AD juvenile pure HSP.

**Funding** The study was financially supported by grants received by the Medical Research Council (grant codes MR/S01165X/1, MR/S005021/1, and G0601943).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Hedera P. Clinical characteristics and recommended Treatment: hereditary spastic paraplegia. 1993;1–31. <http://www.ncbi.nlm.nih.gov/books/NBK1509/>.
- Schüle R, Wiethoff S, Martus P, Karle KN, Otto S, Klebe S, et al. Hereditary spastic paraplegia: clinicogenetic lessons from 608 patients. *Ann Neurol*. 2016;79:646–58.
- Turnbull C, Scott RH, Thomas E, Craig C, Hamblin A, Neill AO, et al. The 100,000 Genomes Project: bringing whole genome sequencing to the NHS. 2018;1687:1–7. <https://doi.org/10.1136/bmj.k1687>.
- Farazi Fard MA, Rebelo AP, Buglo E, Nemati H, Dastsooz H, Gehweiler I, et al. Truncating mutations in *UBAP1* cause hereditary spastic paraplegia. *Am J Hum Genet*. 2019;104:767–73.
- Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet*. 2019;51. <https://doi.org/10.1038/s41588-019-0528-2>.
- Robinson PN, Köhler S, Oellrich A, Mouse S, Project G, Wang K, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res*. 2014;24:340–8.
- Bone WP, Washington NL, Buske OJ, Adams DR, Davis J, Draper D, et al. Computational evaluation of exome sequence data using human and model organism phenotypes improves diagnostic efficiency. *Genet Med*. 2016;18:608–17.
- Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. ARTICLE REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99:877–85. <https://doi.org/10.1016/j.ajhg.2016.08.016>.
- Qi H, Chen C, Zhang H, Long JJ, Chung WK, Guan Y, et al. MVP: predicting pathogenicity of missense variants by deep neural networks. *bioRxiv*. 2018;259390. <http://biorxiv.org/content/early/2018/02/02/259390.abstract>.
- Havrilla JM, Pedersen BS, Layer RM, Quinlan AR. A map of constrained coding regions in the human genome. *Nat Genet*. 2019;51. <https://doi.org/10.1038/s41588-018-0294-6>.
- Agromayor M, Soler N, Caballe A, Kueck T, Freund SM, Allen MD, et al. The *UBAP1* subunit of ESCRT-I interacts with ubiquitin via a SOUBA domain. *Structure*. 2012;20:414–28.
- Lin X, Su H, Dong E, Lin X, Zhao M, Yang C, et al. Stop-gain mutations in *UBAP1* cause pure autosomal-dominant spastic paraplegia. *Brain*. 2019;142:2238–52.
- Nan H, Ichinose Y, Tanaka M, Koh K, Ishiura H, Mitsui J, et al. *UBAP1* mutations cause juvenile-onset hereditary spastic paraplegias (SPG80) and impair *UBAP1* targeting to endosomes. *J Hum Genet*. 2019;1055–65. <https://doi.org/10.1038/s10038-019-0670-9>.
- Gu S, Chen CA, Rosenfeld JA, Cope H, Launay N, Flanigan KM, et al. Truncating variants in *UBAP1* associated with childhood-onset nonsyndromic hereditary spastic paraplegia. *Hum Mutat*. 2019;1–9. <https://doi.org/10.1002/humu.23950>.